Functional and Structural Characterization of the Macaque Middle Face Patch

by

Paul Aparicio

B.A. Psychology

University of California, Berkeley 2001

Submitted to the Department of Brain and Cognitive Sciences in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2014

Signature of Author.....
.....
Department of Brain and Cognitive Sciences
December XX, 2013

Certified by.....
.....
James J. DiCarlo
Professor of Neuroscience
Thesis Supervisor

Accepted by....
.....
Matthew A. Wilson
Sherman Fairchild Professor Neuroscience and Picower Scholar
Director of Graduate Education for Brain and Cognitive Sciences

Functional and Structural Characterization of the Macaque Middle Face Patch

by

Paul Aparicio

Submitted to the Department of Brain and Cognitive Science on November 18, 2013 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Neuroscience

ABSTRACT

The middle face patch is a region of cortex in the ventral visual pathway of the Inferior Temporal lobe in the macaque brain. This region has been identified by functional MRI to respond preferentially to images of faces over non-face images, similar to functionally defined face selective regions in the human brain. In this thesis we spatially map the category selective preference of 100's of multiunit sites in the cortical region localized to the fMRI face selective region with a novel X-ray imaging system. We observed evidence for an ~6mm region of cortex that was enriched with sites that demonstrate a category selective preference for images of faces. The number of face selective sites varied across the cortical region, and could peak as high as 96% near the center of the enriched zone to a baseline rate as low as 3% outside the face patch. Sites in the middle face patch displayed significant category selectivity for the conventional images of faces used in the experiment. Approximately 25% of the sites in the patch displayed high selectivity ($d' > 2$) for faces as compared to less than 1% of the sites sampled outside the patch. Given the limited image variability present in conventional image sets, we examined face detection performance in the middle face patch with a computationally non-trivial image set, that was nonetheless simple for human subjects. We found that under these conditions, sites in the middle face patch demonstrated a weak correlation to human face detection behavior. We conclude that the middle face patch is a region of cortex enriched with sites that participate in an intermediate level representation of faces.

Thesis Supervisor: James J. DiCarlo
Title: Professor of Neuroscience, Department Head

Contents

List of Figures

Chapter 1

Introduction

Human beings are fundamentally social animals. We depend on other people in countless ways for almost our entire existence. As infants, our parents must decipher our needs from little more than a few facial gestures and some non-semantic vocalizations. As children, we must recognize whom to listen to and whom to avoid in order to learn and grow. When we become older we must forge alliances with other individuals in order to mature and succeed. All of these behaviors crucially require the fast and efficient recognition of the potentially countless individuals that we will interact with over the course of our lives. Fortunately, nature has provided a unique signature to aid in this endeavor: the face. As a communicative device, the face can provide a wide range of information (Bruce and Young, 1986) about other people (e.g. gender, emotional status, focus of attention). The face also provides us with an efficient way to recognize other individuals in order to keep track of the numerous contingencies and relationships that social individuals must navigate. Is it any wonder that the brain appears to have specialized mechanisms for supporting the visual perception of faces? This thesis will examine the spatial organization of a cortical region in the macaque brain hypothesized to be critical for face processing, and further ask to what extent it supports face detection behavior.

So why should we be motivated to study (and you to read about) this topic? One motivation for studying the neural representation of faces is that the system is directly open to study. What I mean by this is that, at this moment a number of fortuitous conditions are in place to study the neural mechanisms underlying face perception. First of all we have some idea of *where in the brain to look.* While the specific details are an open topic (of which this thesis makes some attempt to quantify), regions that display category selective responses for faces have been localized with brain imaging methods in humans and monkeys. Second *the category selective signal for faces is robust.* Scientific methods are most amenable to producing knowledge when there is a clear phenomena to study. Neurons in the category selective face areas produce a robust differential response to a class of stimuli that are, at least at first pass, obviously discernible. In other words we have a strong general intuition about what makes these neurons "tick." Thirdly, face selective regions in the macaque have *a correspondence to the human brain.* The use of functional magnetic resonance imaging (fMRI) has proven amenable in use on both humans and monkeys, and similar phenomena have been observed in both species using the *same methods.* It is therefore, possible to directly motivate and test experimental hypothesis across species, taking advantage of the different levels of study each species is uniquely positioned to study.

A second motivation to study the face system in monkeys is that in the domain of faces, the brain must solve the same complex and rich set of problems that have been well formulated in the domain of general object recognition. The everyday visual scene that humans (and monkeys) must navigate through is a cacophony of colors and textures that form objects, places and people. The brain needs to generalize over the potentially infinite variability in the visual input in order to recognize any given thing in the world. This problem can be considered a major problem in operationalizing object recognition (DiCarlo et al 2012). While it is not clear if the brain will solve the general object recognition problem in the same way that it solves face recognition,

understanding how the brain solves face recognition will be one tenable example of a solution, and is therefore, an important domain of research to pursue. The remainder of the chapter will review some of the key findings in previous research that have characterized the neural basis of face recognition, and explicate the problems that this thesis will explore. A major organizing theme of this review is in understanding the spatial organization of the putative face selective areas in the brains of monkeys and humans. I want to know why we presume to know that face perception is localizable to one (or a few) discreet areas, and how we know where to look for such a signal in the first place.

1.1 Prosopagnosia

The first often described piece of evidence that there is a discernible locus in the brain that specifically affects behavior with faces comes from the study of neurological patients who exhibit Prosopagnosia, or "face blindness." The term was first used by Bodamer (1947) to distinguish a specific type of visual agnosia, though reports of human patients with deficits in recognizing faces had been reported in numerous studies as far back as 1844, before the term was actually coined (as reviewed in Mayer and Rossion, 2013). Prosopagnosia could be summarized as a deficit in recognizing individuals from visual images of their faces without low level visual impairments. A classic prosopagnosia patient would not be able to recognize a friend from a line up if all the people in the line up wore the same clothes, were not allowed to move or speak, and had their heads covered so as to only reveal their faces. While exclusive disruption of visual behavior with non-face objects (e.g. normal performance on visual behavior with face images) has been reported in the visual agnosia literature (Moscovitch, Winocur and Behrmann, 1997), prosopagnosia itself is extremely rare. It is usually accompanied by other high level visual deficits such as achromatopsia, or selective loss

in color vision (Bouvier and Engel, 2006), and topographical disorientation, or selective deficits in spatial navigation behavior (Mayer and Rossion, 2013). Furthermore, prosopagnosia can accompany a wide range of behavioral performance with both face and non-face images, and is often associated with spatially expansive anatomical injuries. Prosopagnosia is defined by the inability to recognize *familiar faces* with visual information. Face blind patients have been reported to be able to perform face detection tasks and display subtle behavioral deficits on non-face object tasks. Calder and colleagues reported that extensive testing with one Prosopagnosic patient revealed that the lower half of the face was most informative in a detection task (e.g. the mouth) for the patient, while normal controls found the upper right area of the face most informative (Calder et al., 2000). This finding suggests that prosopagnosic patients can adopt non-optimal strategies to access domain specific information in the completion of behavioral tasks. Task strategy can be important for interpreting behavioral results, particularly in accounting for performance differences in behavior utilizing face and non-face categories. Behavioral tasks with faces are typically subordinate (distinguishing Jack from Jill), while similar tasks with non-face objects can be basic (distinguishing cars from dogs). Category specific deficits could then be the result of tasks differences with non-face objects (distinguishing basic level exemplars can be easier than distinguishing between subordinate exemplars), or with behavioral tasks that implicitly create or allow such differences as a behavioral strategy (i.e. distinguishing between cars and trucks). Examination of prosopagnosic patients in carefully controlled tasks has revealed subtle difference in RT measurements with patients to non-face distractors and reduced performance with both face and non-face objects when task difficulty was manipulated by reduced stimulus presentation time or memory load (Gauthier, Behrmann and Tarr, 1999). Close examination of the behavioral literature with prosopagnosic patients suggests a heterogeneity in performance of face

related behavioral tasks, which mirrors the spatial heterogeneity of the anatomical lesions observed in prosopagnosia.

While prosopagnosia is typically thought of as a unitary disorder, Damasio and colleagues (1989) identified three classes of prosopagnosia from metadata on symptoms and crude anatomical localization. These subdivisions were classified as: (1) *Associative*, defined as an impairment in identifying individual faces exclusively in the visual modality, and with bilateral lesions common to areas Broadman areas 17,18 and 37 (striate, peristriate, and occipitotemporal cortex). (2) *Amnesiac Associative*, defined as impairment in identifying individuals from stimuli in any modality, and with common lesions to areas 38, 20, 21, 22 (temporal pole, infero, middle and superior temporal cortex. (3) *Appreciative* defined as having a more general visual deficit, including deficits in identifying individuals from images of their faces. These subjects have common lesions in areas 39 and 37 (angular and occipitotemporal cortex). As might be expected, neurological insults that produce prosopagnosia are diffuse and widespread, though there is evidence that some areas may contribute more than others. An anatomical study that compared the lesion locations of over 50 reported cases of prosopagnosia, reported several suggestive findings (Bouvier and Engel, 2006). First, while the overlapping locations were widespread, there was a large overlap in the anatomical region where the Occipital Face Area (OFA) is typically observed with fMRI. The OFA (discussed more below), is an fMRI defined region of cortex that responds preferentially for images of faces in human subjects and if typically localized on the inferior occipital gyrus (though see Tsao, Moeller and Freiwald, 2008). Secondly, there were few lesions that overlapped the anatomical region commonly identified as the Fusiform Face Area (FFA, a second region of cortex typically localized on the fusiform gyrus in human subjects using fMRI, discussed more below). The authors contend that this second result may be because of a bias in their sample, as they were looking for patients that

exhibited prosopagnosia and/or achromatopsia, which usually results from more superior lesions (i.e. field of views that typically do not include the ventral fusiform gyrus). In summary, Prosopagnosia patients can demonstrate heterogeneity in both the behavioral deficits with visual representations of faces and in the anatomical location of the lesions responsible for their injury. This brief review of Prosopagnosia seems to suggest that rather than a single localized insult being responsible for "face blindness", prosopagnosia is a name for a diverse set of symptoms, and may have its etiology in damage across a larger network of areas. As will be reviewed next, functional MRI in human and monkeys have also revealed a number of areas in the ventral visual pathway that are activated when the subject views images of faces rather than non-face objects. These results suggest that rather than a single cortical locus being responsible for our visual knowledge about faces, there are a small number areas forming a face network.

1.2 Functional MRI of category selectivity in humans

Early studies in human subjects using whole brain imaging techniques have demonstrated that a number of higher level cortical regions of the ventral stream could be responsive to images of faces depending on the task and contrast used to reveal the activation. For example, in an early PET study, voxels that had a greater response when subjects had to indicate the identity of a face image as opposed to its gender, were found in three areas: the mid fusiform gyrus, the temporal pole, and the medial temporal lobe (Sergent, Ohta and MacDonald, 1992). Similarly, regions that demonstrated an increase in activation when subjects had to match faces across views, in contrast to matching spatial position across rotations, included a number of cortical regions including the posterior, middle, and anterior fusiform gyrus, orbital frontal and prefrontal cortex in PET and fMRI (Clark et al., 1996; Haxby et al., 1994). With fMRI,

regions on the mid fusiform gyrus, superior temporal gyrus, and the occipital temporal sulcus were found to be active when subjects passively viewed intact versus scrambled faces (Puce, Allison, Gore and McCarthy, 1995). While these studies demonstrate that the exact spatial location of activity that is found in relation to images of faces can vary depending on the task used, they also suggested that activation in the mid fusiform gyrus might be robust across tasks. These early imaging studies were motivated by a range of phenomena. Searget and colleagues were interested in the spatial ambiguity to assigning a cortical locus for the deficits observed in prosopagnosia patients due to the spatial variability of neurological damage across many patients (Damasio, Tranel and Damasio, 1990; Sergent et al., 1992; Sergent and Signoret, 1992). Haxby and colleagues (Clark et al., 1996; Haxby et al., 1994; 1991) were motivated to investigate possible human homologies for the broadly defined functional streams postulated in macaques (Ungerleider and Mishkin, 1983). Finally, McCarthy and colleagues (McCarthy, Puce, Gore and Allison, 1997; Puce et al., 1995) were interested in the spatial extent of selectivity for faces and objects discovered in their electrophysiological studies of the temporal lobe in humans. Critically, these studies approached face processing in the context of an explicit task. This component would differ in future studies which would focus on the definition and spatial location of face selectivity per se.

Later studies using whole brain imaging methods to examine the neural basis of face processing in humans subjects built upon earlier findings by refining the task conditions used to probe areas that responded to images of faces. As opposed to devising tasks that would elicit activity from areas performing key aspects of face behavior, new studies sought to establish a spatial location, reproducible across subjects, for face selective regions. Areas where the functionally measured signal responded more when the subject viewed images of faces than when viewing images of non-face distractors. Previous studies had suggested that the mid fusiform gyrus might be responsive over a

wide range of tasks. When directly tested, this area was found to be responsive in a majority of subjects and consistently gave a stronger response to image of faces over non-face distractors, such as hands or intact whole objects, on a number of different tasks (Kanwisher, McDermott and Chun, 1997). This area was designated as the fusiform face area (FFA), and became one of a number of category selective areas that would be identified using fMRI (Epstein and Kanwisher 1998; Downing, Jiang, Shuman and Kanwisher 2001). Moving away from previous efforts to establish correspondence between behavioral tasks and localized brain activity, later studies have similarly sought to establish other category selective areas. The Occipital Face Area, or OFA (Gauthier, Skudlarski, Gore and Anderson, 2000; Halgren et al., 1999) is a functionally defined cortical regions typically localized on the inferior ventral occipital-temporal surface, posterior to the FFA. While the exact function of the OFA is a matter of active research, several lines of evidence suggest that the OFA is hierarchically earlier in the visual processing stream when compared to the FFA.

The OFA appears to be sensitive to the presence of face parts in an image, but not the arrangement of the parts, nor the identity of the face image. Rotshtein and colleagues (2005) found that the FFA, but not the OFA, was sensitive to a change in identity perceived across two images, but not when the same amount of physical change between the two images subjectively maintained the identity across the two images. The OFA, on the other hand, responded significantly whenever there were physical changes across the two images. Similarly, the FFA and not the OFA was sensitive to changes in the spatial position of the eyes, nose, and mouth, though both areas were sensitive to the presence of these face parts (Liu, Harris and Kanwisher, 2010). Further evidence that the OFA is sensitive to the presence of face parts comes from a TMS study that demonstrated deficits in discriminating faces when a face part had been changed, but not when the spacing between face parts had changed with stimulation of the OFA

(Pitcher et al 2007). This study also demonstrated that the deficit occurs 60-100ms post-stimulus onset, suggesting a temporally early role for the OFA in face processing.

A third commonly described face selective area is in the Superior Temporal Sulus (STS). The fSTS has been implicated in detecting eye gaze direction in images of faces, and is generally hypothesized to be important in representing the face's changeable aspects (Haxby, Hoffman and Gobbini, 2000). FMRI experiments that had subjects direct their attention to either the identity or the direction of gaze across images of faces found that the fSTS and not the FFA responded more to the eye gaze condition than to the identity condition. The exact opposite results were seen in the FFA (Hoffman and Haxby, 2000). Studies using dynamic stimuli (short 3s movie clips of face, bodies, places, and objects) have also found that dynamic face stimuli activated the fSTS to a significantly greater extent than static images of faces (Pitcher, Dilks, Saxe, Triantafyllou and Kanwisher, 2011). This effect was not true for dynamic images of other categories, nor in the FFA, suggesting that different category selective areas may be specialized for dynamic and static face stimuli.

The functional differences in the category selective areas typically localized in human subjects has prompted a general hypothesis for face processing within the ventral visual pathway. The face system hypothesis (Haxby et al., 2000) states that a structural (or shape based) representation of faces is built up from low level image features in the posterior regions of the ventral visual pathway, culminating in the OFA. The OFA then sends face specific information forward to the FFA for analysis on the invariant aspects of faces (i.e. identity) and to the fSTS for the changeable aspects of faces (i.e. Social information). While the face system hypothesis (FSH) places broad functional descriptions onto large areas of cortex, there are potentially an infinite number of ways that populations of neurons could produce signals that would be functionally

parsimonious with the descriptions formulated in the FSH (e,g, representing the structural, changeable, or identity based information from faces). In this sense, the FSH is not a model of how the visual system creates signals that can support human face behavior, but rather it is a model for the spatial organization of the key areas involved in representing faces in the ventral visual pathway. The model makes strong predictions about where the brain is creating key signals that can be used to support human behavior with images of faces. Given that the key claim of the FSH is the spatial localization of face selective neural activity to only a few discreet locations in the ventral visual pathway, it is not surprising that a number of recent studies have begun to re-examine the definition of the FSH framework by exposing new areas of cortical tissue that meet the same criteria for inclusion as face selective, and by exposing the regularities in the variance between subjects in localizing the standard FSH areas.

A number of neuroimaging studies have observed additional patches of cortical tissue that putatively meet the same criteria for face selectivity as the areas originally described in the FSH; a cluster of voxels that responds more to images of faces than to images of non-face distractors on average (Kriegeskorte, Formisano, Sorger and Goebel, 2007; Rajimehr, Young and Tootell, 2009; Tsao et al., 2008, Pinsk et al 2009). Previous work had also observed additional clusters of voxels that responded to images of faces. These observations were generally considered individual variation because they were not reproducible across subjects, or the activation could have been related to an underlying mental function related to the task (i.e. memory, attention, feature extraction, subordinate level object recognition) rather than selective for faces per se (Kanwisher et al., 1997; Puce et al., 1995). In the past, activation of a region of cortex on the fusiform gyrus had been observed across a number of studies utilizing different behavioral tasks with images of faces (described earlier). These observation suggesting an implicit type of "behavioral reproducibility" criteria in the definition of face

selective; a criteria that has eroded in recent studies examining category selectivity for classes of behaviorally relevant categories. While one of these additional face patches has been linked to the anatomical organization of face patches observed in monkeys (Rajimehr et al., 2009; Tsao et al., 2008), the number, location, and organization of monkey face patches in still a topic under active investigation (Pinsk et al., 2009). Advances in neuroimaging technology could, in principle, account for the visualization of previously unobserved areas selective for images of faces. Higher field strengths, novel pulse sequences, and new hardware have allowed researchers to sample activity at higher resolution and the ability to image areas (like the anterior regions of the temporal lobe) that produced unreliable signal due to biophysical limitations in the past. In addition to a proliferation in the number of cortical areas (as defined by functional imaging methods) that are dedicated to face processing, reevaluation of the FSH has also come from trying to make sense of the individual spatial variance observed in the number and location of face selective patches.

Previous studies have generally assigned a cluster of face selective voxels localized in an fMRI scan to one of the standard FSH patches, usually without regard to the fine scale neuroanatomical location of the activation. As Weiner and colleagues have argued (Weiner and Grill-Spector, 2010), the activation of additional face patches or large scale anatomical variation in the observed location of standard face patches (even between subjects in the same study) are often described as individual variation across subjects. This results in a large degree of inter-subject spatial variability in the localization of category selective areas. Weiner and colleagues (2011; 2012) have used multiple anatomical criteria (gyrus/sulcus location) and functionally defined landmarks (retinotopic boarders, the location of hMT+), in combination with functionally selective activations to more precisely localize functionally defined areas. They observed spatially reliable alternating face and body selective regions extending from the anterior

fusiform gyrus (similar to the anterior FFA reported in the studies listed above) through the middle fusiform gyrus (i.e. the approximate location usually associated with the FFA) and extending to the lateral surface around the hMT+ complex (including the putative OFA). These results were reliable in their group of subjects, even after three years.

In summary, the use of fMRI to study the neural basis of face processing in humans has resulted in an (expanding?) network of cortical areas (e.g. face patches) that are hypothesized to be exclusively responsible for face related visual behavior in the brain. Similar to the review of prosopagnosia, what seemingly began as a single putative "face region" (e.g. the area damaged in prosopagnosia or the FFA) has become a rich complex network of areas responsible for a broadly defined narrative of operations (i.e. detection, identity, social recognition) on images of faces. Resolving how (or if) to incorporate these additional areas into the FSH view may require a new narrative of operations. The idea of a limited set of domain specific regions in the brain responsible for face behavior, however, has received renewed interest due to the spatial localization of similar face selective patches in the macaque using fMRI. The next section of this chapter will review these recent studies in non-human primates.

1.3 Functional MRI and face selective patches in monkeys

Functional imaging in the awake non-human primate has been used to reveal large scale spatial structure for behaviorally important categories of images. As with human subjects large scale organization has been observed for faces (Bell, Hadj-Bouziane, Frihauf, Tootell and Ungerleider, 2009; Logothetis, Guggenberger, Peled and Pauls, 1999; Pinsk, DeSimone, Moore, Gross and Kastner, 2005; Tsao, Freiwald, Knutsen, Mandeville, and Tootell, 2003), Bodies (Bell et al., 2009; Pinsk et al., 2009; Tsao et al.,

2003) and Places (Bell et al., 2009; Nasr et al., 2011; Rajimehr et al., 2009; Rajimehr, Devaney, Bilenko, Young and Tootell, 2011). For images of faces, up to six category selective clusters have been observed with fMRI along the ventral visual pathway in each hemisphere, though the exact number and anatomical location can vary between subjects (Pinsk et al., 2009), as well as between studies.[1] While the patches identified with fMRI are defined by a greater response when the subject views images of faces over non-face distractors, the different patches have been shown to vary in the degree of face selectivity observed (Tsao et al., 2008). The posterior patches being less selective (i.e. the magnitude of the fMRI defined face selectivity metric) than the more anterior patches, conforming with hierarchically based theories of object recognition in the temporal lobe (Fukushima, 1980; Miyake and Fukushima, 1984; Riesenhuber and Poggio, 1999; Serre et al., 2007a). Of the six putative patches selective for faces identified by Tsao and colleagues (2008), three are localized along the convexity of the STS, while the other three are localized medially, either along the fundus of the STS, or anterior of the Anterior Medial Temporal Sulcus (AMTS). Though there is currently no generally accepted naming scheme for these six patches (partially because not all groups observe six patches), the naming scheme designated by Tsao and colleagues (Tsao et al., 2008) will be used in this thesis. According to that scheme there are three patches along the convexity of the STS (identified generally as "lateral"), named according to their posterior to anterior position: Posterior Lateral (PL), Medial Lateral (ML) and Anterior Lateral (AL). A similar naming scheme is used for the medial patches; from posterior to anterior, Medial Fundus (MF), Anterior Fundus (AF), and Anterior Medial (AM). Anatomically, some of the patches can be related to previous divisions of the temporal lobe based on anatomical methods. The patches that run along the convexity of the STS anatomically intersect the three anterior to posterior

[1] [1] It is worth noting that more than six patches have also been observed (Tsao et al 2009; Moeller et al 2008). At this time these supplementary patches have been regulated to the concept of individual variation.

divisions of IT by Felleman and VanEssen (1991) with PL, ML, and AL located in areas PIT, CIT and AIT respectively. It is not currently 100% clear at this point what the exact pattern of connectivity is for the neurons localized to the fMRI identified patches. As described in the next section, recent studies using simultaneous fMRI and microstimulation targeted to sites in the six face selective patches of the macaque suggest that these neurons have direct connectivity with each other (Moeller, Freiwald and Tsao, 2008).

Simultaneous micro-stimulation and fRMI in awake behaving monkeys has provided some evidence that the fMRI identified face patches form a network of areas. Moeller and colleagues (2008) monitored the face selective areas with functional MRI while locations in and out of the face patches received micro-stimulation. They observed that the activation induced by stimulation into a face selective patch included large portions of other face selective patches in addition to activity that spread around the stimulation site. They also observed some activity at locations not strictly localized to an fMRI identified patch (e.g. induced activity at other patches typically spread out beyond the boundary of the border of the patch identified with fMRI, see also Figure S3). Stimulation in a control areas outside of a patch, on the other hand, induced fMRI activity around the stimulation site in addition to a number of discreet locations outside of any of the face patches. The study used this technique to probe the connectivity of patches ML, AL and AM. Stimulation in the cortex localized to ML produced activation in all of the other category selective areas, at least to some extent across the subject population. While no one subject activated all of the other patches in response to stimulation at ML, stimulation in ML across all subjects resulted in the activation of all the other patches. Similarly, micro-stimulation localized to AL and AM produced activation across the subject pool to the other patches except PL. One of the major

conclusions of this work was the hypothesis that the fMRI identified face patches formed a domain specific network for faces.

Though these results are provocative, the extent to which only these areas are active in face processing is not clear from the data. As noted previously, some of the induced activity did not strictly fall upon cortex that was identified with fMRI as category selective. It is certainly conceivable that the discrepancy in most cases could extend from experimental error in registration between functional and anatomical images or in the localization of the electrode tip. A more difficult problem, however, extends from the uncertainly involved in relating how the selectivity of the fMRI signal is related to the spiking response of the neurons in the tissue localized to the fMRI signal. In other words the fMRI patch may be spatially larger than the cluster of neurons assumed to be responsible for the fMRI signal (or it could be smaller etc.). It is also not clear that, if there were in fact neurons outside the putative face patch system (e.g. fMRI activated areas) that did in fact contribute to face processing, whether they would show up at all in this experiment, as it is not currently known how many putative face neurons would need to be clustered together in order to show activation in fMRI (some data collected in this thesis may speak to this issue in chapter 2). Finally, given that at least some of the face patches can be localized to an established system of cortical regions (PIT, CIT and AIT), it is not clear in what sense the pattern of activation might be considered a unique cortical circuit. It is also not known what exact spatial subregions of IT constitute the default feed forward circuit for object processing in the temporal lobe. It is possible that at least some of the face activations are contiguous with sub-regions in the standard hierarchal feed forward scheme involved in all object recognition. In such a case, rather than a unique domain specific network, the patch activations could be explained as biases in spatial representation extending from lower level features or spatial biases in cell clusters extending from learned associations in experience. While the face selective

activations in IT likely contribute to face recognition behavior, the extent to which they should be considered a face *exclusive* network is open for debate given the paucity of data available on this topic. In conjunction with the proposed homologies between humans and macaques, the face network hypothesis has also prompted a qualitative classes of models to describe their function in a manner similar to humans.

While there has not been a quantitative attempt to functionally corroborate the fMRI defined regions between the macaque and human (but see Tsao et al., 2008), it is widely speculated that the macaque middle face patch (MFP), or area ML, is homologous to the human FFA (but see Ku, Tolias, Logothetis and Goense, 2011). Evidence for this hypothesis stems from its relative position to the putative PPA in both species (Nasr et al., 2011; Rajimehr et al., 2009), as well as the location of the MFP relative to well defined anatomical landmarks across both species (Tsao et al., 2003). Additionally, an area of cortex corresponding to the MFP (i.e. in the same qualitative region of the STS) has been localized in at least one hemisphere in all of the functional neuroimaging studies to date that have sought to localize areas selective for images of faces over non-face distractors (Issa and DiCarlo, 2012; Bell:2009fm; Pinsk et al., 2005; 2009; Tsao et al., 2003; 2008). Given the proposed homology, putative position in the face patch network, and the report that nearly all of the neurons in the MFP were observed to be selective for faces (an observation that will be discussed shortly), a qualitative model with two major claims has been proposed for the face patch system (Moeller et al., 2008; Tsao and Livingstone, 2008). The first major claim is that the face network is gated by a domain specific filter. While it is not entirely clear what this restriction would mean, we can minimally operationalize a weak form of this as requiring the neural response from neurons in the face patches to correlate with face detection behavior. This is seen as a crucial element for this class of models for at least two reasons (Tsao and Livingstone, 2008). First, face detection acts to limit the domain of possible inputs to the network,

reducing the computational load on the system. Face detection would reduce the domain of the inputs to later upstream areas, effectively constraining the solution space for the computation of higher order features. Secondly, a face patch system limits the spatial location in the brain for face processing. Specifically, activity in the normal healthy brain that is related to face behavior is hypothesized to occur exclusively within the network of face patches. Similar to the FSH in humans, the detection of structural feature information for faces is hypothesized to occur in the most posterior face patch, sending neural activity related to the detected faces forward to the MFP (Tsao and Livingstone, 2008). From the MFP, structural information about faces is sent forward to the anterior patches for representing identity and social information. This class of models makes two major experimentally testable predictions: First, neural activity in the MFP should be well correlated with face detection behavior, and secondly all of the neurons localized to the MFP should have spiking responses that reflect information processing for images of faces. While several recent studies have targeted the MFP in neurophysiological experiments, neurons that respond primarily to images of faces have been observed since the earliest investigations that targeted neurons in the temporal lobe, as reviewed in the next section.

In summary, functional MRI studies in the awake behaving macaque have identified several large scale regions of cortex that responded preferentially when the monkey viewed images of faces than when viewing images of non-face objects. While the exact number and location of these patches has been variable across studies from different laboratories, this thesis focuses on the middle face patch. The MFP has been localized on the convexity of the STS in area CIT of the macaque temporal lobe across multiple studies. Due to these factors the MFP is hypothesized to be homologous to the human FFA.

1.4 Neurophysiological studies of face cells in the temporal lobe of monkeys.

Neurons selective in their response to visual images of faces were first observed by Gross and colleagues (Gross, Rocha-Miranda and Bender, 1972), who simply noted that some cells could only be driven by complex images, such as photographs of faces or trees. A later study from the same laboratory reported on the activity of a number of neurons in the upper bank of the STS in area TE, or the superior temporal polysensory area (STP) that seemed to only respond to faces, whether they were images of human or monkey faces, photographs, drawings, and actual faces (Bruce, Desimone and Gross, 1981). These neurons responded vigorously to intact faces, less so if the face was disrupted by covering the eyes in the image, and not at all if the face was completely disrupted by scrambling the image. These neurons were recorded in an area that was not exclusively visual, and the latency for responding was quite long (200-300ms). Cells with similar response properties were discovered in a number of additional locations, including the fundus (Perrett, Rolls and Caan, 1982), ventral surface, and lower bank of the STS (Desimone, Albright, Gross and Bruce, 1984). Similar to the selectivity reported for neurons in the upper bank of the STS, these neurons responded robustly to any image of a face (monkey, human, photographs, pictures, or filtered images). Unlike the visual response of neurons in posterior visual areas, the exact position and size of the face did not seem to be crucial in the response of the neurons (indicating that they exhibited some tolerance). These neurons also seemed to be category selective in that they responded only weakly or not at all to any of the other complex images or grating tested. Interestingly, Desimone and colleagues reported that face parts in isolation always elicited responses reduced in magnitude, while Perret and colleagues observed a few examples (n=48) where cells responded equally as well to an isolated image part as to the whole face (such as an eye or mouth; see figures 9 and 10). This result suggested

that these these neurons could respond maximally to a complex object that was not a whole intact face. Both groups reported that while the neurons in their samples were sensitive to 3D rotations of the head, they were largely insensitive to in-plane rotations of the head. These early reports exposed some of the basic features for what would later become termed "face cells:" they were category selective for images of faces; responding robustly to the images of faces that were used in these studies and weakly to the non-face images examined. These cells also exhibited view tuning, responding selectivity to frontal or profile views of the face.

While initial studies held a more qualitative view of these neurons, later studies would begin to quantitatively differentiate neurons with response preferences for faces. *Face selective cells*, were first defined as cells that responded more to images of faces than non-face images, though some studies had more explicit criteria. For example, face cells were cells that responded at least twice as much in magnitude to the optimal face image in the stimulus set than to any other non-face stimulus tested (Rolls and Baylis, 1986; Rolls, Baylis and Leonard, 1985). This criteria was used as a screen to find "face cells" in order characterize the response properties of this (presumably homogenous) class of neurons. Other studies would come to elaborate on the basic findings of those first studies, creating a collective characterization for the properties of face neurons in IT. Early reports often reported on a limited set of face exemplars, and therefore either did not elaborate on the variability between different face images (Bruce et al., 1981; Desimone et al., 1984), or reported anecdotal accounts (Perrett et al., 1984). Later reports would emphasize the wide range of responses observed to different face identities. Baylis, Rolls and Leonard (1985) found that the discriminability between the best and worst face could range over a d' of 0.5-8, suggesting that putative face neurons might not be encoding semantic or category membership per se, as opposed to an unknown shape based feature dimension common in images of faces. In order to

examine this hypothesis, Yamane and colleagues modeled the response of single face neurons with a simple linear model that weighted the distance between the major features of the face (Yamane, Kaji and Kawano, 1988; Yamane, Komatsu, Kaji and Kawano, 1990). While they arbitrarily chose which facial features to measure, these models produced correlations with their data on the order 0.6-0.8, and predicted the response of the neuron better than simple contrast or color models. In their study, the weights were often non-zero on only a few of the features, suggesting that each cell keyed in on a subset of all the possible features. While intuitive and provocative, it is not clear if such a set of features could distinctly represent a realistic number of exemplars in a population of these neurons. Another issue with this study is that these correlations might be overestimates of their model's true performance as there was no attempt to cross-validate the model performance. Nevertheless, this study illustrated that putative face neurons are sensitive to the structural features that make up a face image, and not "whole face" detectors. Different neurons in their study often had weights for particular features which coincides with the observation that a single feature (for example, an eye or mouth) could drive a neuron as well as the whole face (Perrett et al., 1982; 1984). This intuition has led to the general idea that specific configurations of facial features are represented across a distributed population of face cells. Computational studies that have directly attempted to investigate the representational capacity of face neurons found that individual neurons carry a distributed code for face identity that ranged from 0.3-0.5 bits/neuron (Abbott, Rolls and Tovee, 1996). This study used neurons recorded across neural populations that would today be considered unique.

Another major characteristic that many studies would elaborate on is the view tuning demonstrated by face selective cells. Initial observations suggested that "face" neurons were tuned to a few characteristic viewpoints: frontal and profile views (Desimone et

al., 1984; Perrett et al., 1982; Perrett & Harries, 1988; Perrett et al., 1985). Later studies with larger sample sizes, however, found that a majority face cells responded maximally to a single view of the head. More importantly, these neurons could be tuned to a wide range of specific views (Hasselmo, Rolls, Baylis and Nalwa, 1989b), though there was some bias for cells to be tuned to views of the head face forward, left and right profile, and the back of the head (Perrett et al., 1991). This study also found that a small proportion of their sample was bimodally tuned to isomorphic views of the head (i.e. left and right profile). The width of tuning (the amount of rotation in order for the cell to respond at 1/2 of the maximal response) for a majority of cells was <60 degrees, though a small proportion of the cells had broad tuning widths (>90 degrees). Face neurons have also been observed to be selective for 3D rotation in the vertical plane (e.g. head looking up or down) in addition to the direction of the eye gaze (Perrett et al., 1985). Eye gaze sensitivity was reported in a minority of face neurons tuned for view, and the effect was graded and often additive to view tuning. For example, a cell might be tuned to the frontal view of a face, and further respond maximally when the eyes in the image are directed forward, gradually reducing the response as the gaze is averted, though still responding more than if the head was rotated to the profile view. Similar view and gaze tuning has been reported for putative face neurons in the anterior STS (De Souza, Eifuku, Tamura, Nishijo and Ono, 2005; Eifuku, De Souza, Tamura, Nishijo and Ono, 2004).

Spatially, Perret and colleagues reported that penetrations which contained a cell tuned for one view of the head were 3-8x more likely to have other neurons tuned to the same view than from what would be expected from a random penetrations in IT (Perrett et al., 1984). The same study estimated the probability of finding a penetration that recorded a neuron whose response was selective to the same view of the head as a function of horizontal distance. The magnitude of the probability for distances less than

1mm was ~0.2, reliably greater than chance. This result suggested clustering by head pose along the surface of cortex out to 1-2mm. Similar results would be reported using a optical imaging methods (Wang, Tanaka and Tanifuji, 1996; Wang, Tanifuji and Tanaka, 1998). The region of cortex activated for five different poses of a doll head were found to be spatially adjacent and overlapping on the cortical surface in 4 animals, suggesting a spatial map for head pose. The activated spots averaged 0.58mm and 0.38mm in width (e.g. they were oval in shape, and smaller than the pose tuned clusters described by Perret), and the entire length of the overlapping regions was ~1.28mm. One novel hypothesis that stems from the observation of gaze tuning is that the view tuned head might contribute to tracking the direction that others are allocating their attention (Perrett, Hietanen, Oram, Benson, & Rolls, 1992).

Finally, a number of studies have argued that face cells exhibit tolerance and selectivity in their response to faces in a manner similar to what would be expected from the behavior of a human subject (Perrett, Mistlin and Chitty, 1987). Sensitivity to color and contrast inversion (similar to the behavioral effects observed with contrast reversal and mooney images in human subjects) have been reported for face cells (Perrett et al., 1984) but see (Rolls and Baylis, 1986). Face selective cells have also displayed some invariance to the effects of variable sources of illumination (Hietanen, Perrett, Oram, Benson and Dittrich, 1992) responding selectively to their preferred head pose under different illumination conditions (light from below, above, to the left or right). The 21 neurons analyzed in the study were sampled from a region likely near ML and MF. While single neurons did not typically maintain their response across all lighting conditions, the average response across all of the neurons did. Early reports implied that size and position have very little effect on the response to face cells (Desimone et al., 1984; Perrett et al., 1982; 1984). Studies that have explicitly examined the size and position tolerance of face selective cells in IT have been in general agreement. Cells

selective for faces that were recorded in either the STS or ventral surface of anterior IT were found to maintain their response out to 12 degrees from the fovea (Tovee, Rolls and Azzopardi, 1994). PCA analysis revealed that the majority of the variance was explained by sensitivity to identity, and that position information reached its maximum early (between 80-100ms post stimulus onset) while the amount of identity information peaked later in the response (100-220ms). Rolls and colleagues also examined the size tolerance exhibited by a population of face selective cells recorded on the ventral surface of anterior IT and in the anterior fundus of the STS (Rolls and Baylis, 1986). On average these neurons were reported to tolerate ~3 octaves of size change before the response to the optimal stimuli was reduced by half its maximal amount.

Large scale organization for face selective cells was rarely discussed in these early studies with a few notable exceptions. Perret and colleagues have suggested both clustering of cells tuned for particular views of the head and face on the order 1-2 mm in the thickness of the cortical ribbon, and for a larger "patchy" organization for faces in the upper bank of the STS extending 3-6mm in size (Harries and Perrett, 1991; Perrett et al., 1984), where the density of putative face neurons ranged from 20-80% (Perrett et al., 1988). Harries and colleagues (1991) argued for the existence of periodic patches of face selective cells in the upper bank of the STS, each patch extending 3-6 mm in the anterior to posterior direction. Extensive neurophysiological sampling was performed in four animals targeted to the fundus and upperbank of the STS. Recordings were made along a 15mm extent from posterior to anterior. They found evidence for at least one patch, and argued for two patches separated by a 3mm zone where no face neurons were observed. This was done by pooling sites spatially into 3mm strips along the anterior-posterior extent and looking at the proportion of penetrations that contained at least one face selective cell versus the proportion that would be expected by chance. Retrograde tracers placed in the Inferior Parietal Sulcus (IPS) produced revealed

periodic patterns of label in the upper bank of the STS. Spatial Fourier analysis conducted on the labeled tissue in IT cortex revealed that the spatial location and number of labeled regions was similar to the frequency of face clusters estimated from the neurophysiology data, suggesting that the face patches in the STS feed forward to dorsal regions along IPS.

Another spatial organization that was becoming evident from a number of studies was a divisions between face selective populations in anterior STS and the ventral surface of anterior IT. A report that examined the identity of face images versus the emotional expression displayed by the face in the same image found that neurons in the anterior STS were more sensitive to expression while neurons on the ventral surface of anterior IT were more sensitive to the identity of the image regardless of the expression (Hasselmo, Rolls and Baylis, 1989a). Similarly, Young and Yamane (Young and Yamane, 1992; Yamane et al., 1988) examined the encoding of a population of face neurons recorded from either anterior STS or the ventral surface of anterior IT. Using either a structural or familiarity model, they found evidence that anterior IT was better fit to the structural model, while anterior STS was better correlated with the familiarity model. A similar finding for anterior IT was found in a more recent study that examined identity and pose in neurons localized to either anterior IT or anterior STS (Eifuku et al., 2004). Using a multidimensional scaling procedure, this study reported that anterior IT neurons were more selective for the identity (disregarding pose information), while the opposite was true for the anterior STS neurons.

In summary, studies of face selectivity in IT have described a number of properties prior to the use of fMRI to describe face selective cells in the macaque. Putative face cells were found to have a few major characteristics. First, face cells were predominately view tuned, and largely tuned to one, or two isomorphic views of the head and face.

Second, face cells generally exhibited some tolerance to changes in size, position, contrast, color, and lighting. Third, many face neurons respond variably to different face images, suggesting a distributed feature based encoding by individual cells. Fourth, there was some evidence for structure at two scales in the spatial organization for view tuned cells. There was evidence for small local organization for view (perhaps column like, 1-2 mm), and a periodic "patchy" organization (3-6 mm) in the upper bank of the STS. Finally, there was some evidence for specialization of function by discreet populations of face selective cells. Anterior IT, for example, was implicated by many studies as selective for visually based identity. While these studies have laid the groundwork for understanding how the brain processes images of faces, the use of fMRI in the macaque has added spatial localization as an important variable in understanding the cortical mechanisms of face processing. Modern studies have begin to consider the large scale spatial organization in guiding neurophysiological examination of face processing in the ventral visual pathway.

1.5 Neurophysiology of face cells in the macaque middle face patch

Neurophysiology experiments that have explicitly sampled the spiking response of neurons localized to the fMRI identified face patches have begun to reinterpret previous studies in relation to the striking organization visualized with fMRI. Given the speculated homology between the MFP and the FFA, initial studies have focused on characterizing the activity of neurons localized to the macaque middle face patch. Neurophysiology targeted to areas ML and MF has revealed that these neurons are well driven with stimuli depicting the frontal views of the head (face), an interesting observation given the importance of head pose in previous reports. Spatially, earlier studies had found evidence for clustering of putative face cells estimated to be 3-6mm in anterior to posterior extent in the upper bank of the STS (Harries and Perrett, 1991;

Perrett et al., 1984). These clusters were larger than the "face columns" suggested by optical imaging studies (Tanaka, 1996; 2003; Wang et al., 1998). One important difference between previous and recent descriptions of the spatial organization of face selective cells is the idea of *purity*, or the proportion of face neurons in a face patch. This idea has emerged from recent neurophysiological studies of the fMRI identified face patches, where nearly every visually driven neuron (~97%; or 90% if only excitatory) encountered in the experiment demonstrated at least a 2x face selectivity; meaning that nearly every neuron sampled responded at least twice as much on average to images of faces than to other non-face distractors (Tsao, Freiwald, Tootel and Livingstone, 2006). Previous estimates of clustering considered the area of face selective units to be a type of enrichment, because the probability of finding other view tuned cells was not considered to be 100%. This is in contrast to the cortical module idea hypothesized in recent studies because these studies do hypothesize that the face patches have a near a 100% purity. Further these results were used to suggest that similarly localized areas in human subjects would be purely composed of face selective cells as well, providing the basis for considering the MFP as a distinct cortical area dedicated to processing images of faces (Tsao et al., 2006; Tsao and Livingstone, 2008).

Later experiments that have sampled from MF or ML have also reported a high purity for the number of face selective cells in the MFP, though the results have been mixed. Further studies from Tsao, Freiwald and colleagues have reported the purity of face selective cells in the MFP to range from 84-90% (Freiwald and Tsao, 2010; Freiwald, Tsao and Livingstone, 2009; Ohayon, Freiwald and Tsao, 2012). Similar studies carried out by Bell and colleagues (Bell et al., 2011) observed that only ~41% of neurons in an fMRI identified face selective patch located +5-6mm AP (likely area MF from Tsao and colleagues), were face selective. While this estimate of the purity is much lower than that obtained by previous studies, it might be parsimonious with the previously

described "patchy organization" reported by Perret and colleagues. The discrepancy between the two estimates may come from a variety of sources including differences in targeting, differences in fMRI localization, or even differences in sampling methods. Bell and colleagues argue that previous reports of high purity in the middle patch were due to biases in sampling, because earlier reports recorded from only a few spatial locations which may have restricted their samples to a limited region of the fMRI patch. Their methods provided a more unbiased estimate of the purity by sampling widely from the region of cortex estimated by fMRI to be category selective for faces. It should be noted that nether group could broadly sample the area localized by fMRI because neither group tried to estimate the spatial extent of the patch using the same selectivity measure used to estimate the purity. Differences between these estimates could also result from a variety of sources including methodology and registration error. Similarly, there could be a genuine physical difference between the selectivity localized to a given cortical location by fMRI and the true selectivity derived from the spiking responses of neurons in the tissue itself. Direct estimation of the spatial extent of the middle patch and its purity from physiological samples could resolve this issue, and was one of the major goals of this thesis work.

Studying the spiking responses of neurons in the middle face patch has been an area of active research. MFP neurons have been identified as sensitive to the semantically defined feature elements of the face, in addition to the contrast relationships between large areas of the face, and the 3D pose of the head. Previous studies of face selective cells have similarly suggested that some face selective neurons could be driven as well by a single facial feature (such as the eye) as it could by the whole face image (Perrett et al., 1982; 1984), and that the response of putative face cells were sensitive to the presence of one or more semantically defined face elements (Yamane et al., 1988; 1990). The psychological literature had also observed that face perception can be critically

dependent of the spatial relationship between parts of the face (Tanaka and Farah, 1993). Direct examination of the spiking response of neurons in the MFP has provided evidence that (1) MFP cells have a ramp like tuning to semantically defined facial features, (2) are typically sensitive to three feature elements, and (3) their response can be modeled as a linear summation of the feature sensitivities (Freiwald et al., 2009). Using cartoon images of faces with well defined facial features (cartoon stimuli elicited a response 84% of the maximal response to images of genuine faces), Freiwald and colleagues (2009) found that putative face neurons in the MFP could be sensitive to the presence of up to four face features and their interaction. Interestingly, the sum of the response by these neurons to the individual parts was typically greater than the response to the whole face (fig 2C), suggesting that inhibition played a key role in the response of single neurons to whole face images. The study also manipulated the parameters of the facial features across a range of values (i.e. one feature was the size of the pupils, which went from very small to very large) independently for each neuron studied. Neurons were typically sensitive to three feature elements, and their response over the parameter range was typically ramp shaped, meaning most neurons responded maximally (and minimally) to the extremes of the feature dimension tested, with a smooth response between the extremes. One conclusion of the study was that face detection occurs in the MFP because these neurons can respond to the presence of individual, or combinations of different, face parts in the visual stimuli. It was hypothesized therefore, that face detection can take a type of short cut, or heuristic, that only looks for the combination of a few prototypical face parts to determine if something the viewer would identify as a face is present in the visual input. This solution, however, switches the essential problem of face detection from detecting the presence of whole faces to detecting the presence of face parts (and also makes predictions about what sorts of features should confuse human face detection behavior). As discussed earlier, a major problem in implementing face detection lays in

understanding how the response of MFP neurons signal the presence of a face (or parts of a face) under *real world image variation*. Face detection algorithms (i.e. hypothesis for face detection) on still image scenes, under limited viewing conditions, have been shown to perform robustly in this domain using brute force search methods and relatively simple low level feature elements (Viola and Jones, 2001). However, even these, state-of-the-art systems fail under conditions where the face in the images varies in pose by ±15°, parts of the face are occluded, or with illumination variation across the image (Viola & Jones, 2004), all of which are viewing conditions that can be trivial for human observers. Therefore, it is not clear if the response properties described for MFP cells (e.g. that the cells respond to the presence of 1+ face parts under the viewing conditions examined in the Freiwald et al 2009 experiment) are sufficient to explain face detection behavior under the normal range of image variation encountered on a daily basis by living organisms. This point may emphasis the need to operationalize the idea of face detection in a different way.

Another key finding from previous studies was that a majority of face neurons were view tuned (De Souza et al., 2005; Desimone et al., 1984; Eifuku et al., 2004; Hasselmo et al., 1989b; Perrett et al., 1985; 1991; 1992; Perrett and Harries, 1988). This seemed to be true across populations of neurons localized to many different regions of the temporal lobe. Sensitivity to the 3D pose of the head has also been observed in the response of MFP cells (Freiwald and Tsao, 2010). This study investigated the response of neurons localized to the MFP (i.e. MF and ML), and areas AL and AM to 8 views of the head (face forward, 45° and 90° isomorphic views, in addition to tilt up and down), across 25 different individuals. In the MFP, neurons typically responded over a few of the views from the profile to profile rotation, though they responded best to one pose. In contrast, AL cells typically responded best to the isomorphic views of the head with 45% of the sample preferring a single identity, and in AM 73% of the cells tended to respond in an

invariant manner across multiple views of the head to the same identity. While previous reports had identified face cells that respond to isomorphic views of the head (Perrett et al., 1991), this report differs in 2 key ways. Earlier studies reported sampling in the upper bank and fundus of the STS and those reports did not note the high dependence of identity in cells tuned to isomorphic views. While the tendency to encounter identity selective units was observed over multiple studies in the anterior part of IT, it would appear that 2 different populations of face cells may have been responsible for previous reports.

Another feature model that has been used to test MFP neurons comes from computational studies of face detection The ratio template model for face detection takes advantage of the observation that forward posed images of faces tend to hold specific contrast relationships between different regions of the face across different illumination sources (Sinha 2002). For example, the ratio of the intensity stemming from regions that included the eyes and regions that included the forehead area is stable across illumination sources emanating from both above and below the head. Though face cells have been observed to shows tolerance to image variance from lighting source directions (Hietanen et al., 1992), it is not clear from this previous study how these cells could accomplish tolerance to light variation or how this property changes across different areas of the temporal lobe. Ohayon and colleagues specifically examined how well the Sinha model could account for the response of MFP neurons. Using random combinations of contrast across 11 regions of the face, (essentially outlining semantically defined regions of the face) this study examined the response to neurons localized in the MFP and found that ~50% of the cells in their sample were modulated by the contrast parameters. Given that a number of the contrast stimuli outlined key semantic features of the face image, and previous studies had shown that MFP cells can respond additively to the combination of face parts, the range of responses observed to these

stimuli was comparable to previous reports. Importantly, when tested with natural images that had the contrast relations predicted by the model, the response of MFP cells to genuine faces increased with the number of contrast relations, while MFP cells were unmodulated by the non-face images that contained the same number of contrasts relations. This suggests that the semantic designation of the image was still the dominant factor in the response of the cells. These results indicate that the ratio-template model is insufficient to account for the response of MFP cells.

In summary, recent reports that have specifically targeted the response of neurons in the fMRI identified face patches have corroborated a number of earlier reports on the properties of face selective units in the temporal lobe and, importantly, begun to relate these findings to the cortical areas localized with fMRI. Studies that have targeted neurons localized to the MFP, have found that these cells exhibit sensitivity to facial view in addition to the presence and number of semantically defined physical features on the face. Finally, a number of studies have reported a high purity of face selective cells localized to the MFP (but see (Bell et al., 2011). For these reasons, it has been theorized that the MFP acts as a domain specific filter, detecting faces from the visual input through the presence of face parts and direct template matching (Freiwald et al., 2009; Moeller et al., 2008; Tsao et al., 2006; Tsao and Livingstone, 2008).

1.6 Thesis Overview

In this thesis we will explore the spatial structure of face selectivity and its relation to face detection behavior in the macaque middle face patch. Our goal in this work is two-fold: (1) to provide a detailed characterization of the spatial extent of the face selective signal in the cortical tissue localized by functional imaging, and (2) to characterize the

face detection abilities of sites in the MFP by grounding estimates of neural sensitivity to behavioral performance. It is also our hope that this work will provide useful practical information for the localization of the MFP.

In Chapter 2, we aim to spatially characterize the macaque middle face patch. In particular we want to estimate the size and purity of the fMRI identified MFP. To accomplish this we will examine the response of 100s of multiunit sites localized to the middle face patch (or area ML from Tsao, Freiwald and colleagues) using a novel X-ray imaging system. This is a meaningful contribution because the hypothesized (maximal) purity of the face patches have become a key theoretical property in hypothesis of how the ventral visual pathway organizes and encodes information about images of faces. Estimates of the purity of the MFP have played an important role in hypothesizing a closed system for face processing in the macaque brain (Moeller et al., 2008; Tsao et al., 2008), and is the basis for speculating the MFP's homology to the human FFA and status as a distinct cortical area (Tsao et al., 2006). While previous reports have used fMRI to guide sampling from face selective populations, the spatial resolution of these techniques has been limited and inconsistent. Furthermore, no attempt has been made to actually measure the true spatial extent that putative face selective cells occupy in the cortex, despite the fact that reliable estimates of the purity depend on the actual size of the region. The boundary of the fMRI defined face patch can be modulated by a variety different factors. Some of these include: the analysis parameters (i.e. threshold estimates), noise from a variety of sources, the use of contrast agents, and registration of the functional data to anatomical representations of the subjects brain. Estimating of the size and purity of the MFP from more direct measurements of cortical neural activity will resolve outstanding discrepancies in the literature on the purity of the MFP. We conclude from our data that the middle face patch is ~6mm in diameter (on inflated 2D coordinates) and has a purity that varies as a function of distance from the center of the

patch, that can range from a proportion as high as 96% in the center of the patch, to a baseline proportion as low as 3% outside of the patch.

A preliminary underlying assumption of pervious work is the presence of a compulsory and explicit face detection stage. In chapter three, we aim to explicitly examine face detection performance in the physiologically defined MFP. In order to examine this issue, we correlate the response of MFP neurons to the performance of human subjects performing an explicit face detection task. This work is a meaningful contribution because many current theories about how the fMRI identified face patches are organized to represent knowledge about faces explicitly, or implicitly, assume a face detection stage. While neurons in the MFP are defined based on their selective response to images of faces, the face images conventionally used in these experiments have operationalized only a small portion of the variability that genuine face images can have in daily human experience. We directly examine the response of MFP neurons to face detection behavior in human subjects, examining directly the role that MFP neurons play in face detection. We find that MFP neurons can be poorly correlated with face detection behavior in human subjects, as the median correlation between our sample of MFP neurons and human face detection behavior was $\rho \cong 0.18$. This thesis will conclude in chapter 4, with a brief review of the major findings from chapters 2 and 3, and a discussion of these results in the context of the background detailed in chapter 1. FInally, we will discuss the overall limitations of our work, and suggest future experiments that might overcome some of these limitations.

Chapter 2

Spatial structure in the macaque middle face patch

2.1 Introduction

This chapter will deal exclusively with our efforts to spatially characterize the category selective signal for faces observed with fMRI (Bell et al., 2009; Pinsk et al., 2009; Tsao et al., 2003) by mapping the spiking responses of multiunit sites in the cortical tissue of the macaque temporal lobe. The work described in this chapter was done in collaboration with Elias Issa, who collected the data from Monkey 2 (or M2, see methods). In this research, we were guided by two overarching goals. Firstly, we wanted to be able to describe the structural characteristics of the middle face patch (i.e. How big is the patch? How pure is it? Where is it?). Secondly, we wanted to describe the category selectivity of the entire MFP with a direct neural measure in relation to the tissue outside the MFP (i.e. How much more selective is the MFP than regular IT? How many more face selective sites are in the patch than out?). In order to answer these questions, we devised methods to spatially model the category selective sites sampled from the MFP. We looked to the human and monkey fMRI literature on face processing for inspiration to formulate first order spatial models for our data.

In human subjects, the fusiform face area (FFA) is believed to play a crucial role in face processing (Kanwisher et al., 1997; Kanwisher and Yovel, 2006). The FFA is a cluster of

voxels localized with fMRI, that responds greater on average when a subject views images of faces than when viewing non-face distractors. The FFA can also be coincident with anatomical insults resulting in prosopagnosia, or face blindness (Damasio et al., 1990; Sergent and Signoret, 1992). Anatomically, face preference in the FFA has been observed to fall off gradually within an extent of ~3 mm on the cortical surface (Spiridon, Fischl and Kanwisher, 2006). Hypotheses about the structure of the FFA places emphasis on the purity of the category selective response in the patch, arguing that the responses of putative neurons in the FFA should be domain specific to human face behavior (Kanwisher, 2000). An ongoing debate in the literature directly concerns the purity of category selective neurons in the FFA. Some studies have shown that voxels localized to the FFA contain weak, but reliable information about non-face objects, suggesting the presence of non-category selective neurons (Haxby et al., 2001). Counter theories suggest that the response to non-face images in the FFA reflects behaviorally irrelevant information stemming from weak activity to non-preferred responses along an unknown feature dimension (Spiridon and Kanwisher, 2002; Williams, Dang and Kanwisher, 2007).

Similar fMRI methods has been used in humans have identified several regions of cortical tissue along the ventral visual pathway of the macaque that also respond selectively to images of faces (Logothetis et al., 1999; Tsao et al., 2003). While the size, exact number and location of these face selective patches can differ between subjects (Pinsk et al., 2009) observations from several laboratories robustly localize a middle face patch (MFP) on the convexity of the Superior Temporal Sulcus (STS) in posterior TE (Bell et al., 2009; Pinsk et al., 2005; Tsao et al., 2003). The similarities between the set of fMRI identified face selective patches described in humans and macaques has been used to argue for the evolutionary importance of faces as an important category of visual images (Tsao et al., 2008; Tsao and Livingstone, 2008). Owing to the robust fMRI

response, and the anatomical location of the MFP, several researchers have proposed a direct homology to the human FFA (Bell et al., 2009; Nasr et al., 2011; Rajimehr et al., 2009; Tsao et al., 2003; 2008, but see Ku et al., 2011).

While human, and later monkey studies, have approached face processing using fMRI, early neurophysiology studies had established the presence of face selective neurons, or neurons that responded more to images of faces than to non-face distractors, in the temporal lobe of the monkey. (Desimone et al., 1984; Perrett et al., 1982; 1984; Tanaka, Saito, Fukada and Moriya, 1991; Yamane et al., 1988). These studies spatially characterized face selective neurons as distributed throughout area TE (a downstream region of the ventral visual pathway, located on the ventral surface of the temporal lobe), with some evidence for column like clustering on the order of 1-2 mm (Perrett et al., 1984); and periodic "patchy" clustering (Harries and Perrett, 1991) which consisted of repeating spatial zones, 3-6mm in anterior to posterior spatial extent, separated by 3mm regions along the length of the upper bank of the STS. An extensive study (Baylis, Rolls and Leonard, 1987) that included more than 2600 neurons recorded throughout IT reported that the greatest proportion of face selective units in any cortical area was found to be ~20% located in the ventral surface of central IT, and on the convexity of the STS (TE3, TEa and TEm respectively[2]). Optical imaging has also demonstrated an example of spatial organization for head pose that extended ~1.5mm along the cortical surface (Wang et al., 1996; 1998).

Experimental studies that have specifically targeted the fMRI-defined MFP for neurophysiological recordings have reported conflicting estimates on the purity of category selective neurons. Tsao and colleagues originally reported that nearly every

[2] Though it should be noted that these regions were anatomically defined and much larger than fMRI activations. For example TEa (convexity of the STS), used by Baylis et al 1987, included almost the entire length STS.

cell encountered (~97%, or 90% for excitatory only cells) exhibited face selectivity (Tsao et al., 2006), though later reports from the same group have ranged from 82-94%: 94% (Freiwald et al., 2009), 90% (Freiwald and Tsao, 2010); 82% (Ohayon et al., 2012). Using similar methods, Bell and colleagues observed that the proportion of category selectivity sites in the MFP was ~42%, much lower than previous estimates (Bell et al., 2011). Given the average observed volume of the MFP (~70 mm^3: Tsao et al., 2008), purity estimates less than 50% could reflect a similar 3-6mm "patchy" spatial organization as reported for face selective cells in the upper bank of the STS (Harries and Perrett, 1991). Conversely, previous estimates on the purity of category selective sites in the MFP have relied on geometrical projections estimated from the top of the animal's head (e.g. the recording chamber), anatomical MRI, and microdrive readings to estimate the location of recorded samples. This method is problematic because the resolution of spatial sampling is limited. In principle, the effective spatial resolution of the samples could account for the discrepancy reported in previous results: studies that estimated high purity in the MFP could have sampled from a relatively small area of the total patch, inflating the overall estimate of the purity, while studies that found low purity estimates could have inadvertently sampled outside the putative patch. Which calls attention to another issue that has made previous studies problematic; there has been no attempt to estimate the size of the patch itself using the same method to estimate the purity. Spatial bias in the sample of sites used to estimate the purity in the patch becomes difficult to ascertain without spatially sampling the full extent of the patch. Given the density of neurons and the geometry of the cortical tissue that displays a category selective signal with fMRI, it would be difficult to visualize the 3D structure without some simplifying assumptions and/or novel experimental techniques.

To examine these issues, we performed multiunit recordings from 100s of sites targeted to the fMRI identified cortical region that contained the MFP in 2 macaque monkeys. Using a previously described novel X-ray imaging system (Cox, Papanastassiou, Oreper, Andken and DiCarlo, 2008), and 3D cortical models (Dale, Fischl and Sereno, 1999; Fischl, Sereno and Dale, 1999), we estimated the spatial location of our recorded sites on flattened 2 dimensional representations of the MFP region. The X-ray system allowed us to sample the spatial area while 2D mapping allowed for both the visualization and characterization of the functional structure in a simplified 2D format. We found evidence for an area of enrichment ~6mm diameter with a categorical preference for images of faces. The fraction of category selective sites in the enriched area was modeled over the cortical distance on a 2D sheet that contained the fMRI face selective signal. The fraction of category selective sites in the center of the enriched area approached 84%. The fraction of category selective cells gradually fell to a baseline fraction of 3-7% outside of the enriched zone, in good correspondence with some previous reports on the distribution of category selective face cells in the temporal lobe of the macaque (Baylis et al., 1987), but not others (Bell et al., 2011).

2.2 Material and Methods

2.2.1 Subjects

Two macaque (macaca mulatta) subjects, described herein as M1 (male) and M2 (female), were prepared for MION enhanced functional imaging and multiunit neurophysiology as described previously (Op de Beeck, Deutsch, Vanduffel, Kanwisher and DiCarlo, 2008). All procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care and followed the guidelines set forth by the National Institutes of Health.

2.2.2 Data Acquisition

Awake functional imaging: A plastic MRI compatible head-post was attached to the subject's skull under aseptic surgery conditions. Upon recovery the animal was trained using standard operant conditioning methods to fixate in a 3-4 degree window and to adopt a sphinx position while sitting in an MRI compatible chair (Vanduffel et al., 2001). The subject was rewarded for constant fixation in the response window as 5 degree images of face and non-face distractors were presented at the center of the screen . Images were randomly jittered on each trial (+/- 0-2 deg in both azimuth and elevation; uniform distribution). Images were displayed for 250ms, with an inter-stimulus interval of 500ms. Eye movements were monitored with an optical ISCAN system (ISCAN Inc., Woburn, MA). Time points where the animal broke fixation for >250ms were excluded from further analysis. Images were shown in blocks of 3-5 stimulus categories (faces, bodies, places, objects, and scrambled faces) with 20 different exemplars in each category displayed in a random order. Analyses of the functional imaging data was conducted using the FS-FAST toolbox (http://surfer.nmr.mgh.harvard.edu/fswiki/FsFast) and custom written scripts in MATLAB (Mathworks, Natick MA).

MION enhanced functional imaging was conducted on either a 3T Siemens Tim Trio (M1) at the Athinoula-Martinos Imaging Center at MIT or a 3T Siemens Allegra imaging system (M2) at the Athinoula-Martinos Imaging Center at Charles Town, MA. Functional data (TR 3.2s or 3s, 46 or 45 slices, 1.25 mm isotropic voxels with a 10% slice gap) were collected with a custom built, single loop surface coil. Image presentation and water reward was controlled with experimental presentation software: either MWorks http://mworks-project.org (M1); or similar software developed in-house (M2). In brief, functional data was motion corrected across all sessions and co-

registered to the animals anatomical reconstruction with FMRIB's FLIRT software package. Field scans were taken during each session conducted at the MIT imaging center and used to correct magnetic field distortions with FSL's FUGUE software package (Jenkinson, Beckmann, Behrens, Woolrich, and Smith, 2012) in M1. Data from M2 has been reported previously (Issa and DiCarlo, 2012; 2013; Op de Beeck et al., 2008).

Multiunit electrophysiology: At the conclusion of functional imaging experiments, the animals were prepared for neurological recording by the placement of a plastic, MRI compatible recording well (18" diameter; Crist Instruments Inc.), under aseptic conditions, and targeted to the middle face patch. The chamber was placed so that the MFP area described by fMRI could easily be accessed. M1: Right side, Horsley-Clarke center AP coordinates +3mm, with a 9 degree angle. M2: Left side, Horsley-Clarke center AP coordinates +14mm, with a 8 degree angle. The animals were trained to sit upright in a standard neurophysiology recording chair and fixate in a 2-3 degree response window while viewing images of faces and non-face objects in an RSVP sequence (M1: 7-10 images/trial, 200ms on 100ms off or M2: 12-15 images/trial, 100ms on and 100ms off). Images were repeated 12/image (M1) or 3-5/image (M2). Eye movement traces were collected and monitored with an optical EyeLink system (SR Research Ltd., Kanata, Ontario Canada). Trials where the animal broke fixation were aborted, and only images presented before an eye movement were considered. The first image in an RSVP sequence was always disregarded from further analysis.

Multiunit recording was conducted with single, glass coated tungsten microelectrodes (0.5-0.7 MΩ: Alpha-Omega Co., Alpharetta, GA), amplified with a BAK system (BAK

Electronics, Mount Airy, MD). Neural signals were sampled at 14 or 8 Khz and bandpass filtered with an inline butterworth filter (Krohn-Hite, Brokton MA) between 300hz and 7Khz or 300hz and 4Khz. Experimental stimulus and reward control, as well as data recording and storage, were managed by MWorks software (http://mworks-project.org) on a Mac Pro running OS 10.5-6. Straight (Crist Instruments Inc.) and custom made angular (5 and 7 degree) well grids were used to position a metal guide tube ~5-6mm from the STS. Microelectrodes were lowered using a microdrive (Crist Instruments) and, carefully listening for cell transitions, the electrode was advanced until the distinct sound of crossing the STS was heard.

2.2.3 Data Analysis

fMRI Analysis: Standard univariate methods were used to analyze fMRI data as implemented in FS-FAST. The functional signal was smoothed (2.5 mm) in volume. To localize category selective voxels for neurophysiology experiments, we examined all visually active voxels (voxels that had a significant response to visual images; $p > 10e\text{-}6$: faces, bodies, and objects - scrambled objects). The selectivity to faces at these voxels was used to target electrode penetrations with the use of an X-ray imaging system. The Caret software package was used to visualize activations on inflated surfaces of the two monkeys (Van Essen et al., 2001).

Neural Analysis: The multiunit response to an image was recorded as the spike count rate in a fixed window 60-160 ms post stimulus onset. A global estimation of the baseline (the mean response 0-50 ms post stimulus onset across all image repetitions) was removed from the response to each exemplar. The average response to all repetitions of each exemplar image in a given category (faces of non-face objects) was

used to estimate the mean and standard deviation of the category selective response. Category selectivity for each site was estimated as d':

$$d' = \frac{\bar{X}_{Faces} - \bar{X}_{Objects}}{\sqrt{\frac{\sigma^2_{Faces} + \sigma^2_{Objects}}{2}}}$$

In order to make some of our results directly comparable to the efforts made in previous studies, we also calculated a face contrast metric as a category selectivity index (Bell et al., 2011; Freiwald et al., 2009; Tsao et al., 2006) :

$$fsi = \frac{\bar{X}_{Faces} - \bar{X}_{Objects}}{\bar{X}_{Faces} + \bar{X}_{Objects}}$$

Following from previous studies, we defined sites where (1) $X_{faces} < 0$ and $X_{objects} > 0$; fsi = -1, (2) $X_{faces} > 0$ and $X_{objects} < 0$; fsi = 1, and (3) $X_{faces} < 0$ and $X_{objects} < 0$; fsi = -fsi.

In order to determine how reliable the face selectivity metric was at each site, we used resampling methods to find the correlation between the selectivity measured from random subsamples of the data. Repeated trials to a given image were randomly assigned to 2 pools, and the faces vs. object selectivity measure was calculated for each pool. This procedure was repeated (n = 1000) across all sites and the average correlation between the two sets of measures, corrected for the split-half procedure (Spearman, 1910), served as an estimate of the reliability of the data set.

We wanted to know how various factors impacted the selectivity measurements used in our modeling analysis. To determine this we estimated the average squared error between two estimates of the faces vs. objects selectivity as a function of the distance between the sites that contributed to the measure. The metric was calculated as a

function of the distance (from 300um) between sites in a single penetration as measured on the microdrive. Additionally, we examined the impact that limited numbers of image repetitions and the variability due to trial repetitions had on our selectivity estimates.

To determine the contribution of limited sampling at the same location in the cortex, the trial repetitions at each site were split into two random groups. The average squared difference ($\Delta d'$) in faces vs. objects selectivity estimates between each split was estimated across all sites. To account for the arbitrary pooling of trials, n = 1000 repetitions of the procedure was done, and we report the average value. Noise in the estimate of category selectivity could also be due to the limited set of images used to estimate category selectivity. To account for this variance source, we followed a similar procedure as above, splitting the data at each site into two estimates of category selectivity by subsampling the face and object images used in the experiment. This procedure actually includes the former variability source (variance from trial repetitions), and therefore should be larger in magnitude. To estimate within penetration variability, we examined differences in category selectivity ($\Delta d'$) between pairs of sites given a particular distance between them as measured on our recording microdrive (i.e. no X-ray information was used to estimate the spatial distance between the recorded sites). In all cases we ensured that equal numbers of samples were used to estimate category selectivity in all three cases (repetition, image and within penetration). In some cases, this meant subsampling from one source of variability. For example, the same number of images and repetitions went into the estimates of category selectivity for estimating the variance due to image repetitions as went into estimating the variance within penetration as a function of the distance between sites.

Measurements of purity were estimated by counting the proportion of sites that met a "face selective" criteria (either d' or fsi) as a function of the 2D spatial distance from the center of the MFP (e.g. the radial distance from the center). Multiunit sites were binned together arbitrarily into non-overlapping bins (M1 = 28 sites/bin; M2 = 31 sites/bin) and the proportion of face selective sites in the bin was used as an estimate of the purity. The location of the bin was the average radial distance of the sites allocated to the bin. The resulting purity function was smoothed over 5 bins for presentation in figure 6. As the exact position of the center was dependent on the specific model, we estimated the purity with each of our models.

We examined the distributional form of the faces vs. objects selectivity metric estimated from multiunit sites located in each of the three course regions around the MFP (defined by the model analysis). Each sample was fit to a Generalized Extreme Value (GEV) function. The faces vs. objects selectivity (i.e. d') estimate at each site was collapsed across M1 and M2 according to the isogaussian model. We accessed the goodness of fit for our data to the GEV distribution using a 2 sample KS test, modified for bias in estimating the distribution parameters from empirical data (Clauset, Shalizi and Newman, 2009). The parameter values were used to characterize the distributions.

2.2.4 Spatial Modeling

Cortical surface models: Anatomical models of the white and pial surfaces were estimated from multiple (6-8) high resolution anatomical MRI volumes (500um isotropic T1 weighted anatomical volumes) taken under anesthesia in a MRI compatible stereotaxic frame (Christ Instruments). The mid-layer surface model used in our analysis was estimated by taking the midpoint between the estimated white and pial surfaces. The area of the surface to flatten was chosen arbitrarily by centering a point in the fMRI identified MFP and finding a closed boundary of mesh nodes that were a

specified distance (7mm) from the center. The radius was chosen so that it maximized the cortical surface for the analysis, but did not include face selective activity from anterior or posterior patches. High resolution (i.e. small inter-node distance) patches were created by up-sampling the closed mesh using custom scripts and toolboxes in MATLAB (Wavelet meshes toolbox; Peyre 2007). Finally, the manifold was computationally flattened to preserve geometry with the MR tools MATLAB toolbox (Heeger Lab: http://www.cns.nyu.edu/heegerlab/wiki/doku.php?id=mrtools:top).

X-ray localization and registration: The electrode position for every site sampled was estimated in 3D space using a custom built stereo micro-focal X-ray system (Cox et al., 2008). Briefly, the X-ray system used 2 X-ray sources (Oxford Instruments; Tubney Woods, Abingdon, Oxfordshire UK) and digital image capturers (Shad-O-Snap 1024; Teledyne/Rad-icon Imaging Corp., Sunnyvale CA) positioned around the monkey's head in the recording set up. Six brass fiducials (diameter ~500um) were positioned in known locations on a rigid frame attached to the animal's skull. The X-ray system produced images in 2 known planes that contained both the fiducials in the frame and the electrode tip. Custom software was used to reconstruct the 3D location of the electrode tip, based on the known locations of the fiducials and the geometry of the two X-ray images. Wells drilled into the frame at known positions, were filled with CuSO4 and an MRI anatomical volume was used to register the electrode positions to the high resolution anatomical volume using FMRIBs FLIRT and FNIRT registration tools (Jenkinson et al., 2012).

The recorded sites were co-localized to the 3D anatomical volume of each subject. This registration was carried out in two steps, a linear affine registration (Figure 2A left inset)

followed by a nonlinear registration using FMRIB's FNIRT tool. The non-linear registration process was used to adjust for the distortion in shape between the reference volume and the cortical model. These adjustments were applied to the 3D positions of the sampled sites (inset Figure 2A right), generally resulting in small changes in spatial position relative to the linear registration. Each recording site was than projected to the closest orthogonal node on a surface manifold created from each subject's anatomical volume. We discarded any recorded sites that moved >1250um from the original 3D location to the projection site on the mid layer surface model (~14% Monkey 1; ~25% Monkey 2). Spatial analysis were conducted on high resolution flattened 2D surfaces of the area around the fMRI identified MFP. The location of each recording site on the 2D surface was recovered as the node identity after flattening.

Model fitting: We fit a gaussian and a simple circle ('Box Car') model by least squares to the estimated 2D spatial locations and the faces vs. objects selectivity at each recording site. Evaluation of the models was conducted with custom written code in MATLAB and modified toolboxes (Mathworks; Mineault 2011). In brief, each model predicts a selectivity value for each site given its spatial 2D location. The gaussian and box car model produce a weight value (the prediction weight) based on a set of parameters, which include a 2D center position (x_0 and y_0), and a measurement of dispersion (i.e. the radius of the MFP). The prediction weights are then fit to the data by a least squares procedure. The model parameters and linear weights were estimated by a corse to fine brute force search of the parameter space, minimizing the error between the estimated selectivity and the observed selectivity of the data. The domain of the spatial parameters was limited by the size of the cortex defined by the flattening procedure (e.g. the radius or center position could not extend to a position outside the

flattened mesh). The standard error of the parameter values were estimated by bootstrap methods.

The Non-linear models used to estimate the prediction weights (which ranged from [0, 1]), were defined as (1) a circle, or "Box Car" model (M_{Box}), or (2) a 2D spatial gaussian model ($M_{gaussian}$). The dispersion parameter for M_{Box} was simply the radius to the center (e.g. the 2D distance from each site determined if the site was in or out of the patch), while the dispersion parameter for $M_{gaussian}$ was the FWHM of the gaussian (either 1 parameter for the isotropic model or 2 parameters for the full gaussian). There was an additional rotation parameter (θ) for the full gaussian model that defined the angle of rotation about the 2D frame. In the case of M_{Box} the prediction weight was defined as 1 for sites whose 2D distance to the center was less than the radius parameter and 0 for all other sites. The prediction weights for $M_{gaussian}$ were set to the height of a standard gaussian distribution based on the model parameters and the spatial location of each site.

To examine the 3D layer information in our data, we created upper and lower layer cortical surface models. These models were generated following similar procedures as described previously. The upper surface was located at 80% of the total thickness of the cortical ribbon, which would approximately be 500um from the pial surface. Similarly a "lower" surface model was created at 20% of the cortical thickness, or approximately 500um from the estimated white matter border. In order to uncover any possible differences between the upper and lower layer sites in our data, we limited the analysis to sites that moved less than 500um from their original 3D position to their projected

location on the either the upper or lower layer models (M1: $n_{upper} = 269$, $n_{lower} = 70$; M2: $n_{upper} = 111$, $n_{lower} = 365$).

2.3 Results

MION enhanced, awake fMRI was conducted in two macaque subjects (Monkey 1 and Monkey 2) to localize the middle face patch (MFP) in the Temporal lobe (Figure 1). Conventionally posed and cropped images of unfamiliar conspecific faces and familiar everyday objects were used in both the functional imaging and the neurophysiology experiments. Unthresholded t-maps for all object driven voxels (see methods) that responded greater to images of faces, demonstrate three large clusters of voxels along the posterior, middle, and anterior convexity of the Superior Temporal Sulcus (STS). In both subjects the posterior patch was located near the inferior occipital sulcus on the gyrus between the Posterior Middle Temporal Sulcus (PMTS) and the STS in area PIT (Posterior Inferior Temporal cortex). The MFP extended across the lip of the STS, localized near the end of the PMTS in CIT (Central Inferior Temporal cortex). The anterior patch was large and localized on the gyrus between the STS and the Anterior Medial Temporal Sulcus (AMTS) in AIT (Anterior Inferior Temporal cortex; AIT, CIT, and PIT after Felleman and Van Essen 1994) . As reported previously we observed a number of other clusters in the STS (possibly corresponding to MF and AF or AM; (Tsao et al., 2008) though we choose here to focus on the three lateral patches. The MFP was approximately similar in spatial extent across the two animals, and was observed to begin at AP +5.5mm in Monkey 1 and AP +7mm in Monkey 2. In both animals the activation extended across ~5 mm and predominately covered the crown of the Superior Temporal gyrus.

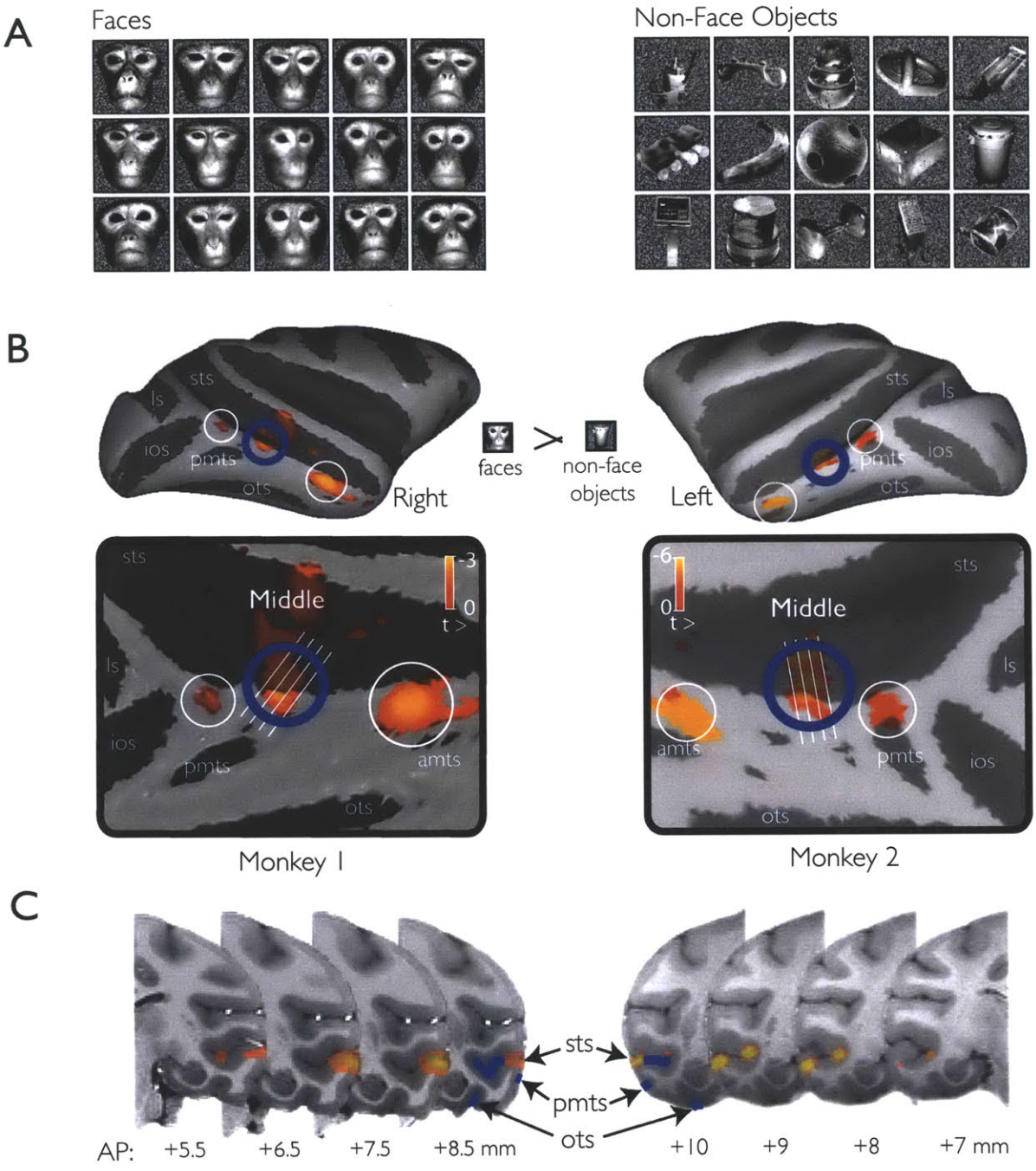

Figure 1. MION enhanced fMRI localization of face selective patches in the Macaque Temporal Lobe. Conventional images (A) of macaque faces and familiar everyday objects used to localize areas in the Inferior Temporal cortex that responded selectively to images of faces. Three discrete patches (B) were found on the convexity of the Superior Temporal Sulcus in inflated and flattened surface models of the Temporal lobe. We examined the fMRI face selective signal in the right (M1) and left (M2) hemispheres of the two subjects. (C) The Middle Face Patch was observed consistently across both animals on the convexity of the STS (i.e crown of the Superior Temporal Gyrus). Right and left hemispheres shown in coronal slices.

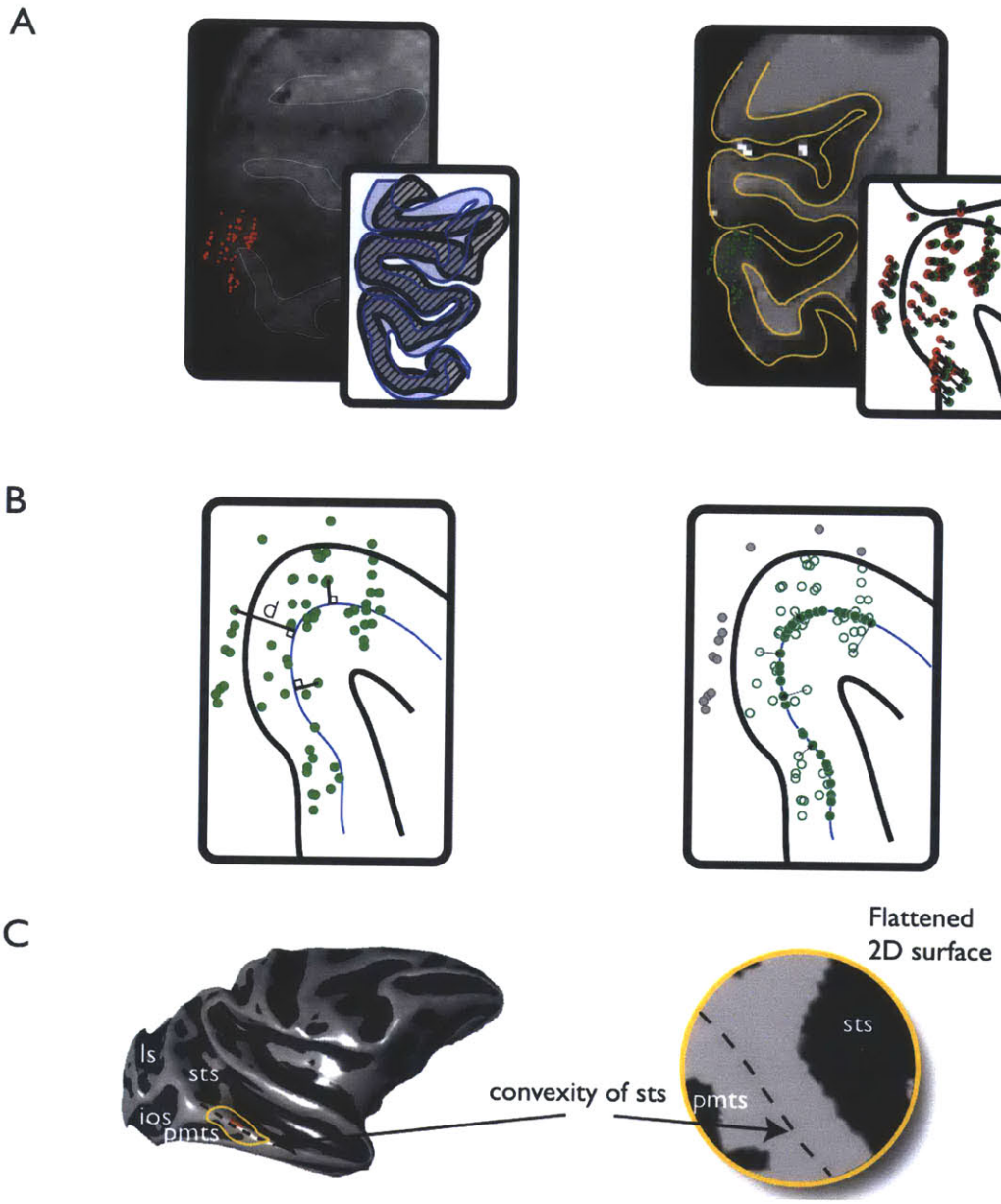

Figure 2. Spatial registration and analysis methods. The 3D spatial locations of all sampled in the MFP region were estimated with a custom built stereoscopic X-ray imaging system. The 3D locations were registered (A: Left) to a high resolution (500um) anatomical MRI. To improve the overall accuracy of the registration (A: inset), non-linear registration using FMRIB's FNIRT registration tool was performed (A: Right), and the resulting transform was applied to the estimated positions of the recording sites (A: Inset). The spatial position of each recording site was projected to the closest orthogonal node of a high resolution mid layer mesh of the cortical surface (B: Left). Sites that moved a distance greater than 1250um from their original 3D position to the 2D surface manifold were excluded (B: Right) from further analysis. An arbitrary portion of the Temporal lobe was selected for all spatial analysis (C: Left). This area was selected by choosing the approximate center of the fMRI MFP activation in each monkey and including all nodes within an arbitrary geodesic radius (7mm). The radius was chosen to be approximately the greatest distance that would not extend into face selective activation at posterior (PL) or anterior (AL) locations.

Neurophysiological experiments were conducted with standard single-electrode methods, but were targeted in and around the location of the MFP using a custom designed X-ray imaging system. That system allowed us to (1) reconstruct the 3D position of the recording electrode tip at each site sampled in IT cortex and (2) project their positions to a 3D Freesurfer cortical model of each subject's brain. A summary of the steps involved in the registration to the cortical anatomy is shown in Figure 2 (see Methods for details). The spatial extent of the region where sites were sampled was arbitrarily chosen to maximize the spatial extent of the anatomical coverage without extending so far as to encroach upon the anterior or posterior face patch regions observed with fMRI. Under these constraints, a 14 mm region centered on the MFP (e.g. on the inflated surface) in each monkey was chosen for inclusion in this study (Figure 2C).

The face selectivity at each recorded site was measured as d', and all sites were projected from their original 3D X-ray estimated position to a midlayer surface representation for each monkey. The surface was then computationally flattened (see methods). As shown in Figure 3, there is an enriched zone of sites with a response preference for images of faces over non-face distractors. Similar to the fMRI, the enriched zone fell along the Superior Temporal gyrus in CIT. Hereafter, we refer to this physiologically-measured region of enriched face selectivity as the physiological MFP (pMFP), as its relationship to the fMRI-defined MFP is the not the goal of this study, but see Issa et al (2013).

While there was a clear enrichment for face preferring sites in the pMFP, it is also evident that there is variation in the category selective signal throughout the pMFP. Some sites displayed a stronger preference for faces than other nearby sites (note different sized red circles in Figure 3), and some sites showed inverted category

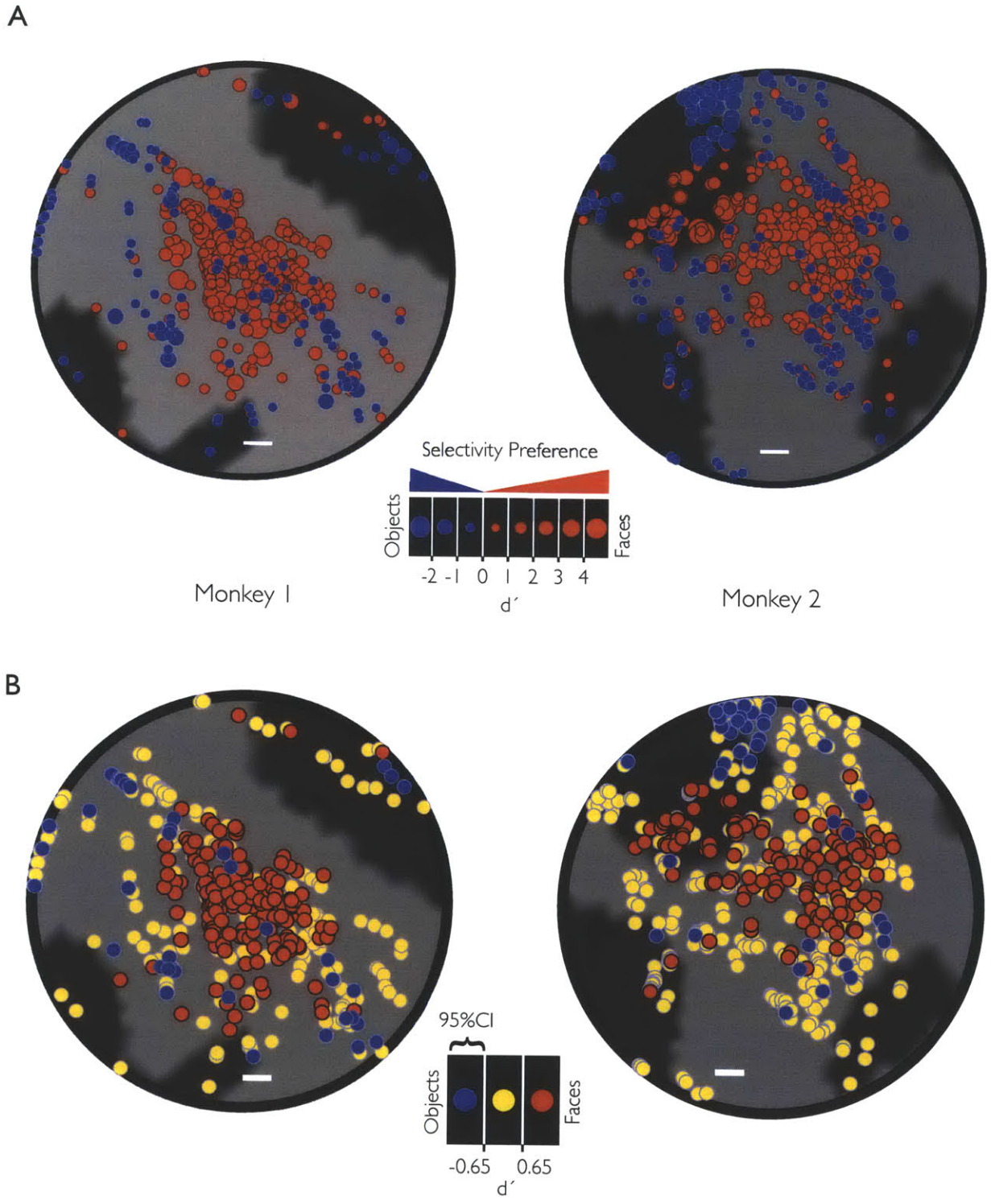

Figure 3. The physiologically defined MFP (pMFP) is a region of cortex with an enrichment of category selective sites. (A) Flattened 2D regions of the Temporal Lobe highlight category preference (red = faces; blue = objects) across the spatial region of the fMRI defined MFP. The size of the circle indicates the d' value at each recorded site. (B) Category preference is indicated by color as above, but yellow indicates sites where the 95% confidence interval was not greater than the estimated 2x criteria for face and object selectivity.

selectivity -- they clearly preferred non-face images over faces (note large blue circles in Figure 3, see legend). This variation in measured face selectivity (d') is replicable in that it is not due to well-known physiological "noise" (Poisson spiking variability). We estimated the reliability of the category selectivity for faces at each site by randomly subsampling (n = 1000) our data over trials into two equal groups, and correcting for the number of splits (Spearman, 1910). The corrected correlation (e.g. reliability) for our data was $r > 0.96$ in both animals. As a more stringent examination of the data we asked which sites were face or object selective, as defined by previous criteria (the site responded at least twice as much to their preferred class on average--see figure 8 for the estimate of this value). We then estimated the 95% confidence interval for d' at each site by boot-strap methods. Each site that had a d' estimate that was greater than the 2x threshold *and* whose 95% confidence interval also did not include the 2x threshold were considered category selective (depicted as non-yellow circles in Figure 3B).

To characterize the spatial profile of the pMFP, we fit 3 different spatial models to the unthresholded 2D data (Figure 4). These assumptions were based on implicit models from the fMRI literature; (1) a simple "in vs. out" module or "box car" type model, (2) a circular (isotropic) gaussian model, and (3) a three parameter gaussian model. All three models were parameterized by a center position (x and y). In addition, the "Box Car" model was parametrized by a single radius parameter (i.e. In vs. out), while the isogaussian model was parameterized by a measure of dispersion (sigma), and the full gaussian model had two parameters of dispersion and an angle parameter to allow rotation on the 2D cortical surface. These models were fit to the data by linear regression. In Monkey 1 all three models resulted in approximately similar fits to the data ($r^2 \sim 0.3$), while Monkey 2 demonstrated a slightly better fit with the higher parameter models (r^2: 0.33). In each monkey, the absolute center of the model was consistent over all three models (Figure 5). The models also made very similar

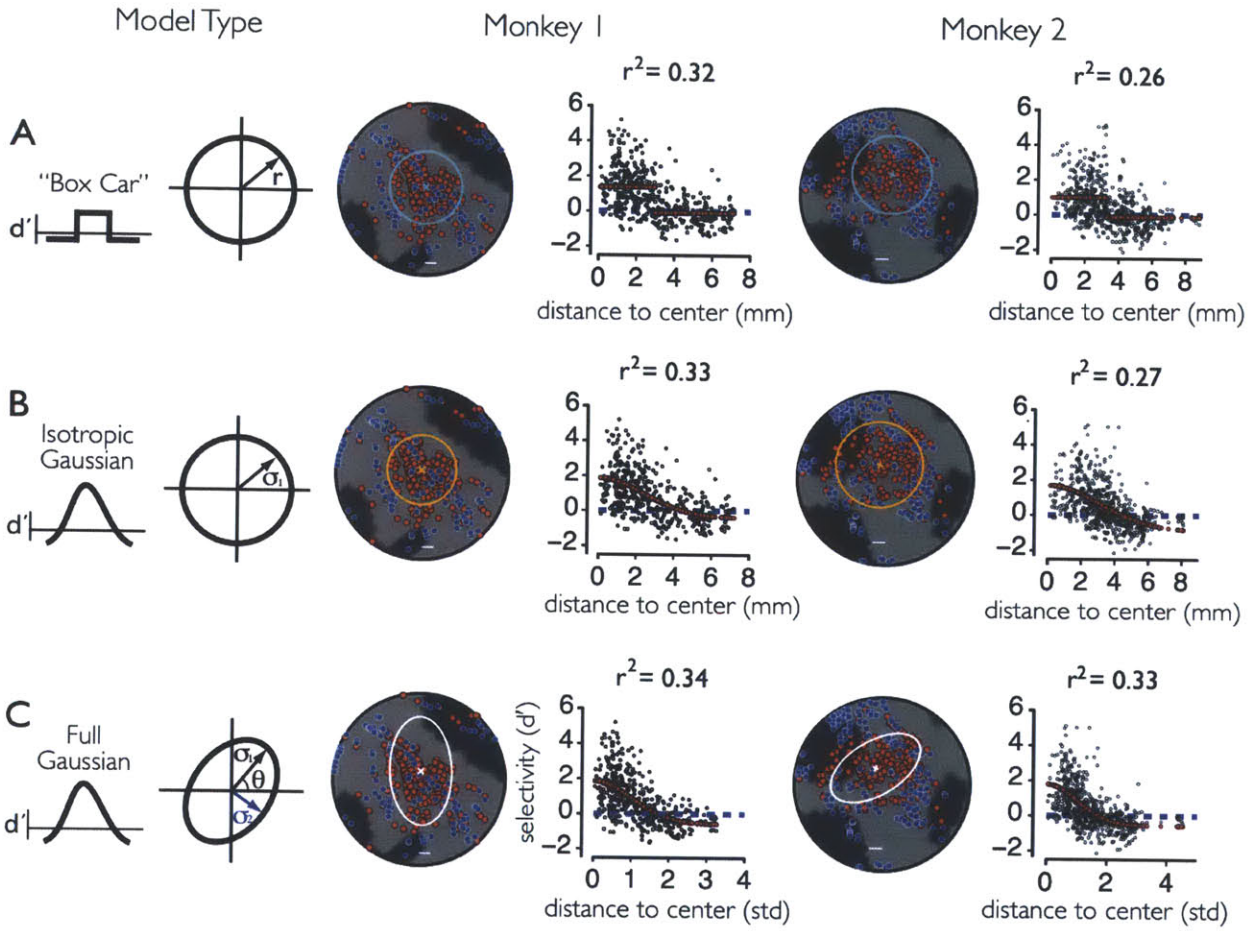

Figure 4. 2D Models of the category selective spatial region. Three models for the spatial structure of the pMFP were used to fit the category selectivity and 2D spatial position over all multiunit samples. Column 1 summarizes each model and their parameters (see methods). The 2D schematic depicts the range of estimated selectivity values as a function of 1D spatial position and the two linear estimators for the model, while the parameters for the non-linear predictor formulas are described in the adjacent diagram. Columns 2 and 3 summarize the results in each subject. Flattened maps displaying the data are depicted as well as the best fit model. The selectivity at each site was collapsed across radial distance to the estimated center of each patch (black dots) alongside the model prediction for each site (red dots). The number of standard deviations are plotted along the abisca for the full gaussian model as this isocontours of this model were not radially symmetric.

predictions on the size of the enriched face zone. The radius parameter of Model I was taken to be the estimated size of the patch under a module type model (e.g. the diameter of the MFP in the 2D flattened space would be twice the size of the estimated radius), while the FWHM for the gaussian distributions in Models II and III were taken as a similar estimate for the size of the "patch." Table 1 tabulates these parameters as well as the model fits for the parameters. It is notable that, while all models well-

capture the mm-scale spatial profile of the physiological MFP (e.g. see Figure 3), the overall fits of all the models ($r^2 \sim 0.3$) were low. For example, examination of the 1D collapsed model plots (Figure 4) show that there is a large range of values between the actual selectivity in the data at a given spatial position and the selectivity predicted from the model. This is consistent with the visual observations described above, and suggests that none of the models examined were able to capture the detailed spatial structure of the physiological MFP (see Discussion).

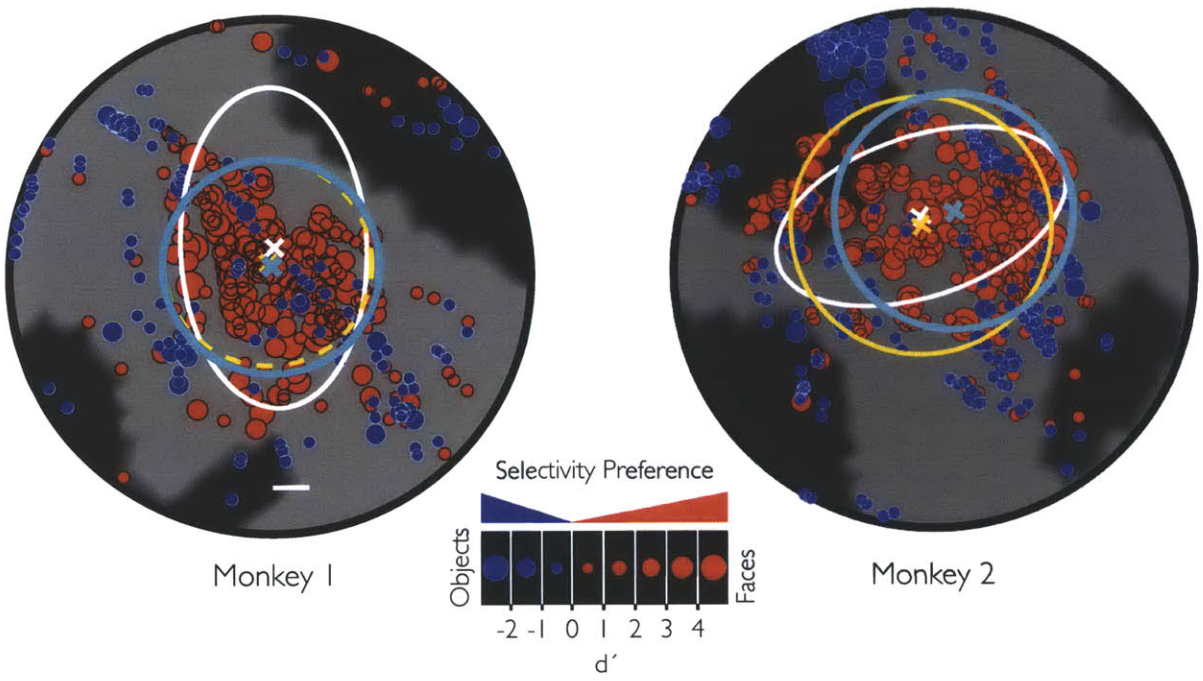

Figure 5. Low frequency 2D spatial models of the pMFP are largely consistent. The spatial location and category preference for each recording site is localized on 2D flattened surfaces of the fMRI identified MFP region. The different best fit models from our analysis are largely consistent on the position of the center and the spatial extent of the enriched spatial region.

To more closely examine this feature of the data (Figure 6) we characterized the variance in selectivity as a function of the distance between pairs of recorded sites (see methods). In all cases, we calculated the average squared error ($\Delta d'$) across all pairs of sites, binned by the distance between the sites (see methods). The minimal amount of

variance we could expect would come from measuring the selectivity at the same site independently (i.e. a "pair" of sites at a distance of zero mm). To estimate this quantity we measured the difference in selectivity measured from random splits of the data. Another source of variation that could affect category selectivity estimates for faces could stem from the limited number of exemplars used to test face and object selectivity. To address this issue we randomly assigned the images used in the experiment into two pools and measured the resulting selectivity estimate to the two image groups and each site. These random splits were done multiple times (n=1000). The variability observed in the model plots (Figure 4) is derived from the fits in 2D space, and therefore does not reflect the true variability between sites, because, by definition, the projection/ flattening procedure moves sites across small distances. To account for these biases, we estimated the average squared error in face selectivity estimates between sites recorded on the same electrode, as measured on our recording microdrive. Previous studies that have examined the correlation in response between sites to a given image set have found evidence for local similarity of shape preference extending 600-800um (Kreiman et al., 2006; Sato, Uchida

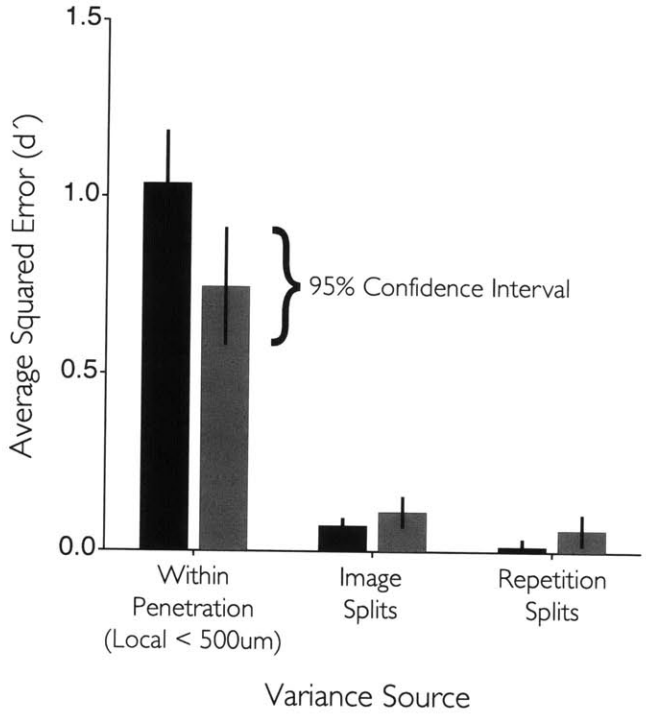

Figure 6. Variance in selectivity estimated at nearby spatial locations is greater than expected in the MFP. The bar graph depicts that average squared error between face selectivity estimated at sites <500 um in distance from each other, recorded from the same electrode. The average variance expected due to the limited number of image exemplars or trials is also depicted for comparison. Errorbars indicate the 95% confidence interval on estimate of the average error.

and Tanifuji, 2009). While category selectivity measures differ somewhat from previous shape preference metrics, previous studies that have specifically used face stimuli have found similar correlations between the multiunit sites in optically defined "face spots" that were < 600um apart (Sato et al., 2009). In our data (Figure 6) we found that the average squared error between face selectivity estimates measured at local sites within a penetration was greater than would be expected due to (1) the variance in category selectivity estimated from the limited number of image exemplars used to estimate category selectivity and (2) response variability, or the response variance observed over multiple presentations the same stimulus. This result suggests that the local variability observed in our spatial modeling analysis in the MFP is genuine variability that is not accounted for by the low spatial frequency signal predicted in the models we examined.

Previous estimates of the proportion of face selective sites in the MFP -- referred to here as "purity" -- have found conflicting results. A number of studies (Freiwald et al., 2009; 2010; Ohayon et al., 2012; Tsao et al., 2006) have reported that the number of face selective units in the MFP was high (greater than 84%), while others studies (Bell et al., 2011) have reported more modest purity estimates (less than 50%). To attempt to resolve this issue, we measured the purity of the MFP as a function of distance from the center of the patch (Figure 7). Since all three models examined produced very similar center locations, we limited our analysis to the two simplest models: the isotropic gaussian model and the module or "box car" model. We defined a neuronal site as "face selective" if it had a d' > 0.65 (response to face images vs. non-face images). To check the dependence of our results on the choice of "face selective" definition, we also used a contrast metric (FSI) similar to what has been used in previous studies to estimate face selectivity (Freiwald et al., 2009; 2010; Tsao et al., 2006). FSI is constructed so that it ranges from -1 to +1, and sites that had an FSI (face selectivity index; see methods) greater than 0.33 were considered to be "face selective" (note that, we used

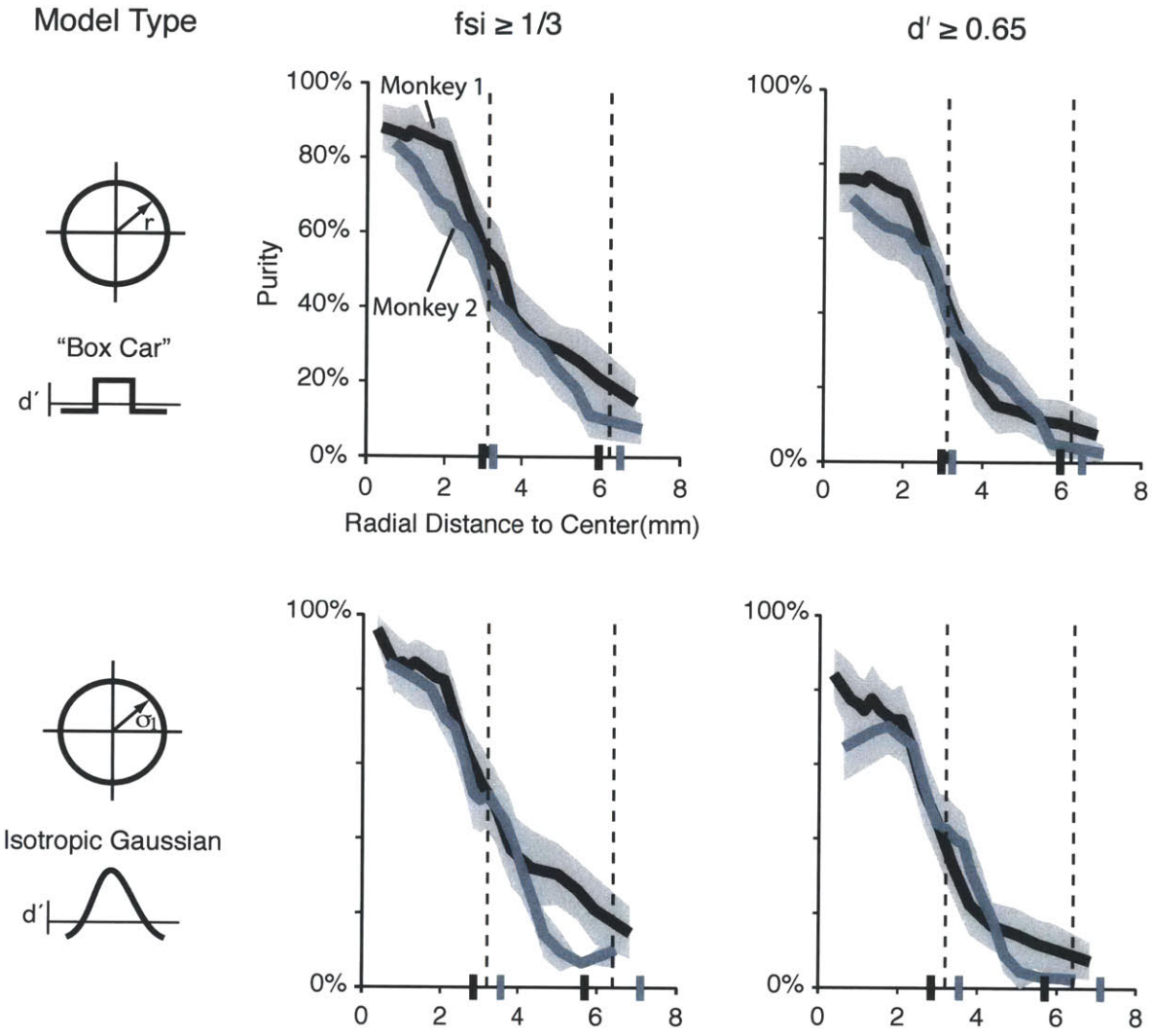

Figure 7. Purity estimates in the pMFP are sensitive to the metric and spatial extent used to define the region. Estimates of the number of category selective multiunit sites in the MFP (the purity) could range from 94% to 58% depending on the distance from the center of the patch. These ranges varied due to both the model and metric used to define the patch. Estimates using FSI tend to increase the purity in and out of the patch. Overall, the purity falls off gradually from the center of the patch to cortical regions outside the patch.

the median d′ from the range of values that corresponded to the fsi = 0.33, see inset Figure 7). We found that the purity of the MFP ranged from near 96% (M1, fsi metric and gaussian model) at its center, to less than 50% for much of its outskirts.

Not surprisingly, the purity profile of the MFP was dependent on the selectivity metric used to define "face selective." Specifically, looser definitions of "face selective" led to

higher reported purity inside the MFP, but also led to higher reported "face selectivity" outside the MFP. To quantify how the purity changed over distance relative to the size of the patch, we defined the spatial region contained within 1 radius from the estimated center of the pMFP (for the "Box Car" model, or 1/2 of the FWHM for the gaussian

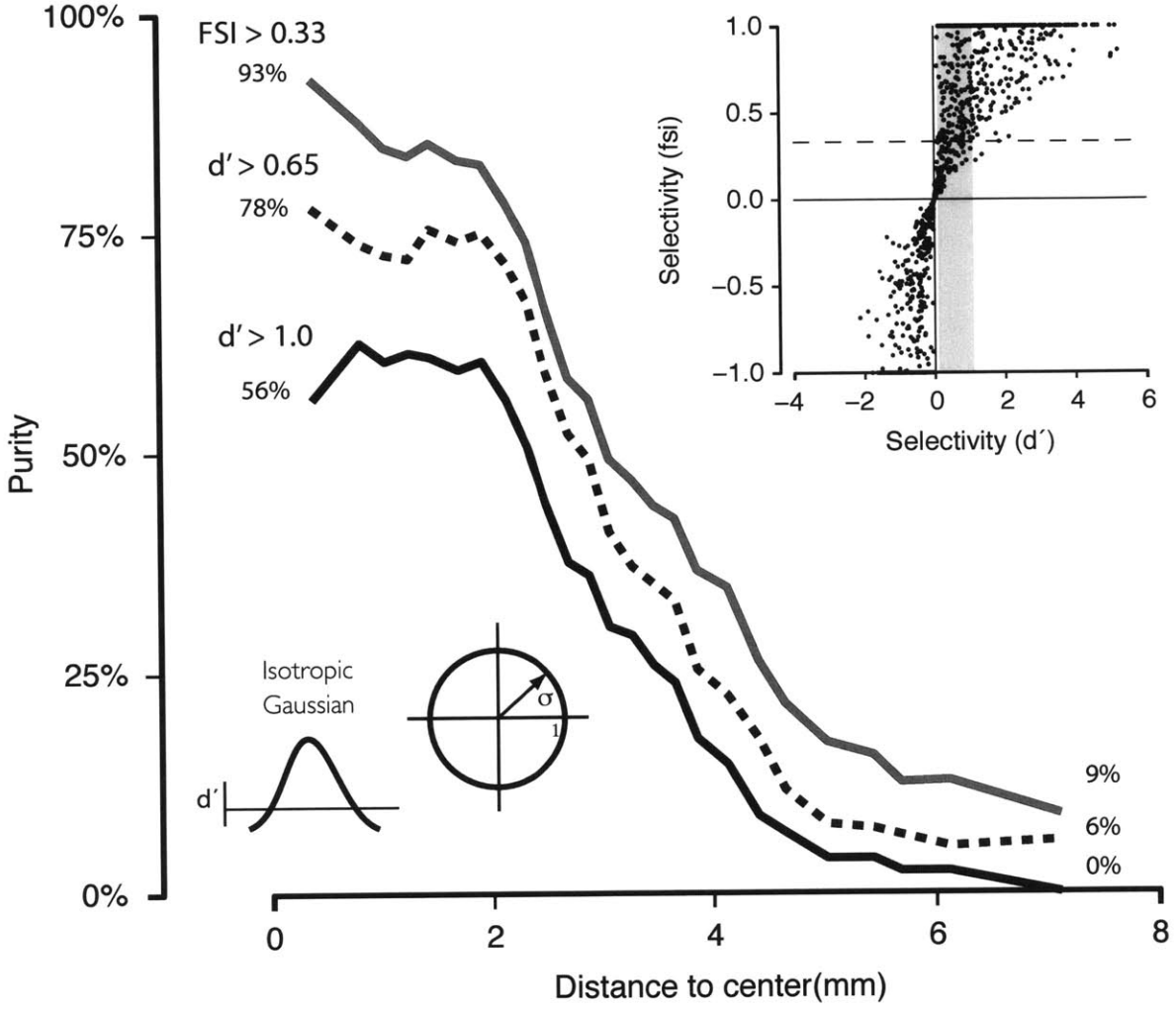

Figure 8. Estimated purity across subjects is largely consistent. We estimated the purity collapsed across both animals using the estimated center form the isotropic Gaussian spatial model. The empirical purity function across both animals is remarkably consistent in shape across different selectivity estimators. Lines show d' estimates of the purity across different thresholds (Line width, thicker to lighter corresponds to higher to lower thresholds) for category selective inclusion (i.e. how a face site is defined). A similar shape was observed with the FSI index that has been used in previous studies (Tsao et al 2006; 2008; Friewald et al 2009; 2010). Not surprisingly, weaker thresholds resulted in higher estimates of the purity near the center and far outside the pMFP.

model) to be "in" the patch. The annular spatial region just outside this, but less than twice this distance of the radius/FWHM was defined to be "near" the patch. Finally, the annular spatial region greater than twice the radius/FWHM from the estimated pMFP center was considered "far" from the patch. Either model or metric produced similar overall results. Using the "Box Car" model the purity "in" the MFP averaged 67% across both subjects (d': M1 = 74%; M2 = 60%; average FSI: 76%, M1 = 84%, M2 = 68%). The purity in the "near" region averaged 16.5% across both subjects (d': M1 = 16%, M2 = 17%; average FSI: 28%, M1 = 32%, M2 = 24%), and the purity "far" from the MFP averaged 6% (M1 = 8%, M2 = 4%; average FSI: 12.5%, M1 = 14%, M2 = 11%).

A strong form of the modular hypothesis of the MFP would argue for purity differences across a boundary defining the MFP, and perhaps a different form for the distribution of face selectivity measured at sites inside the MFP (e.g. arising from selectivity mechanisms unique to the MFP tissue). A weaker hypotheses would, in contrast, predict a gradual increase in the number of face selective sites as one moves from outside the MFP toward its center (e.g. no clear boundary) and a common distribution form for neural sites inside and outside the MFP, with simply a shifted mean value (i.e. such that the average face selectivity is higher inside than outside the MFP). To examine this issue, we analyzed the distribution of face selectivity in the three regions in and around the MFP: "in", "near" and "far" as defined previously. These regions are spatially normalized across the monkeys by model, so the data was collapsed across subjects (Figure 9). The empirical distributions of the face selectivity metric for the neural populations in and around the MFP demonstrate three findings. First, the average face selectivity of the population of neuronal sites "in" the MFP is about 2 times higher than the population of sites "far" from the MFP, consistent with the previous analyses above (Fig. 4). Second, we found that the selectivity distributions from all three spatial areas were significantly non-normal (KS test = [0.41, 0.13, 0.29], in, near,

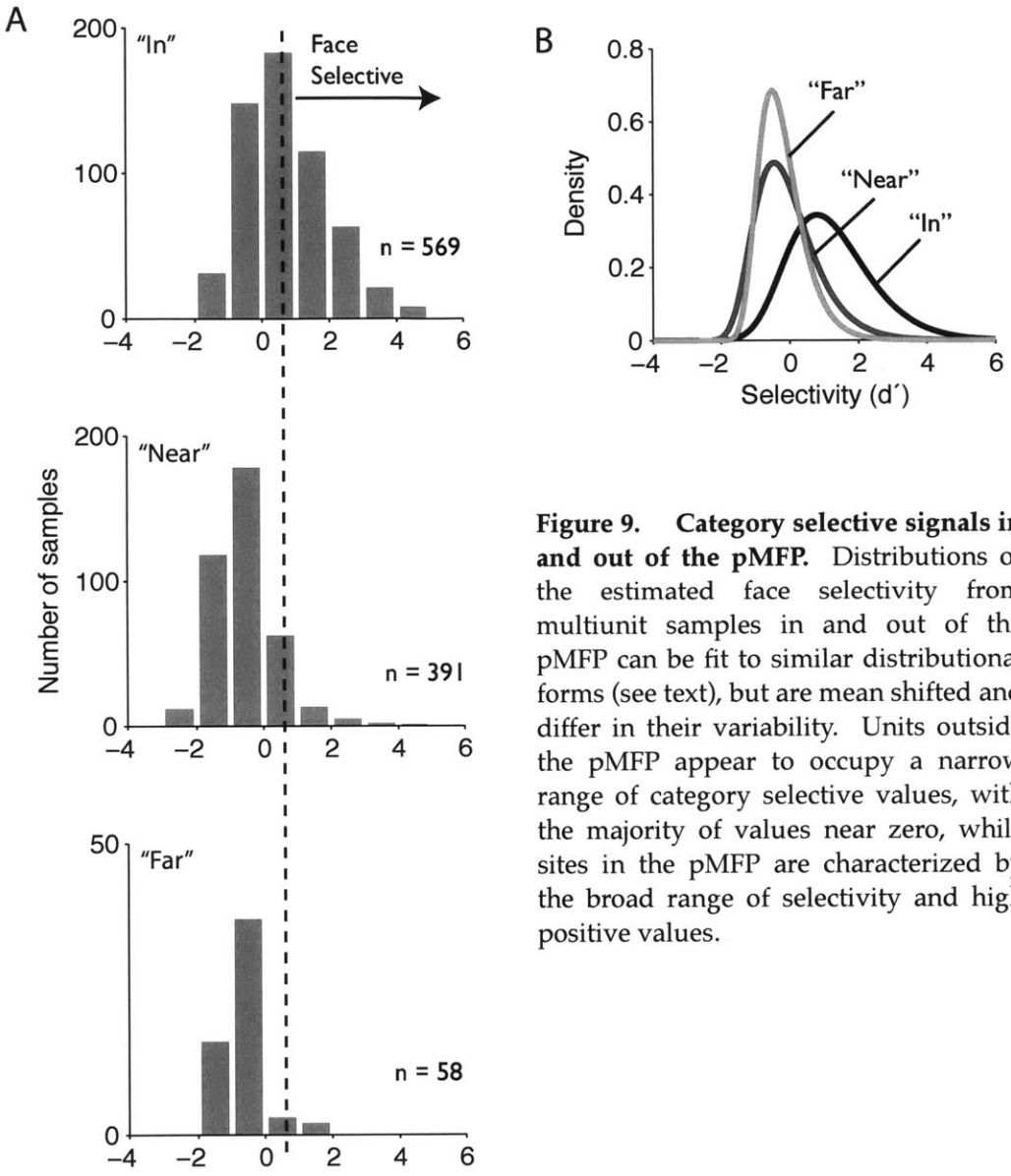

Figure 9. Category selective signals in and out of the pMFP. Distributions of the estimated face selectivity from multiunit samples in and out of the pMFP can be fit to similar distributional forms (see text), but are mean shifted and differ in their variability. Units outside the pMFP appear to occupy a narrow range of category selective values, with the majority of values near zero, while sites in the pMFP are characterized by the broad range of selectivity and high positive values.

and far respectively, and all $p < 0.001$). In order to compare the face selectivity distributions in each spatial region, we fit each distribution to a Generalized Extreme Value (GEV) function. As expected the means between each distribution were different ($\mu_{In} = 0.69$, $\mu_{Near} = -0.48$, $\mu_{Far} = -0.52$). The estimated standard deviation between the distributional forms also differed between the three spatial regions ($\sigma_{In} = 1.08$, $\sigma_{Near} = 0.78$, $\sigma_{Far} = 0.54$). The GEV distribution has a third parameter (**k**) that controls the form of the distribution, this parameter was small and negative over the three distributions

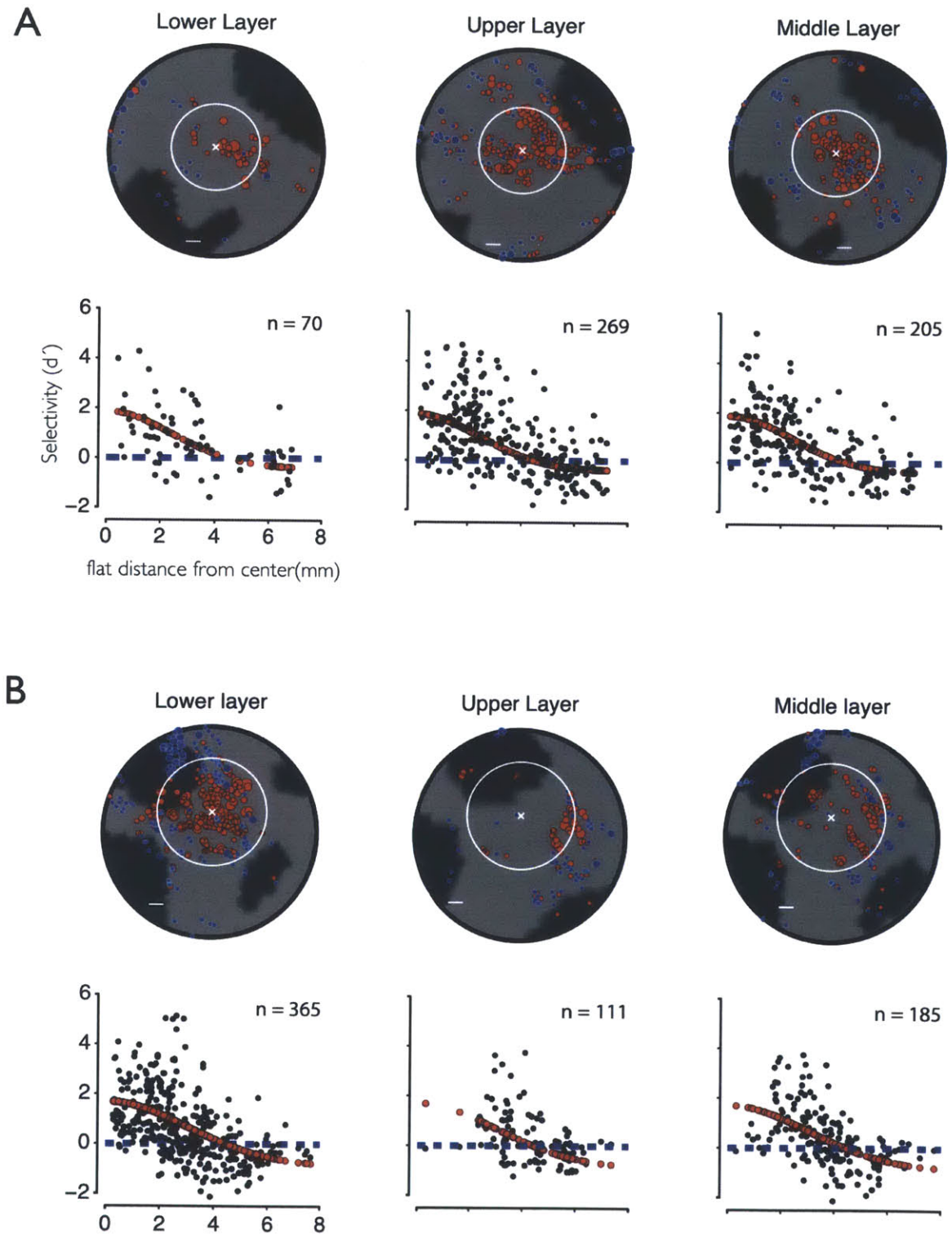

Figure 10. Upper and Lower regions of the cortical ribbon show similar category selectivity profiles. Dividing sites that could be reliably localized in the "upper" and "lower" layers from our data did not reveal structure different from the main results. We also reprojected sites to the middle layer using a more restrictive criteria for localization. Despite these changes, the average category selective signal in the pMFP did not change.

(k_{In} = -0.10, k_{Near} = -0.03, k_{Far} = 0.03). Finally, we found that the sites with the highest degree of face selectivity were only found in the MFP. For example, we found that ~24% of sites "in" the MFP (i.e. or about a quarter of the sites) had strong face selectivity ($d' > 2$). In contrast, we found no such sites in the "far" spatial region, and based on the distributional fits, we predict that less than 0.65% (i.e. approximately 6 out of every 1000 sites) of sites in the "far" region would have equally high face selectivity. In that sense, the inside of the pMFP is nearly 50 times more "face selective" than the outside of the pMFP. There is nearly a 50-fold enrichment for sites in the pMFP that have a high face selectivity (d'>2), and these sites constitute about a quarter of the sites in the pMFP.

We considered the possibility that projection to a 2D surface might hide 3D spatial structure (e.g. information from cortical layers). To examine this issue, we projected our data to two surfaces at two different depths in the estimated cortical ribbon underlying the MFP (Figure 10). Each recording site was projected to either an "upper" or "lower" surface exclusively. For comparison we also projected to the middle layer as was done in the original analysis, but with a similar projection restriction. While the data from first monkey seemed to predominantly have been sampled from the upper layers, and the data from second monkey was largely sampled from the lower layers, there was no evidence that either layer substantially differed from the mid layer model in terms of its selectivity for faces (Figure 10). The average category selectivity was approximately stable across the estimated "layers" in the MFP (Figure 11).

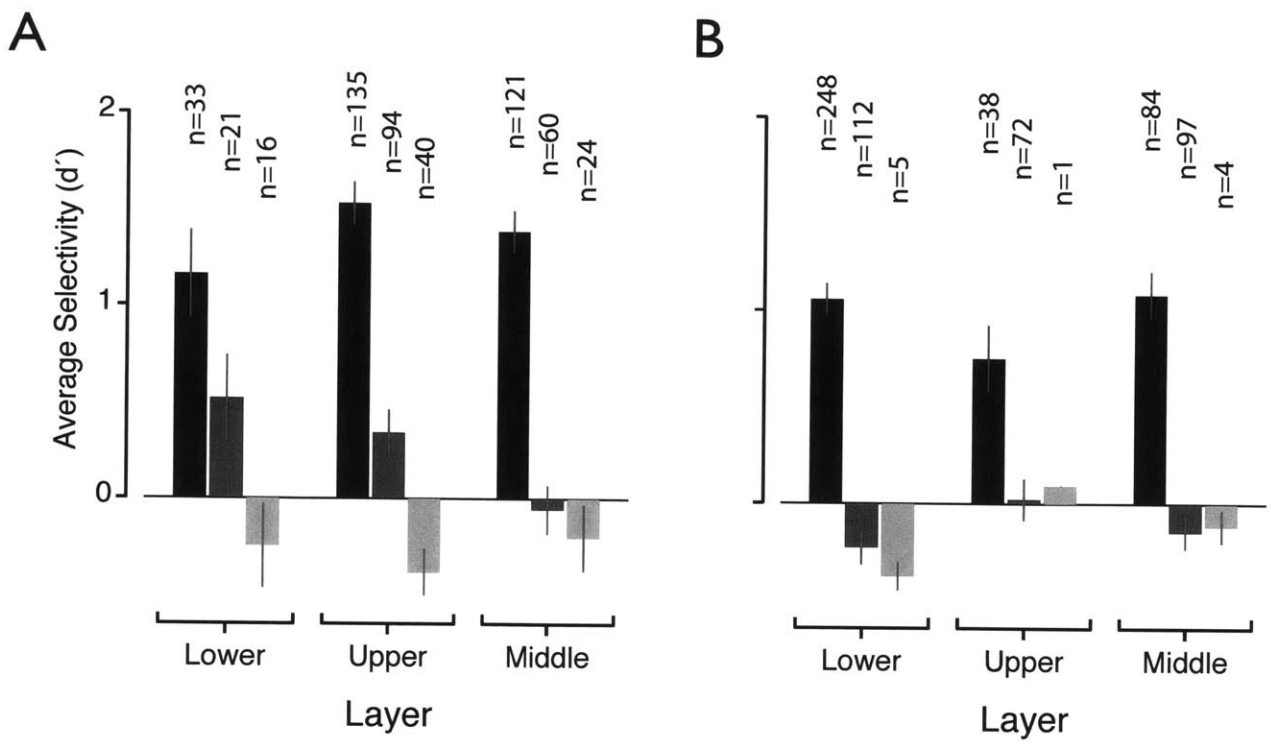

Figure 11. Upper and Lower layer estimates show similar average selectivity estimates. Separating sites by upper and lower layers produces similar selectivity estimates of the average selectivity in the "In", "Near", and "Far" regions (black, dark gray, and light gray bars respectively), in either subject (A: Monkey 1; B: Monkey 2).

2.4 Discussion

We analyzed the data from 1041 multiunit sites in two macaque monkeys in and around the fMRI identified middle face patch, primarily on the convexity of the STS in area CIT. We found evidence for a spatial region of cortex where a large fraction of sites responded preferentially (d'>0.65) to images of faces over non-face objects. Estimating the size of the MFP, we found that the enriched zone is best described as a ~6 mm area (on a 2D sheet) where the purity of face selective sites in the zone could range from a peak as high as 84% at the center and 5% for sites outside the MFP. The characterization depending somewhat upon the definition of "face selectivity" and the model used to estimate the location of the patch.

We choose to examine intuitive models for the spatial layout of the large scale face selective activations found with fMRI in non-human primates. Similar activations in humans have been used as evidence for domain specific modules - finite areas of cortical tissue where all of the neurons in the theoretically defined region are hypothesized to exclusively support face behaviors. Here, we choose to examine simple first order approximations of the concept by focusing on two aspects of the modular hypothesis: the boundary of the module and the purity of the category selective response. In a strong form, the modular hypothesis incorporates a definite spatial boundary between neurons inside and outside the module, where all of the neurons inside the patch are category selective (i.e. high purity). A weaker form of the hypothesis could be viewed as allowing for flexibility in terms of the boundary and purity of the category selective response in the module. Specifically, we examined a cylinder or module type model (i.e. MFP is defined by hard boundaries), and a gaussian model (category selectivity is more concentrated at a specific location and becomes less concentrated with distance from the center). Limiting the domain of the model space to three low spatial frequency models, we found that all of the models that we tested fit our data equally well; and were far from a perfect quantitative fit (i.e. r^2 observed was ~0.3). Examining the residuals to the model fits suggested that at any given spatial distance from the center of the MFP, a wide distribution of selectivity could be observed. Even inside the MFP, some sites were highly face selective and others showed inverted face selectivity -- preferring objects over faces. This wide range of neural selectivity was not explained by either the spiking response variability of neurons or the variance due to the limited set of image exemplars used in our study. Considering previous work demonstrating tuning for 3D head pose in face selective neurons (Desimone et al., 1984; Perrett et al., 1985; Freiwald et. al., 2010) one possibility might be that the high spatial frequency structure observed within the MFP in our data represents organization for some obvious real world image feature such as head pose,

as observed with optical imaging studies (Wang et al., 1996; 1998). Another possibility is that putative feature columns with tuning for an unknown shape dimension are more and less correlated with the images of faces used to probe the MFP (Tanaka, 2003; Tanaka et al., 1991).

Our modeling analysis suggests clustering of category selectivity on the order of 6 mm. This estimate is larger than previous estimates of physiological clustering in IT based on the selectivity to shape based image features. Previous studies have typically reported to be <1mm in spatial extent (Kreiman et al., 2006; Sato et al., 2009; Tanaka, 1996; 2003). Earlier reports of the spatial structure in IT have observed that neurons close enough together to be recorded simultaneously on the same electrode (<300 um) responded maximally to qualitatively similar stimuli (Fujita, Tanaka, Ito, and Cheng, 1992). Accounting for the angle the electrode made with the cortical surface, the same study estimated that cells recorded along a penetration had markedly different stimulus preferences when the lateral distance was estimated to be > 400 um. Experiments with intrinsic signal optical imaging found that several activity spots could be identified on the cortical surface in response to a single image. Images were arbitrarily simplified to maximize the response of a sample neuron in neural recordings that targeted the area prior to optical imaging. At least one of the observed spots included the location where the sample neuron was identified (Wang et al., 1996; 1998). The observation that several activity spots could be identified with a single image suggested that image features were distributed throughout IT and that similar features clustered in ~400 um "spots," or feature columns (Tanaka, 1996; 2003). While these initial studies were criticized for the arbitrary simplification procedure used to identify the columns, recent work using a moderately sized set of random images has found similar results. Sato and colleagues (2009) recorded single and multiunit spiking responses from optically identified spots in TE. They found that the correlation between the response of isolated single units

recorded from inside a spot was generally low to the same image set. In contrast, the same analysis between isolated single units and the average multiunit response within a spot, to the same set of images, was significant in the upper layers (I-IV). This was not true for single units and multiunits recorded from different spots (though the closest spots were > 600um apart). This result suggests that the response of isolated single units will be correlated to multiunit samples within a spot (i.e. at distances less than 600um). While this study did not explicitly target fMRI identified face patches, they found no differences in the properties of spots that were mainly activated to images of faces (over non-face distractors), suggesting shape based clustering of ~600um regardless of category.

Evidence of category specific spatial clustering for faces has been observed with both single unit electrophysiology and optical imaging. Perret and colleagues (1984) found evidence for small scale clustering of putative face cells by head pose on the order of 1-2mm. Neurons that were tuned for at least one view of the head (or face) were more likely to be observed along the same penetration if a neuron with similar tuning had already been recorded from that penetration. Additionally, penetrations that contained large numbers of face preferring cells were often estimated to be within 2mm of each other, suggesting spatial clustering along the cortical surface for face selective cells. While the spatial extent reported in this study is larger than the putative feature columns suggested by optical imaging, the measurement error associated with the spatial localization of their recording sites was not estimated. Estimation error could in principle account for the discrepancy observed between previous studies. However, one optical imaging study with single face exemplars did find spatial structure in the activation to specific views of the head (e.g. face), ranging from profile to frontal plane (Wang et al., 1996; 1998). Critically, the activated spots for each view were located adjacent, and occasionally overlapping each other, forming a systematic map on the

cortical surface when the image of the head was rotated in depth from left to right. Taken together, these activations suggest that category selectivity organized by image view along the cortex can extend up to 1-1.5mm. In contrast to these studies, our data provides evidence for the clustering of face selective neural responses of approximately ~6mm. A critical element for future studies to distinguish is the extent to which other encoding hypothesis can fit the spatial extent of the enrichment observed here.

The computational methods used in our study simplified the spatial geometry of the cortical tissue near the fMRI identified face selective region by projecting the depth dimension to a 2D cortical sheet. Therefore, one possible aspect of cortical organization that is not well covered by our data is the laminar distribution of selectivity in the MFP region. While we found no evidence for differences between lower and upper segments of the cortical ribbon, the spatial resolution of our X-ray system is limited in the ability to generate reliable spatial separation of sites at this level of resolution. Sites sampled in the two layers were not sampled equally in the two animals. Subject 1 was mostly sampled from the upper layers, while Subject 2 was mostly sampled from the lower layers (figure 6). This was likely due to differences in the data acquisition procedures between the 2 subjects. Despite the spatial sampling bias between the two animals, the distribution of category selectivity between the two animals was similar, and provided no support for differences between the upper and lower layers in our data set. Laminar electrodes localized in crucial areas around the MFP (I.e. "in" , "near", and "far") might provide some insight into the laminar dynamics of the MFP.

We detail a systematic study on the spatial organization of category selective responses covering the cortical tissue localized to the fMRI identified middle face patch. We sampled over 1000 multiunit sites in a 14mm area of central IT cortex in 2 monkeys. While previous studies have sampled the neural responses in the MFP, our work is

different because we were able to broadly sample the fMRI identified region with a relatively high degree of spatial precision (e.g. in comparison to previous efforts). This was possible due to the use of a custom designed X-ray imaging system that allowed us to estimate the 3D position of each recording site, and the use of 3D cortical models of the subject's brain to co-register recording sites to an anatomical reference. Previous studies have relied on geometrical projections from structural MRI. The spatial error inherit in this projection procedure makes fine resolution sampling of the cortex nearly impossible at the spatial scales of interest (1 mm or less). By sampling a large area of cortex we were able to examine a number of models that attempted to describe the spatial structure of the category selective signal for faces at the MFP.

In summary, we found that the cortical tissue localized to the MFP contains an enrichment of sites with a significant category selective response for faces, as reported previously (Bell et al., 2011; Tsao et al., 2006). In addition, we were able to estimate that the enriched category selective area extends ~6 mm on the cortical surface. While previous reports had mixed observations on the purity of the fMRI identified region (ranging from as high as 97% to as low as 41%), we reconciled those results by showing that peak purity estimates could approach 84% at the center of the MFP, while the number of face selective sites gradually fell to ~5% outside the MFP. Overall, we report that the average purity in the estimated spatial envelope of the MFP is ~67%. Our results suggest that spatial error in sampling from the face selective region could explain the discrepancy observed between previous studies. We also extend previous studies by showing comparative measures of face selectivity both inside, and outside the MFP using the same methods and stimuli. Further, we observed a common distributional form for face selectivity measured in neural populations "in" and "far" from the MFP. Neural populations "in" the MFP could be characterized by the high incidence (50x) of high category selective sites. Finally, our data provided evidence for

the presence of high spatial frequency structure in the MFP; structure not predicted by simple modular (e.g. low spatial frequency) hypothesis for face areas.

2.5 Table 1

Model Type	Model Parameters	Best Fit M1	r^2 M1	Best Fit M2	r^2 M2
"Box Car"	Selectivity Weights: [W_{in}, W_{out}]; 2D center [x0, y0]; radius [r]	[1.46, -0.09] [0.01, 0.13] 2.98 mm	0.32	[1.13, -0.14] [1.56, -0.94] 3.25 mm	0.26
Isotropic Gaussian	Selectivity Weights: [W_{in}, W_{out}]; 2D center [x0, y0]; fwhm [$2\sigma\sqrt{2\ln(2)}$]	[2.27, -0.40] [-0.09, 0.24] 5.70 mm	0.33	[2.58, -0.88] [1.33, 0.01] 7.10 mm	0.27
Full Gaussian	Selectivity Weight: [W_{in}, W_{out}]; 2D center [x0, y0]; $fwhm_x$ [$2\sigma_x\sqrt{2\ln(2)}$]; $fwhm_y$ [$2\sigma_y\sqrt{2\ln(2)}$]; angle [θ]	[2.45, -0.58] [0.07, 0.66] 8.90 mm 5.10 mm 92°	0.31	[2.35, -0.55] [1.58, -0.03] 8.20 mm 4.40 mm 99°	0.33

2.6 Figure Legends

Figure 1. MION enhanced fMRI localization of face selective patches in the Macaque Temporal Lobe. Conventional images (A) of macaque faces and familiar everyday objects used to localize areas in the Inferior Temporal cortex that responded selectively to images of faces. Three discrete patches (B) were found on the convexity of the Superior Temporal Sulcus in inflated and flattened surface models of the Temporal lobe. We examined the fMRI face selective signal in the right (M1) and left (M2) hemispheres of the two subjects. (C) The Middle Face Patch was observed consistently across both animals on the convexity of the STS (i.e crown of the Superior Temporal Gyrus). Right and left hemispheres shown in coronal slices.

Figure 2. Spatial registration and analysis methods. The 3D spatial locations of all sampled in the MFP region were estimated with a custom built stereoscopic X-ray imaging system. The 3D locations were registered (A: Left) to a high resolution (500um) anatomical MRI. To improve the overall accuracy of the registration (A: inset), non-linear registration using FMRIB's FNIRT registration tool was performed (A: Right), and the resulting transform was applied to the estimated positions of the recording sites (A: Inset). The spatial position of each recording site was projected to the closest orthogonal node of a high resolution mid layer mesh of the cortical surface (B: Left). Sites that moved a distance greater than 1250um from their original 3D position to the 2D surface manifold were excluded (B: Right) from further analysis. An arbitrary portion of the Temporal lobe was selected for all spatial analysis (C: Left). This area was selected by choosing the approximate center of the fMRI MFP activation in each monkey and including all nodes within an arbitrary geodesic radius (7mm). The

radius was chosen to be approximately the greatest distance that would not extend into face selective activation at posterior (PL) or anterior (AL) locations.

Figure 3. The physiologically defined MFP (pMFP) is a region of cortex with an enrichment of category selective sites. (A) Flattened 2D regions of the Temporal Lobe highlight category preference (red = faces; blue = objects) across the spatial region of the fMRI defined MFP. The size of the circle indicates the d' value at each recorded site. (B) Category preference is indicated by color as above, but yellow indicates sites where the 95% confidence interval was not greater than the estimated 2x criteria for face and object selectivity.

Figure 4. 2D Models of the category selective spatial region. Three models for the spatial structure of the pMFP were used to fit the category selectivity and 2D spatial position over all multiunit samples. Column 1 summarizes each model and their parameters (see methods). The 2D schematic depicts the range of estimated selectivity values as a function of 1D spatial position and the two linear estimators for the model, while the parameters for the non-linear predictor formulas are described in the adjacent diagram. Columns 2 and 3 summarize the results in each subject. Flattened maps displaying the data are depicted as well as the best fit model. The selectivity at each site was collapsed across radial distance to the estimated center of each patch (black dots) alongside the model prediction for each site (red dots). The number of standard deviations are plotted along the abisca for the full gaussian model as this isocontours of this model were not radially symmetric.

Figure 5. Low frequency 2D spatial models of the pMFP are largely consistent. The spatial location and category preference for each recording site is localized on 2D flattened surfaces of the fMRI identified MFP region. The different best fit models from our analysis are largely consistent on the position of the center and the spatial extent of the enriched spatial region.

Figure 6. Variance in selectivity estimated at nearby spatial locations is greater than expected in the MFP. The bar graph depicts that average squared error between face selectivity estimated at sites <500 um in distance from each other, recorded from the same electrode. The average variance expected due to the limited number of image exemplars or trials is also depicted for comparison. Errorbars indicate the 95% confidence interval on estimate of the average error.

Figure 7. Purity estimates in the pMFP are sensitive to the metric and spatial extent used to define the region. Estimates of the number of category selective multiunit sites in the MFP (the purity) could range from 94% to 58% depending on the distance from the center of the patch. These ranges varied due to both the model and metric used to define the patch. Estimates using FSI tend to increase the purity in and out of the patch. Overall, the purity falls off gradually from the center of the patch to cortical regions outside the patch.

Figure 8. Estimated purity across subjects is largely consistent. We estimated the purity collapsed across both animals using the estimated center form the isotropic Gaussian spatial model. The empirical purity function across both animals is

remarkably consistent in shape across different selectivity estimators. Lines show d' estimates of the purity across different thresholds (Line width, thicker to lighter corresponds to higher to lower thresholds) for category selective inclusion (i.e. how a face site is defined). A similar shape was observed with the FSI index that has been used in previous studies (Tsao et al 2006; 2008; Friewald et al 2009; 2010). Not surprisingly, weaker thresholds resulted in higher estimates of the purity near the center and far outside the pMFP.

Figure 9. Category selective signals in and out of the pMFP. Distributions of the estimated face selectivity from multiunit samples in and out of the pMFP can be fit to similar distributional forms (see text), but are mean shifted and differ in their variability. Units outside the pMFP appear to occupy a narrow range of category selective values, with the majority of values near zero, while sites in the pMFP are characterized by the broad range of selectivity and high positive values.

Figure 10. Upper and Lower regions of the cortical ribbon show similar category selectivity profiles. Dividing sites that could be reliably localized in the "upper" and "lower" layers from our data did not reveal structure different from the main results. We also reprojected sites to the middle layer using a more restrictive criteria for localization. Despite these changes, the average category selective signal in the pMFP did not change.

Figure 11. Upper and Lower layer estimates show similar average selectivity estimates. Separating sites by upper and lower layers produces similar selectivity

estimates of the average selectivity in the "In", "Near", and "Far" regions (black, dark gray, and light gray bars respectively), in either subject (A: Monkey 1; B: Monkey 2).

Table 1. Model parameters. Each model and the parameters estimated for each subject are tabulated.

Chapter 3

Face Detection in the Macaque Middle Face Patch

3.1 Introduction

Patches of cortical tissue that respond preferentially to images of faces have been localized with fMRI in the temporal lobe of macaques across multiple laboratories (Bell et al., 2009; Logothetis et al., 1999; Pinsk et al., 2005; Tsao et al., 2003). While the number, size and anatomical location of these patches can vary between subjects (Pinsk et al., 2009), the largest consistently reported patch is the middle face patch (MFP). Typically observed on the convexity of the superior temporal sulcus in posterior TE, the MFP has been variously identified as the posterior patch (Bell et al., 2009; Pinsk et al., 2005), and area ML by Tsao and colleagues (2008), who reported the average volume of the area in a sample of 9 subjects to be ~70 mm^3. The MFP has been speculated to be homologous to the human Fusiform Face Area (FFA) based on categorical selectivity preference, spatial location, and psychological models of face processing (Nasr et al., 2011; Tsao et al., 2003; 2008; Tsao and Livingstone, 2008; but see Ku et al., 2011).

Neurophysiological recordings targeted to the MFP have reported that nearly every (>97%) visually responsive neuron sampled was face selective (Tsao et al., 2006). Note,

however, that another study reported a much lower fraction of face selective neurons (Bell et al., 2011). The observation of a large region of posterior IT cortex where essentially all neurons are reported to be selective for faces is remarkable because previous work in posterior Inferior Temporal cortex (IT) suggests an encoding of visual shape based on geometric features (Brincat and Connor, 2004; Tanaka, 1996), rather than semantic category. As with previous studies of face preferring cells in IT (Desimone et al., 1984; Perrett et al., 1984; 1985; Yamane et al., 1988), MFP neurons exhibit tuning to a narrow range of poses when the head is rotated in depth, and are sensitive to individual features (e.g. mouth shape, eye size) and their spatial combination (Freiwald et al., 2009; Freiwald and Tsao, 2010).

In the previous chapter, the spatial characteristics of the macaque middle face patch were explicitly examined. Using a novel X-ray imaging system, the spatial area identified by fMRI to be category selective for faces was targeted for neurophysiological recording. Similar to previous studies, we observed an enrichment of sites that responded preferentially when subjects viewed images of faces over non-face distractors. Our data provided evidence for an area, ~6 mm in diameter, where category selective sites had a higher probability of being present than outside the enriched area. While we utilized images similar to what has been previously used to examine the face selective signal identified by fMRI, we did not explicitly examine the category assumption itself. A prevailing view in the literature is that the MFP acts as a detector, or "gateway," for a larger face processing network. By signaling the presence of a face in the visual input, downstream face patches can evaluate more specific attributes of the image, such as identity, emotional state, gender, etc. (Tsao and Livingstone, 2008). However, a more cautious interpretation may be warranted. Previous ('conventional') image sets comprised a limited set of face and non-face exemplars. Such an image collection only sparsely samples a high dimensional image

space, and does not probe the boundaries in the image space of when an image is perceived as a face. Thus, previous studies have produced only a limited characterization of genuine face selectivity, and the existing data are consistent with the alternative interpretation that MFP neurons, like the rest of posterior IT, respond to simpler geometric features that might only be correlated with the semantic identification of a faces (e.g. some other shape based non-semantic feature). To examine this issue, we compiled a set of natural images that spanned a spectrum of semblance to human faces using the output of a computational face recognition system (fig 1, experimental image set highlighted in grey; see Methods). Critically, human observers accurately classified the true face images correctly (>96% accurate), but rarely identified a face in any of the distractor images (<10%). We assessed neural spiking responses in the MFP, with the goal of determining how many sites had categorical responses for faces with our experimental image set. Consistent with previous work (Tsao et al., 2006), we found that many sites in the MFP responded strongly to images of faces when tested with a conventional image set. However, using an image set chosen to be more challenging for detection based on shape based features (yet trivial for human subjects to recognize) we found only a weak correlation between the responses of neural sites localized to the MFP and human face detection behavior.

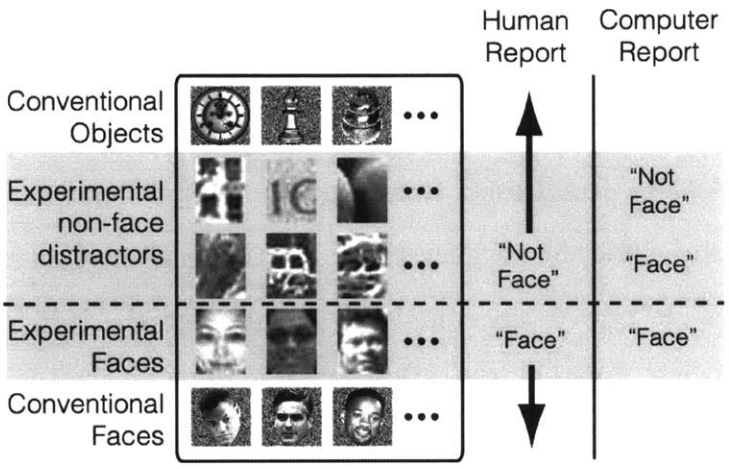

Figure 1. Experimental design. Images conventionally used to examine face selectivity in the MFP have minimal incidental image variability, consisting mainly of frontal views of faces cropped to be similar in spatial envelope. We use a novel image set derived from the output of a computational face detection system and validated by human subjects.

3.2 Materials and Methods

3.2.1 Data Acquisition

Subjects: One male *macaca mulatta* was prepared for MION enhanced fMRI and multiunit neurophysiology as described previously (Issa and DiCarlo, 2012). All procedures we performed according to guidelines from the National Institutes of Health and approved by the M.I.T. Committee on Animal Care.

Awake functional imaging: A plastic headpost was attached to the subject's skull under aseptic surgery conditions. The animal was trained for awake functional imaging as described previously (Vanduffel et al., 2001). Images were presented for 250ms (inter-stimulus interval = 500ms) at 5° in the fovea, randomly jittered on each trial (+/-2 deg in both azimuth and elevation, uniform distribution). Eye movements were monitored with an optical ISCAN system (ISCAN Inc., Woburn, MA), time points where the animal broke fixation for > 250ms were excluded from further analysis. Images were presented randomly in blocks of categories (faces, bodies, places, objects, and scrambled faces, 20/category). Analysis of the functional imaging data was conducted using the FS-FAST toolbox (http://surfer.nmr.mgh.harvard.edu/fswiki/FsFast) and custom written scripts in MATLAB (Mathworks, Natick MA). Anatomical models of the brain were created using the CARET software package (Van Essen et al., 2001).

MION enhanced functional imaging was conducted on a 3T Siemens Trio Tim at the Athinoula-Martinos Imaging Center at MIT. Functional data (TR 3.2s, 46 slices, 1.25 mm isotropic voxels with a 10% slice gap) was collected with a single loop surface coil. Image presentation and water reward was controlled with MWorks software (http://mworks-project.org). Functional data was motion corrected within session and co-

registered to the animals anatomical model. Field scans were used to correct magnetic field distortions with FSLs FUGUE software (Jenkinson et al., 2012).

Electrophysiology: The animal was prepared for awake neural recording by the placement of a plastic recording well (18" diameter; Crist Instruments) targeted to the MFP (Horsley-Clarke center AP +3mm, 9 degree angle). The animal was then trained to sit an upright position and fixate in a 2-3 degree response window while viewing images in a Rapid Serial Visual Presentation sequence (7-10 images/trial; 300ms on 300ms off). Images were presented at 7° on a background of random white noise in the fovea. Eye movement traces were monitored with an optical EYELINK system (SR Research; Ontario, Canada). Trials where the animal broke fixation were aborted, and only images presented before the eye movement were considered. The first image in an RSVP sequence was always disregarded from further analysis.

Multiunit recording was conducted with single, glass coated tungsten microelectrodes (0.5-0.7 MΩ: Alpha-Omega Co., Alpharetta, GA), amplified with a BAK system (BAK Electronics, Mount Airy, MD). Neural signals were sampled at 14Khz and bandpass filtered with an inline butterworth filter (Krohn-Hite, Brokton MA) between 300Hz and 7KHz. Experimental stimulus and reward control, data recording and storage were managed by custom software developed for this purpose (MWorks, http://mworks-project.org). Well grids (straight and angular) were used to position a metal guide tube 5-6mm from the STS. Microelectrodes were advanced (hydraulic microdrive, Crist Instruments) to the STS dorsally. Each day 3-6 sites were recording every 300-600um further along the penetration from the STS. The precise electrode position for every site sampled was determined using a custom X-ray system and projected to a 3D anatomical model (for details on this method see Cox et al., 2008).

3.2.2 Spatial Localization

To localize the MFP, we used a separate mapping protocol and made recordings from 481 sites localized around the fMRI identified MFP. Every site was tested with a conventional set of images containing faces and non-face objects (125 images). Given the 3D coordinate and estimated selectivity of each recording site (d' faces > objects), a sphere was fit to the 3D data that maximized the (face vs. non-face object) selectivity of sites in the sphere. The center of the sphere was considered an estimate of the spatial location for the center of the MFP. Using a previously established criteria (i.e.sites in the patch responded 2x on average to images of faces than to non-face images; Tsao et al., 2006), we determined that locations within 3mm of the estimated center could be established as residing within the MFP. All of the sites analyzed in this study also resided within the pMFP region identified in chapter 2.

3.2.3 Task

Experimental Protocol: 114 of the 481 sites used to estimate the location of the MFP were also tested with our experimental protocol (see chapter 2). Of the 114 sites, 45 were: 1) localized to the estimated MFP (see previous section), and 2) selective by a d' > 0.65 criteria to the conventional images interleaved with the experimental image set (average $d'_{\text{faces > objects}} = 1.65$; range = [0.68, 4.34]).

Images: Images of face and non-face exemplars were obtained from a computational face detection system (Schneiderman, 2004). The image set consisted of the actual face images the system identified as faces ('true hits'), non-face images the system called faces ('false-alarms'), and random patches from image areas the system rejected as containing a face ('true misses"). These images were then evaluated in a face detection task by 18 human subjects (Meng, Cherian, Singal, and Sinha, 2012). Briefly, each image

was presented at the center of gaze for 300ms (subtending 5 deg visual angle). Subjects indicated whether or not the image contained a face. True-hits were almost always detected as faces (~96%), while the non-face images were rarely identified as containing a face ('false alarms' ~8%, and 'random patches' ~2%). This well characterized image set was used to probe face selectivity in the MFP. Conventional images (e.g. frontal views of cut-out faces and objects) were also interleaved with the experimental images. A number of different protocols, each differing in the exact number of exemplars presented from each category, were used in our experiments: these ranged from 20-60 random images; 90-180 false alarms; 15-60 true hits; 5-30 conventional faces, 5-30 conventional non-face. Each image was presented 8-12 times each. To ensure that these choices did not affect our overall result, we conducted our analysis with the response to all of the images collected on a site by site basis, and on a subset of the images common to all protocols at every site. Neither of these choices significantly affected our main results.

3.2.4 Analysis

Multiunit neural signals: The multiunit response to an image on a single trial was estimated as the spike count rate in a fixed window 60-160 ms post stimulus onset, minus the baseline (average response 0-50 ms post stimulus onset over all images). In contrast to the analysis in chapter 2, the mean response to each exemplar over all trials repetitions was than used to estimate the average response to each image (e.g. averaging over repetition variance). To denote the average response for a site (as depicted in figures 2-4) to a particular image, we normalized the response to each image by the standard deviation for that site. Since we were interested in characterizing the response to each image, averaging over repetitions provided a more robust estimate for each exemplar (though it should be noted that variation in the category selectivity

estimate, which were not the focus of this study, directly depend on the number of repetitions to each exemplar). The estimated response of each image and the standard deviation between the response to all of the exemplars in each category (faces or objects) was used to estimate the category selectivity at each site as d':

$$d' = \frac{\bar{X}_{faces} - \bar{X}_{objects}}{\sqrt{\frac{\left(\sigma^2_{faces} + \sigma^2_{objects}\right)}{2}}}$$

The correlation between human psychophysical performance on a face detection task (y) and the multiunit response to each image (x) was calculated as the Spearman Rank correlation (ϱ_{xy}). The correlation was further disattenuated (Spearman, 1904), or corrected for variability that cannot be explained (a.k.a. "noise"; response variation measured over repeated presentations of the same image). The correction was obtained by dividing the raw Spearman correlation (ϱ_{xy}) by an estimate of the reliability of the neural and behavioral data.

$$\rho_N = \frac{\rho_{xy}}{\sqrt{r_{xx} r_{yy}}}$$

The goal of the correction is simply to normalize the correlation computed at each site so that it will reach a value of $\varrho_N = 1.0$ if human judgements perfectly explain neural responses at that site, given the noise levels of the neuronal and behavioral data. The reliability of the neural (r_{xx}) and behavioral data (r_{yy}) for a given signal (w) was estimated by split half correlations (ϱ_{ww}) and corrected by the Spearman-Brown prediction formula (Spearman, 1910):

$$r_{ww} = \frac{N\rho_{ww}}{1 + (N-1)\rho_{ww}}$$

The maximum expected correlation that can be obtained from the signal of interest r_{ww} (i.e. the reliability of the neural or behavioral data) is estimated by taking the average correlation between multiple (n = 1000) draws of N = 2 random splits of the data (ϱ_{ww}) and applying the correction. The Spearman-brown prediction formula is needed because the estimation is made with only half of the data, and is therefore an underestimation of the true reliability.

Single unit neural signals: While we did not explicitly set out to record the activity of isolated single units, we were nevertheless able to sort 85 units from our sample of 45 multiunit sites. We only considered units that fired more than 300 spikes and had a split half reliability correlation across trials greater than 0.3. This resulted in a pool of 42 sorted units. A majority (39/42) of the sorted units from our sample were selective for faces ($d' > 0.65$) when tested with the conventional images of faces and objects used in our experimental protocol. To insure that our results were robust to the quality of the isolation, we iterated our experimental analysis with subsamples of our single-unit pool, where the threshold SNR used to include putative sorted units in the analysis was allowed to vary as a free parameter.

3.3 Results

By definition, sites that exhibit true category selectivity for faces (i.e. putative "face cells"; Tsao and Livingstone, 2008) should respond greater to images that subjects readily identify as a face, than to those similarly identified as non-face distractors. As described above, we examined 45 putative "face sites" in the MFP (Tsao et al., 2006), in that these sites strongly preferred faces over non-face objects when tested with a protocol comparable to that used in previous studies (i.e. "conventional" images). However we found that many of these 45 MFP sites responded weakly on average to

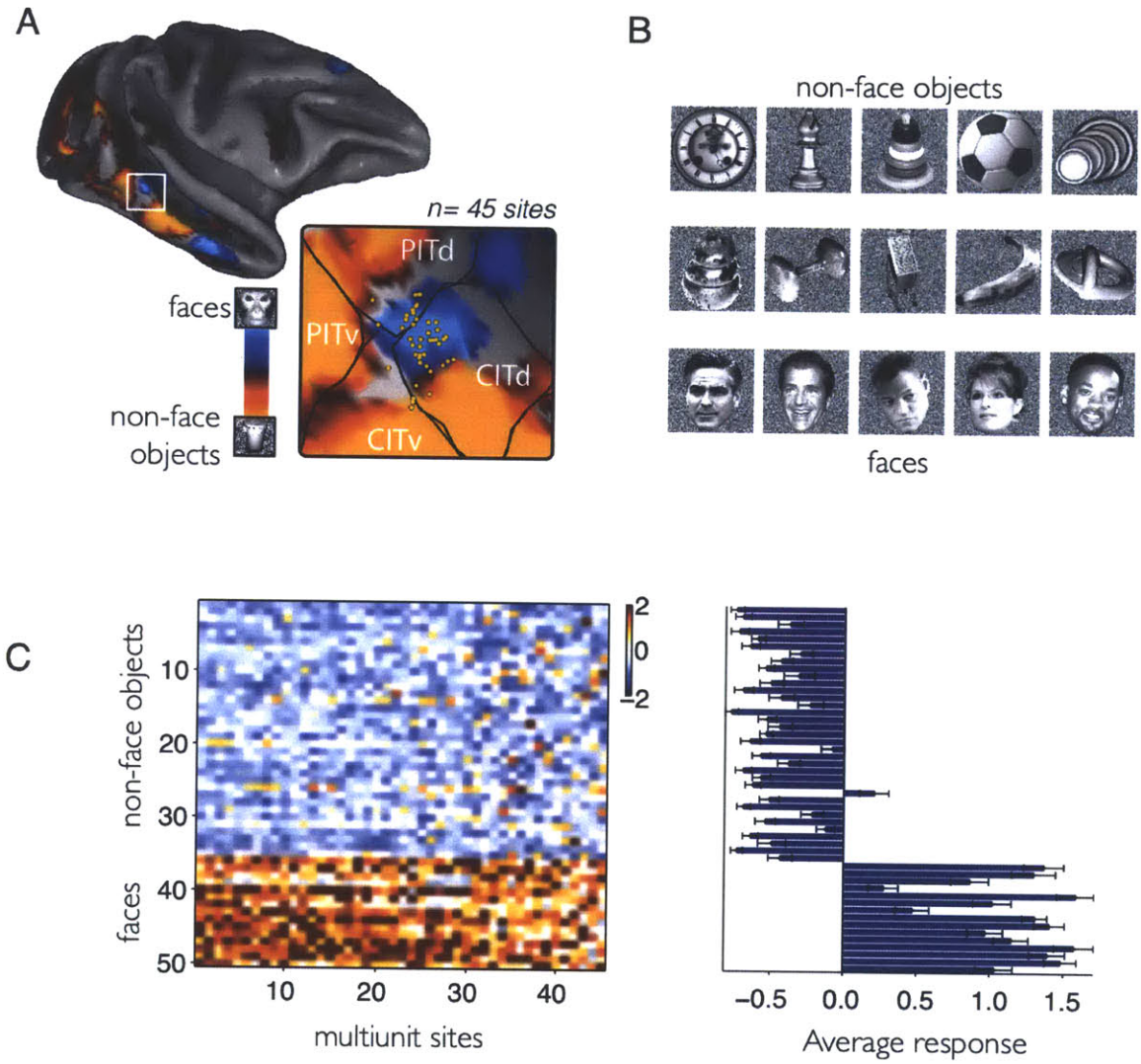

Figure 2. **Sites in the MFP selective for human faces.** **A**, The MFP was localized using MION enhanced fMRI. X-ray localization (see methods) was used to estimate the 3D location of recorded sites (inset shows sites projected to a flattened 2D sheet). **B,** Images of face and non-face distractors used in X-ray targeted electrophysiology experiments to screen for the face selective sites used in the study. **C,** Selected multiunit sites under examination responded selectively to faces from our conventional image set. Average response to each image is normalized by the standard deviation at each site.

face images from our experimental protocol, even when human subjects unequivocally identified these images as faces (fig 2A: average $d'_{\text{faces>objects}} = 0.23$). A site by site analysis revealed that many of the sites reliably responded more to some images that human observers almost never called a face while responding weakly to other images

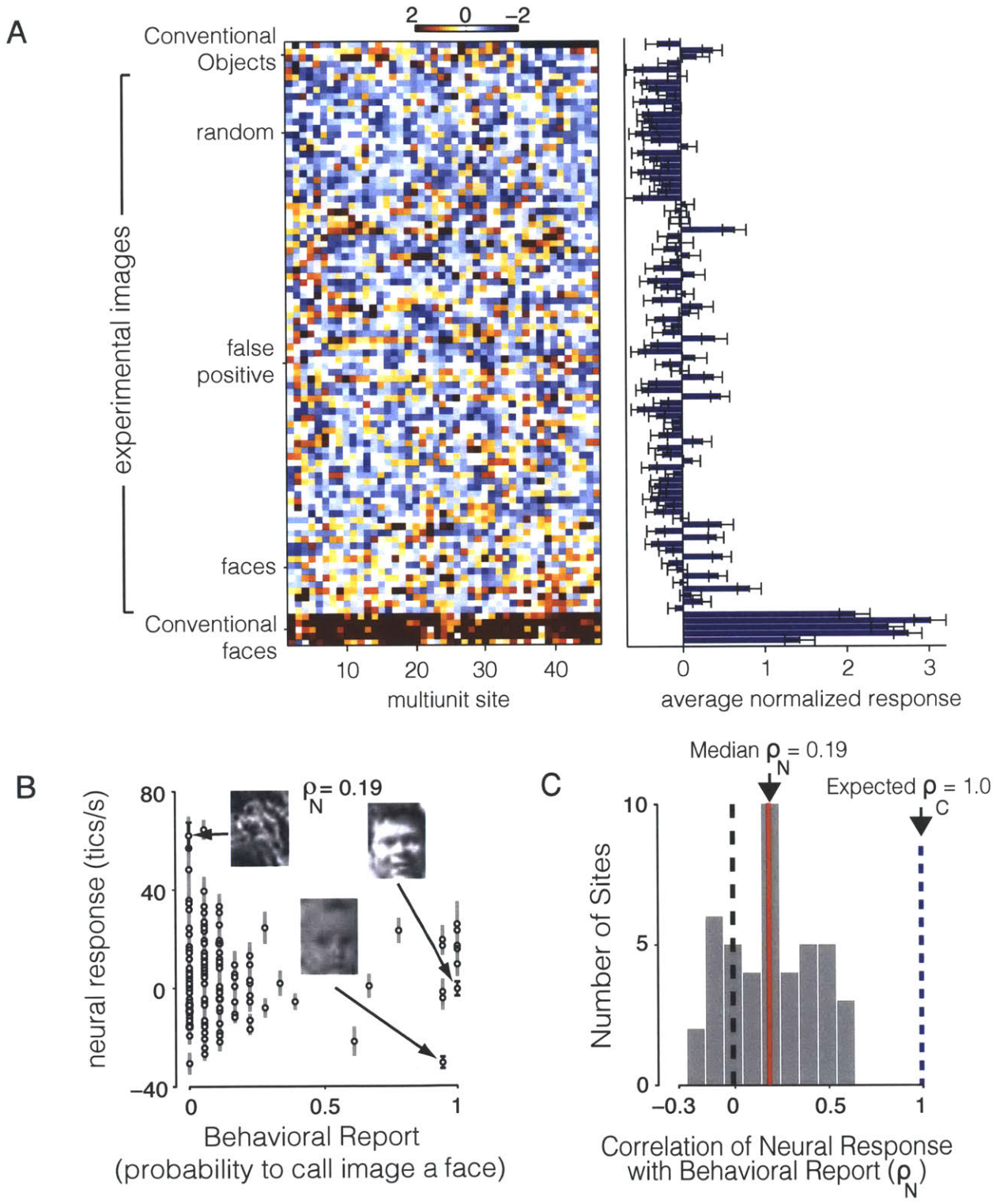

Figure 3. Face selective sites with conventional images correlate only weakly with human face detection behavior. **A**, The average response from each multiunit site to images common to all experimental protocols. **B**, An example site whose neural to behavioral correlation was near the median for the distribution ($p = 0.19$). **C**, The distribution of corrected correlation coefficients between the multiunit spiking and the probability that human subjects rated the same image as containing a face.

that subjects nearly always called a face (fig 2B, arrows), violating an explicit prediction of the face detector hypothesis. To quantify this, we computed the correlation between the response of each neural site and the probability that human subjects would identify the image as containing a face (corrected for noise, see methods). The normalized median site by site correlation ($\rho_N = 0.19$) for our sample was well below what would be expected given a strong face detector hypothesis (ρ_N expected = 1.0). This failure cannot simply be explained by response variability (a.k.a. "noise") as the normalization fully accounts for such variability. The distribution of correlation coefficients ranged from ρ_N = -0.20 to 0.70. Because not all sites were tested with the exact same set of images we reanalyzed our data using only images that were tested at all sites (see methods: Images) which resulted in a similar distribution of correlation coefficients (Median ρ_N = 0.26, range = [-0.10, 0.69]).

One explanation for our results could be that individual neurons in the MFP have unique response profiles, so that the multiunit signal might average out the true face selectivity exhibited by signal neurons. To examine this possibility, we sorted our multiunit data and re-performed our analysis. Figure 3A displays the spiking response to sample images from an example unit that responded strongly to the conventional face images, but only weakly to face images from the experimental set. This behavior was common in our sample of sorted units (fig 3B). To ask if the results of the correlation analysis (Fig. 2b) changed for single-units, we re-performed the analysis with the sorted unit pool. Because there is no universally accepted definition of the signal quality required to deem a unit as a "single unit," we looked at pools of single units with increasingly stringent isolation criteria (SNR, see methods). We observed that the median corrected correlation between the neuronal response and human face detection behavior remained low even with the highest quality single units (fig 3C).

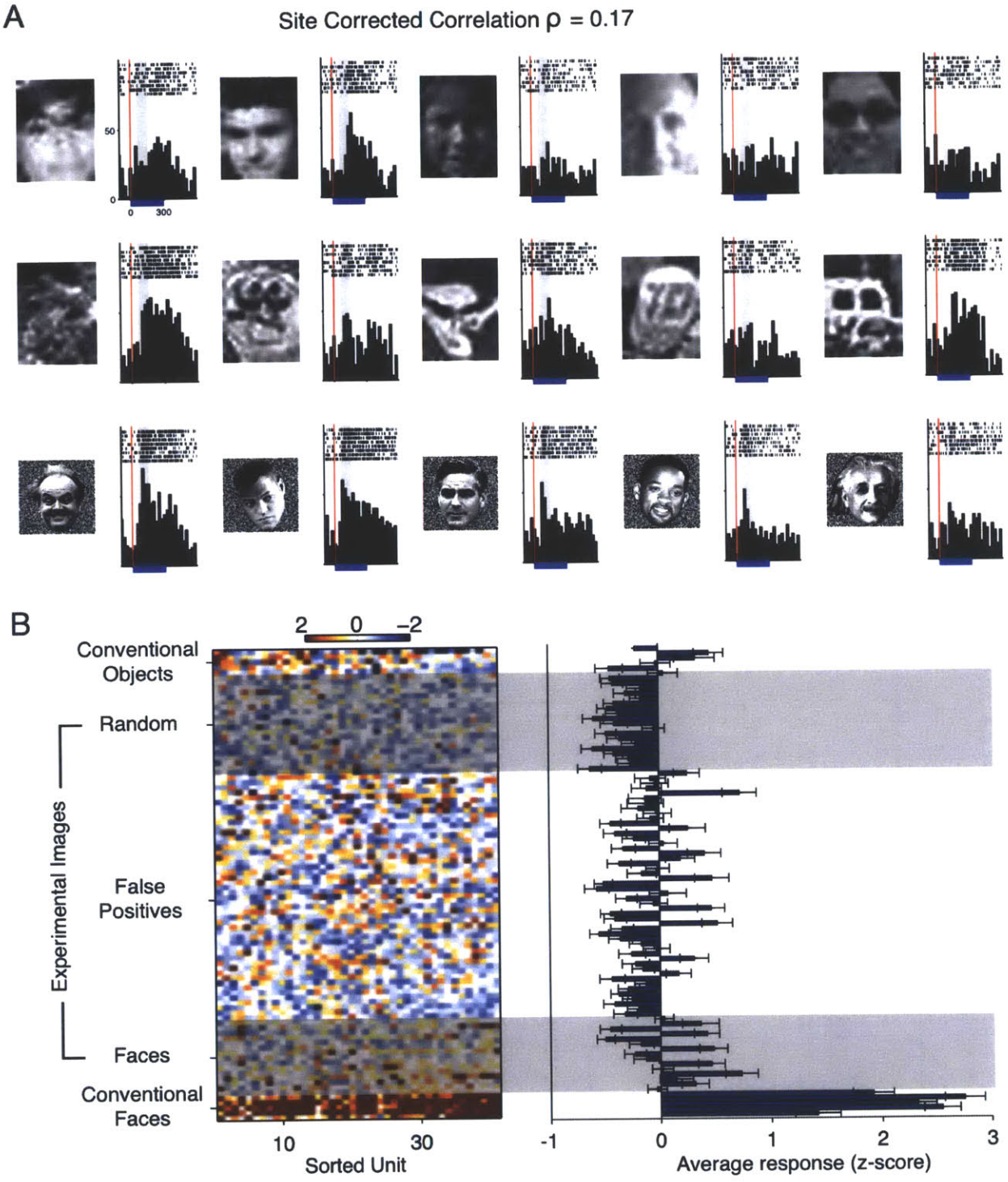

Figure 4. Sorted single units display similar category selective responses. **A**, Response to the 5 best exemplars (highest d') from three image categories for an example sorted unit. **B**, The average response from each isolated unit to the images common to all protocols.

Another possible explanation for our result is that categorical selectivity at MFP sites

develops over time, and that the response window in our analysis poorly captured the true face selectivity of the MFP sites. Previous studies have demonstrated an early and late peak in face selectivity with the LFP signal in the MFP (Tsao et al., 2006). This activity was speculated to be due to an early, feed forward response, and a later feedback response. To examine this possibility we reanalyzed our multiunit data using a sliding 50ms window to calculate the average corrected correlation over time. We observed a time series profile with a single major peak after stimulus onset that slowly returned to baseline. The maximum correlation at the single peak was $\rho_N = 0.21$ which

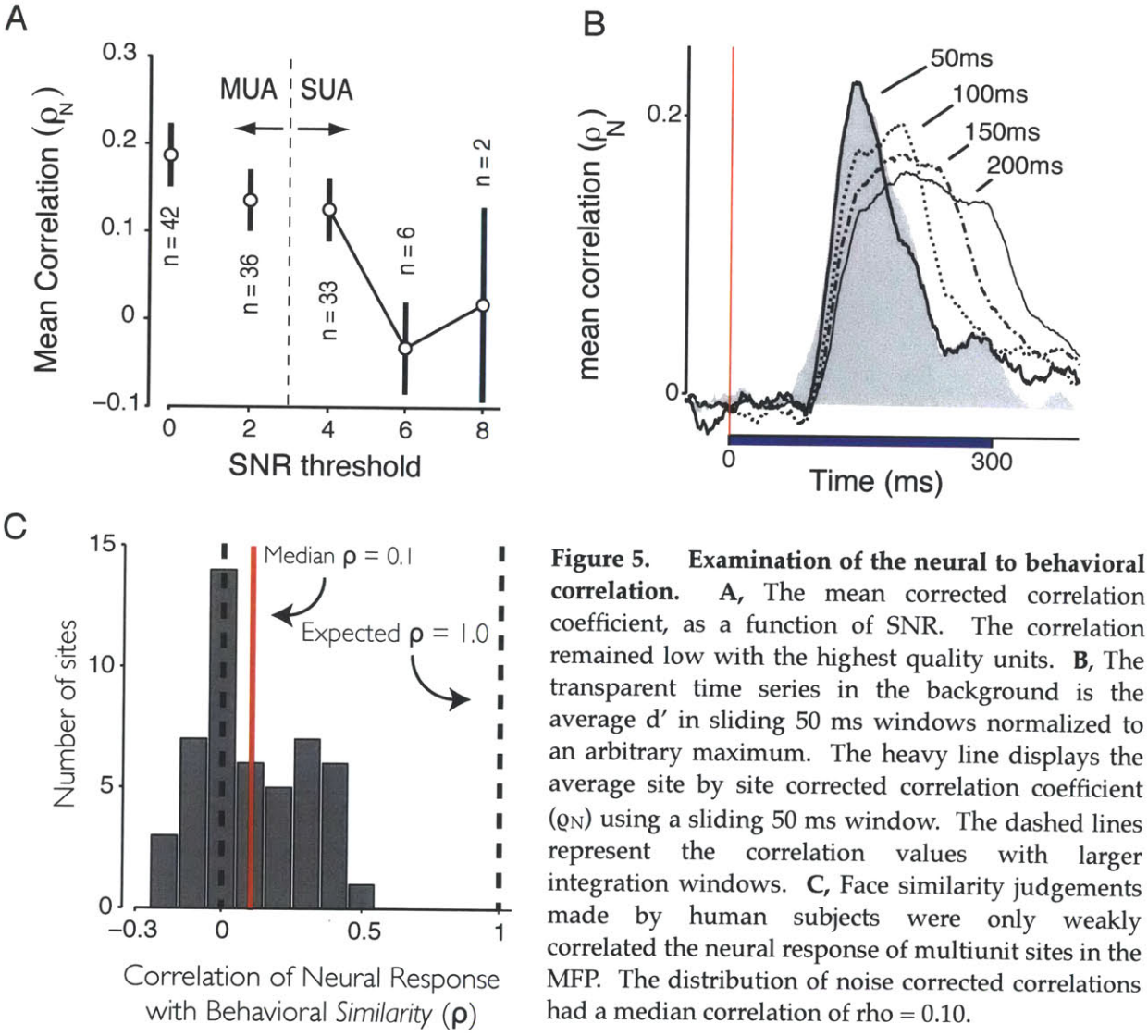

Figure 5. Examination of the neural to behavioral correlation. **A,** The mean corrected correlation coefficient, as a function of SNR. The correlation remained low with the highest quality units. **B,** The transparent time series in the background is the average d' in sliding 50 ms windows normalized to an arbitrary maximum. The heavy line displays the average site by site corrected correlation coefficient (ρ_N) using a sliding 50 ms window. The dashed lines represent the correlation values with larger integration windows. **C,** Face similarity judgements made by human subjects were only weakly correlated the neural response of multiunit sites in the MFP. The distribution of noise corrected correlations had a median correlation of rho = 0.10.

occurred at 141ms post stimulus onset (fig 3D). The elevated response qualitatively returned to baseline (ρ_N = -0.01) by 412ms. While the peak average corrected correlation was modestly elevated in comparison to the median correlation calculated with the standard response window, (0.21 vs. 0.19), the value is still far less than what should have been observed if MFP sites were acting like face detectors. To examine the role that the length of the response window may have had on our results, we re-performed the analysis with a number of different window sizes (100, 150, and 200 ms) in order to test if longer integration windows might result in a higher peak correlation. No window size tested resulted in a greater maximum correlation value than the 50ms window (fig 4D).

Finally, we considered the possibility that the response of sites in the MFP might be correlated with the degree that images from our experimental image sett were considered face-like by human subjects. A related psychophysical experiment also asked subjects to consider how similar each image was to a real face (Meng et al., 2012). We estimated the noise corrected correlation between the multiunit response in our MFP sample to the similarity judgements made by human subjects. Similar to the correlations with face detection performance, behavioral ratings of face similarity were only weakly correlated to the multiunit response of MFP sites on the same images (median $\rho = 0.10$, corrected for noise).

3.4 Discussion

We found that individual MFP sites lack a necessary signature of a face specific area (Cohen and Tong, 2001), in that the neural responses to at least one image set is

markedly discordant with the unequivocal "face" and "non-face" perceptual labels of human subjects. While it is possible that analysis of the animal's face detection behavior may result in a higher consistency between the response of MFP neurons and psychophysical behavior, this explanation seems unlikely for images that are well above threshold (near ceiling) for detection in a face detection task by humans. The true face images employed in our experiments were easily detected by human subjects (near ceiling performance) and could reliably activate the human Fusiform Face Area in related fMRI experiments (Meng et al., 2012). Nevertheless, it is possible that using an image set behaviorally validated by macaque subjects may shift the distribution of correlation coefficients towards higher values.

While previous studies have observed a strong average preference for images of faces in the response of neurons localized to the MFP, these observations were made using a computationally trivial subset of images. In a simplified regime, face detection (e.g. respond more to images of faces than other classes of images) no longer requires the presence of specialized face mechanisms to produce category selectivity, but could result from shape features less complex than a whole face. As an extreme example, "round" preferring neurons might appear to be "face" neurons if we only tested them with a simple image set of faces and non-face objects that did not expose the variation that makes face detection difficult. This is not a technical detail -- on the contrary, it is the presence or absence of such specialized mechanisms for face processing (e.g. that a neuron only responds to a face) that are the key distinguishing features between standard shape based and modular domain specific hypothesis of face processing in the ventral visual pathway. Indeed, the prevailing view in the non-face IT literature is that IT neurons are not tuned to semantic categories, but are tuned to geometric features (Baldassi et al., 2013; Kobatake and Tanaka, 1994) such as curvature (Brincat and

Connor, 2004; Connor, Brincat, and Pasupathy, 2007) and non-accidental shape properties (Vogels, Biederman, Bar, and Lorincz, 2001).

While it might be tempting to conclude that neurons in the MFP respond in a manner consistent with shape based models, rather than semantically as would be expected given the language used to describe these areas (i.e "face patch," "face cell" etc..), the truth may be far more complex, especially given that we have no readily common language to precisely describe intermediate shape representations from semantic categories. In sum, while the MFP is likely to participate in processing images of faces, in light of these findings, it cannot be considered the site of "face cells" that individually convey an explicit, firing rate representation of the presence or absence of a face.

3.5 Figure Legends

Figure 1. Experimental design. Images conventionally used to examine face selectivity in the MFP have minimal incidental image variability, consisting mainly of frontal views of faces cropped to be similar in spatial envelope. We use a novel image set derived from the output of a computational face detection system and validated by human subjects.

Figure 2. Sites in the MFP selective for human faces. A, The MFP was localized using MION enhanced fMRI. X-ray localization (see methods) was used to estimate the 3D location of recorded sites (inset shows sites projected to a flattened 2D sheet). B, Images of face and non-face distractors used in X-ray targeted electrophysiology experiments to screen for the face selective sites used in the study. C, Selected multiunit sites under examination responded selectively to faces from our conventional image set. Average response to each image is normalized by the standard deviation at each site.

Figure 3. Face selective sites with conventional images correlate only weakly with human face detection behavior. A, The average response from each multiunit site to images common to all experimental protocols. B, An example site whose neural to behavioral correlation was near the median for the distribution ($p = 0.19$). C, The distribution of corrected correlation coefficients between the multiunit spiking and the probability that human subjects rated the same image as containing a face.

Figure 4. Sorted single units display similar category selective responses. A, Response to the 5 best exemplars (highest d') from three image categories for an

example sorted unit. B, The average response from each isolated unit to the images common to all protocols.

Figure 5. Examination of the neural to behavioral correlation. A, The mean corrected correlation coefficient, as a function of SNR. The correlation remained low with the highest quality units. B, The transparent time series in the background is the average d' in sliding 50 ms windows normalized to an arbitrary maximum. The heavy line displays the average site by site corrected correlation coefficient (ϱN) using a sliding 50 ms window. The dashed lines represent the correlation values with larger integration windows. C, Face similarity judgements made by human subjects were only weakly correlated the neural response of multiunit sites in the MFP. The distribution of noise corrected correlations had a median correlation of rho = 0.10.

Chapter 4

The macaque middle face patch is a region of cortex enriched in neurons that participate in the intermediate representation of faces.

4.1 Introduction

This thesis has explored the spatial structure and correlation to behavior of the category selective signal for images of faces in the middle face patch. More specifically, in chapter two we described the use of a novel X-ray imaging system to spatially map the spiking multiunit response across the cortical tissue in the region of the fMRI identified middle face patch in two macaque subjects. A flattened representation of the 3D cortical surface was used to visualize and model the spatial organization of the face selective signal measured in our sample of sites. We found evidence for a ~6mm region of cortex, enriched with sites that had a preferential response when the subject viewed images of faces; we defined this region as the physiological MFP (pMFP). This region was not homogenous, rather the proportion of sites that were selective for faces could peak as high as 96% in the center of the pMFP. The purity of face selective cells gradually

decreased from the estimated center of the pMFP, decreasing to a purity that could be as low as 3% outside the pMFP (with variation in those values depending on the model and measurement of selectivity used). Not surprisingly, a weaker definition of face selectivity resulted in a higher proportion of face selective units both in and out of the pMFP. These data suggest that the fMRI activation related to the presentation of face images results from a high proportion (high purity, though not 100% pure), of neural sites that respond more on average when the subject views images of faces than when the subject views images of non-face distractors. The variation in the category selectivity for faces that was measured at neural sites across the entire patch was poorly fit by all of the low spatial frequency models we examined, suggesting the presence of higher frequency spatial structure in the pMFP. Finally, previous studies have reported conflicting estimates on the purity of the MFP. Some studies have reported very high estimates, suggesting that this area is a cortical region dedicated to face processing. In contrast, other studies have reported only a modest estimate of the purity in the MFP, which could be consistent with the patchy or columnar organization previously reported for shape organization in IT, and not necessarily a spatial organization dedicated for faces. We found that the MFP could have a high proportion of face selective sites near the middle of the patch, but that the proportion of face selective sites varied as a function of its distance to the center of the patch.

In chapter three, we examined the correlation between the activity of putative face cells in the MFP and the behavioral performance by human subjects in a face detection task with the same stimuli. We found that the correlation between these two measurements was low, given that human subjects readily identified the images as containing a face (e.g. near ceiling performance in a face detection task). These data suggest that the MFP does not explicitly represent the presence of a face in that the signal derived from multi or single units do not act as face detectors. In contrast to models that require the

explicit detection of faces to act as a gate for signals that are sent to downstream face patches, our results suggest that sites in the pMFP play a relatively early role in the generation of a signal that is correlated with face detection behavior. These results have led us to conclude that the MFP is likely an area involved in the intermediate processing of faces, with an explicit representation presumably somewhere else in the brain.

The remainder of this chapter will present a general discussion for each of the previous chapters. The goal is to relate the thesis work to the general literature discussed in chapter one, placing the work of the thesis in the context of similar endeavors to understand face processing in the temporal lobe of the macaque. The limitations of our methods will similarly be discussed for each chapter. The goal of this portion of the summary will be to draw attention to the limitations our results should have on making inferences from our data. Finally, the chapter will discuss the possible extensions or directions that seem outstanding given the previous literature discussed in chapter 1 and the results of this work.

4.2 A region of cortex consisting of a graded enrichment of sites with a category selective signal for faces.

As with previous studies, we observed numerous sites in the fMRI defined MFP region that demonstrated an increase in their spiking responses whenever the subject viewed images of faces. Unlike previous studies, we found evidence for an enrichment of sites with a category selective signal for faces that extended ~6mm across the cortical surface. We defined this region as the physiological MFP or, pMFP. Our study can be compared to earlier studies that have sampled the MFP in a number of ways. First the term MFP has been used in the literature by some authors to either effectively or explicitly denote

both activations in the middle temporal lobe (e.g. areas ML and MF; (Freiwald et al., 2009; Freiwald and Tsao, 2010; Rajimehr et al., 2009; Tsao et al., 2006), while other authors have included only the lateral activation in describing the characteristics of the tissue localized to the MFP (Bell et al., 2011). In contrast, we examined the neural activity exclusively at the fMRI activation located on the convexity of the STS (i.e. ML). Our results suggest that the MFP does not form one larger patch with the activation commonly observed in the fundus of the STS (i.e. MF) as has been speculated previously (Rajimehr et al., 2009). Secondly, we found that the pMFP had a lateral extent on the surface of the cortex of ~6mm in diameter. Our estimate of the size of the MFP is parsimonious with the volume [3] of the MFP estimated with fMRI (Tsao et al., 2008), and falls in the range of the "lateral clumping" reported in the upper bank of the STS for head pose (e.g. face) selective cells (Harries and Perrett, 1991). While the clusters of neurons selective for head pose in the upper bank of the STS were estimated to range in size from 3-6 mm, regions between these putative clumps were found to contain no head selective cells at all. In contrast, we observed that the number multiunit sites with a category selective signal for faces decreased in frequency over the spatial extent of the pMFP to a low baseline proportion of face selective sites outside the pMFP.

Our results are also consistent with previous estimates on the baseline number of face preferring cells (~10%) outside of large category selective clusters (Baylis et al., 1987), but not with recent studies that found a much higher proportion of face selective cells

[3] In the simplest case, a cylinder with a radius of 3mm along the cortical thickness (estimated to be ~2.5mm) would have a volume of ~70.7mm^3, while the average volume of ML, estimated with fMRI has been reported to be ~70.1 mm^3.

outside the MFP (Bell et al., 2011)[4]. While we did observe that the exact proportion of sites inside and outside the pMFP could vary depending upon the selectivity metric used, it is more likely that the discrepancy between our results and the those of Bell et al. stem from the more reliable spatial resolution of our recording techniques. Slight deflections of the electrode can produce spatial variation related to the depth the electrode is moved in the brain, relative to where the electrode should be from estimating its position geometrically (e.g. from a grid location in the recording chamber). Such deflections are of no consequence to our X-ray imaging system, because the position is estimated from the final position of the electrode tip itself. While there is spatial error in estimating the 3D position from our reconstruction procedure, this has been shown to have a median value less than 150um (Cox et al., 2008) in a skull-based frame, and up to ~500um in a tissue-based frame (Issa et al Cosyne abstract).

4.2.1 Limitations of our work on the spatial characterization of the MFP

As might be expected, the amount of variability induced to our spatial analysis from the surface flattening and visualization methods depended on the amount of distance we allowed sites to travel between their X-ray reconstructed 3D position to the 2D coordinates on the surface model. While we did find that larger distances induced variability in the average selectivity measured at local spatial positions on the 2D sheet compared to the 3D reconstructed positions, wide variation in this variable did not dramatically alter the estimates of the model parameters in our study. Our study simplified the 3D complexity of mapping the MFP by projecting to a flat 2D surface,

[4] This study reported 29-30% of the sites sampled in their object selective area (located outside the MFP) contained 29-30% face selective cells (fig 6). However, it is still, in theory, possible that there exists other smaller clusters of category selective cells outside the pMFP that are too small to be detected by fMRI, and that we did not observe in our data, which could explain the differences in purity reported by the two studies.

which allowed us to collapse across the thickness of the cortical ribbon in a reasonable way. Though the detailed microstructure of the pMFP may reveal important divisions across the layers of the cortical ribbon, the resolution of our X-ray imaging system would have difficulty reliably describing this structure at a meaningful resolution. To overcome this limitation would require a strong relative effect between the layers sampled to overcome the effective spatial blurring filter from the X-ray system. Given the data collected in our studies, we found no evidence for differences in category selectivity across putative upper and lower layers estimates. In this analysis, our conclusions were limited by the fact that the data collected in our two subjects tended to be spatially biased to either the upper (M1) or lower layers (M2). Perhaps a more limiting factor in our analysis was the use of T1 weighted anatomical MRI to define the spatial position of the recorded sites.

The relative positions of the reconstructed 3D positions of the sampled sites on an electrode do not depend on anatomical information per se. The position is reconstructed relative to a skull based reference frame, and projected to an anatomical location based on the position of MRI visible markers at known positions in the frame. It is not known how the shape of the brain changes over repeated electrode penetrations (due to tissue damage, etc.), over the time course of physiological data collection (plasticity), or between the time the fMRI selectivity maps were taken and the time the electrophysiological experiments were conducted (development). All of these factors contribute to limit the accuracy of the spatial localization of our recording sites. We circumvented some of these issues by defining the MFP from physiological samples, though the exact position of these sites in relation to the 3D anatomy is subject to some of the limitations described. Ideally, structural anatomical images would be acquired at multiple time points throughout the recording process in order to model potential shape changes across the data acquisition. In addition 3D models of the deformation induced

by electrode penetrations could account for some of the error induced to the spatial position of our samples. Though these factors would be crucial for studies of fine scale structure in the pMFP, given the general characterization that was the goal of our study, these factors should have minimal contribution to the overall results of our work as reflected in the reliability of our parameter estimates.

Another potential limitation (or difference) between our characterization of the MFP and previous work is our use of the multiunit signal. Previous studies have recorded the response of isolated single units, whereas our work is based on the multiunit response at known spatial locations in the MFP. The multiunit signal likely reflects the activity of neurons within 140-300um from the electrode tip (Logothetis, 2003), whereas single units are typically recorded at distances < 20um from the electrode tip to a cell body (Buzsáki, 2004). The response preference of isolated single units has been shown to be correlated at distances less than 400-600um (Fujita et al., 1992; Sato et al., 2009), while multi-unit recordings have shown correlations in stimulus preference up to ~800um (Kreiman et al., 2006). These results suggest that the multiunit signal should be well correlated at small distances (<800um). Reexamination of our model fits with single units isolated from our multiunit data would be one way to address concerns over the use of multiunit samples in our current data set. However, we did not explicitly set out to record isolated single units, and our data acquisition methods are not optimal for recovering isolated units. This limitation would effectively reduce the number of samples in our study and impact the reliability of our spatial modeling efforts. Given the goal of our study, to spatially characterize the category signal for faces in the MFP, the use of the multiunit signal does not seem to seriously impact the conclusions of our efforts.

4.2.2 Extensions and future directions

Their are several avenues for extending the spatial characterization of the pMFP reported here. The primary motivation for studying the spatial structure for neuronal responses to faces in this area of the temporal lobe is due to the fMRI signal observed in awake macaques to images of faces, and the potential homologies to the human brain. The reliability of this signal is however not well understood, and is variable both across subjects and laboratories (Pinsk et al., 2009; Tsao et al., 2008). Understanding whether these individual differences result from noise inherit in fMRI, are the result of individual cortical geometry or are somehow learned, is an important factor in interpreting the relationship between the fMRI signal and the physiologically defined functional structure. This is critically observed in the case of relating fMRI contrast maps taken in the subject before neurophysiological investigation to the physiologically defined spatial inferences reported on years later. While it was not the goal of our study to relate the physiologically defined category selective signal explicitly to the fMRI defined signal, such questions are of considerable interest to a broad community.

Another limitation of our work was the ability to reliably study the fine scale spatial structure in the cortical depth dimension. Previous work has suggested that category selectivity may not be spatially homogenous across layers. Sato and colleagues observed that the correlation in the multiunit response to a set of images, at different sites in a functionally defined spot, was reliably above zero only for the upper layers (Sato et al., 2009). As our data was collapsed across layers, our model fits might reveal differences between upper and lower layers. While we explicitly attempted to examine this in our data, we found no reliable difference in selectivity estimates between upper and lower layers. Additional limitations come from the X-ray imaging system itself. Deformation of the cortical tissue from drag friction of the electrode could add up to 350um of error to spatial depth estimates, inducing spatial error across estimates

between the two layers. While these limitations are not crucial to our study, they would be important for understanding how the category selective signal is transformed across layers in the pMFP.

4.3 Neural activity in the MFP does not correlate well with human face detection behavior

One underlying assumption in previous work that has sought to characterize the MFP is the assumption that face processing occurs exclusively at the fMRI defined face patches. This idea has lead to the proposal of a broad class of qualitative models. These models hypothesize that face detection serves as a domain specific filter on the visual input to a putative face network (Moeller et al., 2008; Tsao and Livingstone, 2008). Research that has examined face detection behavior has provided some agreement with this idea. Studies with human subjects suggest that detection can be a separable process in object recognition under certain conditions (Mack, Gauthier, Sadr, and Palmeri, 2008). Additional studies (described next) have demonstrated the object and face detection can occur very rapidly, suggesting that the neural correlates of face detection behavior can occur at an early stage in the visual processing stream. Monkeys and humans have been shown to be able to detect complex objects vary rapidly, with reaction times as low as 250 ms for human subjects and 210 ms for monkeys (Fabre-Thorpe, Richard, and Thorpe, 1998; Thorpe, Fize, and Marlot, 1996). Explicit examination of face detection in natural scenes by human subjects has curiously revealed that subjects performed equally well when detecting either faces or animals (median reaction time ~382ms), with only a subtle difference between detecting inverted bodies or faces (Rousselet, Macé, and Fabre-Thorpe, 2003). These authors argued that face and non-face detection might rely upon the same neural mechanisms. In support of this view, face selective

neurons recorded in anterior IT responded no differently to scrambled or whole faces when the images were presented for only a very short duration (16 ms) and backward masked (Rolls, Tovee, Purcell, Stewart, and Azzopardi, 1994). This result is in contrast to the differences in activity that might be expected due to the configural advantage reported in face detection task with human subjects. Similarly, electrical stimulation to face selective cells in monkey anterior IT (+14-21 AP) can most effectively bias face detection at latencies between 50-100 ms (Afraz, Kiani, and Esteky, 2006). As proposed earlier, these studies suggest that face detection occurs very early in the visual processing stream.

While the general idea, that category selective neurons have "detected" exemplars from that category is intuitive, real world object (and face) detection/identification is a complicated and difficult computational problem (DiCarlo and Cox, 2007). The difficulty of the problem is one that summarizes the entire project of object recognition, in that the representation to be detected, or read out, by downstream neurons must generalize over a broad range of identity preserving image variation that includes variability in position, size, pose, lighting, and clutter (DiCarlo, Zoccolan, and Rust, 2012; Riesenhuber and Poggio, 1999). This is a non-trivial problem, and numerous accounts of the visual system have hypothesized a gradual building of tolerance to image variation over multiple hierarchical stages (or visual areas) in the ventral visual pathway (Connor et al., 2007; Serre et al., 2007a; Serre, Wolf, Bileschi, Riesenhuber, and Poggio, 2007b; Wallis and Rolls, 1997). The degree to which the modular hypothesis insulates face detection within a putative face network is one way of distinguishing the modular hypothesis from other standard hierarchal models of object recognition. In standard models, the computational problem stems from maintaing object identity over incidental image variation; category and identity "detection" diverge in the amount of image variation tolerated by the system, rather than the kind of computation needed to

perform the problem (Riesenhuber and Poggio, 2000). In contrast, modular hypothesis suggest that the major computational work (and hence the explicit allocation of cortical resources) stem from distinguishing within class sources of information about faces (Tsao and Livingstone, 1998). Does that face belong to a woman or a man? Are they happy or sad? While there are potentially an infinite variety of ways that a modular face processing system could be instantiated in the brain, the degree to which face detection is computationally distinct from the domain specific knowledge that the face system is hypothesized to recover distinguishes it from standard shape based models. In other words, the presence of something that the subject would call a face in the visual input of the organism is found at the earliest stage of the face network. This would allow for (putatively) advanced domain specific knowledge to be culled from the input beyond the identity of the face (the hypothesized goal of object recognition).

Given the limited types of stimuli that have been used to explicitly examine category selectivity in the MFP, we wondered to what extent sites in the MFP could support face detection when the image set was made challenging for shape based hypothesis (like standard models of the ventral stream) but not for semantic detector models. To find such an image set required that the images be trivial for behavioral face detection but non-trivial on an image or shape basis. In order to find such images, we used non-face images that were readily rejected as faces by human observers (see below), but were falsely detected as faces by a computer face detection algorithm (Schneiderman, 2004). In addition, we used the "genuine," or true, faces that were detected (as a face) by the same face detection algorithm. The falsely detected face images are, by definition, not genuine faces. Yet, they share enough features with the genuine faces that they are difficult to distinguish from each other on the basis of shape (where "shape" here is operationally defined as the representation built by the computer vision algorithm). This set of face images (falsely detected faces and genuine faces) was explicitly tested

with human subjects in a face detection task (Meng et al., 2012). This insured that the genuine face images were readily detected as faces by human subjects, and, conversely, falsely detected face images were trivially rejected as containing a face. Crucially, the falsely detected face images were rarely reported as containing a face by human subjects (<6%), in contrast to the genuine face images, which were almost always identified as containing a face by the same subjects (>95%). The response of MFP sites was only weakly correlated to behavioral face detection performance in human subjects on this set of images. In other words, we found that images which were reliably called a face by human subjects could elicit weak MFP responses, while images that were almost never called a face by subjects could elicit significant responses from sites localized in the MFP. The median correlation across all of the sites in our sample was 0.18, where a value of 1.0 would be expected if the MFP were explicitly gating face detection (as proposed in the modular models reviewed above), while a value of 0.0 would be expected if the MFP had no ability to gate face detection. A similar result was found when we examined the correlation between how similar subjects thought each image was to a face and the neural response to that same image. These results cannot be explained by "noise" in the data, as our analyses fully account for trial-by-trial spiking variability and human inter-subject report variability. We also found that the correlation remained low in single units isolated from our multiunit sample, ruling out the possibility that the multiunit signal averaged out the true magnitude of the correlation. These results suggest that the response of individual MFP neurons are insufficient to account for face detection behavior with our experimental image set in humans subjects; failing to bear out one of the main predictions of the modular MFP "gating" hypothesis for face processing.

4.3.1 Limitations of our face detection study

There are a number of limitations that impact the inferences that can be drawn from our results. A major limitation is the reproducibility of our methods. A basic tenant in science is the idea of reproducibility; showing that the same result can be found using the same methods in a different exemplar. Our study consisted of only one monkey subject, so it is possible that this monkey is an outlier (e.g. not representative of normal monkey face processing). A second major limitation of our study was the use of human behavior in our correlations to the neural activity in the monkey. While speculation on the homology between the human FFA and the monkey MFP has been useful in generalizing domain specific hypothesis on face processing, one might propose a more complicated relationship between the human and monkey face systems. While that proposition would reconcile our results with the modular gating hypothesis (by essentially rejecting the monkey model as a good model of human face processing), such reasoning should extend to inferences drawn from previous results with similar methods (Nasr et al., 2011; Rajimehr et al., 2009; Tsao et al., 2008; Tsao and Livingstone, 2008). Finally, another limitation of our study is that we did not present a positive control, in that a more convincing case could have been made for our conclusion if we had found a sample of neurons in IT, outside the MFP, that did contain an explicit representation of face detection behavior (i.e. whose responses were well correlated with the behavioral judgements of our image set). For example, perhaps the more anterior face patches show such a behavior, and that would then become the site of the "gate" in the modular gating hypothesis. Such a hypotheses would then be quite limited as it would not speak to most of IT (e.g. MFP), shifting the "face processing network" to even further downstream areas (e.g. frontal cortex).

Arguments against the main conclusion of our study from the use of human face detection behavior instead of monkey face detection behavior can solidify into an

attention hypothesis and a hypothesis about fundamental differences in the ways monkeys behaviorally use face information as compared to human subjects. The attention hypothesis would argue that the monkey in our study found the faces in the experimental image set to be uninteresting or not engaging. We do not know what the monkey is actually "thinking" when it views a given image. Mechanistically, this attention hypotheses might mean that the ventral stream network is at an operating point that is very different than the one it would be at if the monkey were actually "doing" the task. There are numerous examples in the experimental literature where changes in task demands (phenomenologically called changes in "attention") have resulted in interactions with neural tuning (reviewed in Reynolds and Heeger 2009). This hypothesis would than claim that the neural response across the image set was attenuated and altered (randomly perturbing the responses to all the images such that responses were no longer correlated with behavior). The fact that the conventional face images interleaved into the experimental protocol resulted in large responses suggests that the monkey was indeed attending to the stimulus location. Furthermore, our RSVP protocol should have mitigated any possible effects that might have arose from the animal only looking at the visual display when a conventional face appeared (see methods). It is also implausible that random factors in the images caused some neurons to respond more to some images than others. We did not observe any particular distractor images that reliably modulated subsamples of MFP sites. Finally, studies have explicitly observed that 2D computerized images of human and monkey faces elicit preferred viewing of the head and face region over other areas of a naturalistic scene in free viewing macaques (Nahm, Perret, Amaral, and Albright, 1997). This would suggest that our experimental image set should have naturally attracted the animals attention over other aspects of image or, the surrounding visual environment. Furthermore, our experimental methods (monkey passively fixating), have been used in numerous studies of face and object processing in the ventral stream.

The other possible explanation for the low correlation we observed in our study might stem from a fundamental difference in the way that humans and monkey subjects perform face detection. The behavioral literature for faces in non-human primates has largely focused on behavior thought to be characteristic of human face recognition behavior such as face inversion and holistic processing. In contrast, a comparative mapping of the factors that affect performance between monkeys and humans on behavioral tasks that rely on images of faces remains unknown. In fact, one could argue that outside of the well known exotic face behaviors (such as the inversion effect and parts-whole context effect) there has been no detailed attempt to compare behavioral face processing between humans and monkeys. Nevertheless, studies of face behavior (described below) have provided some evidence that monkeys process conspecific images of faces in a manner similar to humans, though it is not clear if monkeys treat face images from other species (including human faces) in the same manner as images of faces from their own species.

A number of studies have observed that when monkeys view 2D images of faces they spend the most time viewing the eye region of face images (Dahl, Wallraven, Bülthoff, and Logothetis, 2009; Guo, Robertson, Mahmoodi, Tadmor, and Young, 2003; Nahm et al., 1997; Parr, Winslow, Hopkins, and de Waal, 2000). Furthermore, covering or altering the eye region can significantly degrade face detection behavior in monkeys to conspecific face images (Parr et al., 2000), as well as schematic drawings of human faces (Keating and Keating, 1993). Studies in human subjects have also described the importance of the eye region in face detection (reviewed in Keating et. al., 1993). A case study with a prosopagnosic patient reported that control subjects found the left eye region most informative in a face detection task. In contrast, the prosopagnosic patient relied heavily on the lower half of the face image, particularly the mouth region

(Caldara et al., 2005). While the prosopagnosic patient could perform the task, she was worse than controls. The patient had adopted an efficient strategy using the mouth region of the face stimuli to do the task, while controls (with normal face processing abilities) used the eye region for the task. These studies suggest that monkeys and humans use one similar diagnostic feature from the visual images of faces: the region of the face containing the eyes. However a number of studies have produced conflicting results on face inversion tasks, which would suggest a fundamental difference in the way the specialized face systems in humans and monkeys use visual information from faces.

The face inversion effect is the observation that visual behavioral tasks become more difficult for human subjects when face images (more than for non-face objects) are inverted (Yin, 1969). An effect due to inversion is considered a hallmark of face processing in humans, suggesting that similar effects in monkeys would lend support for the idea that monkeys and humans process face images in a similar manner with similar neural mechanisms. While a number of studies have failed to find an inversion effect in macaques, others have found inversion effects (reviewed in Parr, 2011). Recent studies have reported that the inversion effects can be species specific (Dahl et al., 2009), or that macaques have inversion effects that are not face specific (Parr, 2011). For example, using a preferential viewing adaptation paradigm, Dahl and colleagues (2009) found that monkeys tended to spend more time viewing the eye regions of monkey faces and less time viewing the eye regions of human faces. The opposite was true for human subjects: they spent more time viewing the eye regions of human faces and less time viewing the eye regions of macaque faces. When the images of macaque and human faces were inverted, however, both human and macaque subjects viewed the eye regions of both species equally. These results were used to argue that human and macaque subjects use similar viewing strategies when viewing images of conspecific

faces but not the faces of other species. Face inversion acts to disrupt the specialized viewing strategies for conspecific faces, and forces the subject to use a similar default strategy for both humans and monkeys face images. In contrast, Parr and colleagues (2011) found that in a match to sample task rhesus macaques showed an accuracy cost for inverted faces (both macaque and chimpanzee), inverted houses, and a trend for inverted shoes. This was true even when task performance was normalized for generalized inversion affects from non-face objects. The study also examined the same set of tasks with chimps, and unlike the rhesus macaques, the chimps demonstrated a normalized accuracy cost only for inverted conspecific faces, suggesting an evolutionarily more primitive face system for rhesus monkeys but not chimps. While these studies suggest that rhesus macaques may not process faces in the same way as humans, the fact that macaques do not show an inversion effect under the task conditions present in those experiments, does not necessarily mean that monkeys do not use configural information in processing images of faces. Perret and colleagues (1988) also found that macaques displayed no differences in reaction time or accuracy in detecting upright or inverted faces. However, when these experimenters designed a task that explicitly required the monkeys to more directly use configural information (meaning that simply looking for discriminative parts of the image would be an insufficient behavioral strategy), both monkeys and human subjects displayed an inversion effect. This result was used to argue that macaques had been using a part based strategy for solving the conventional face inversion task.

The studies discussed above suggest that human and macaques use similar strategies when viewing faces (eye scan paths and the saliency of the eyes in the image). However, it remains unclear if humans and monkeys, as characterized by face inversion effects, process face information similarly or if macaques and humans have different strategies for "other race" face images. Recent functional imaging (Bell et al., 2009;

Pinsk et al., 2005; Tsao et al., 2003) and neuophysiology studies (Bell et al., 2011; Freiwald et al., 2009; Freiwald and Tsao, 2010; Tsao et al., 2006) would suggest that neurons in the MFP respond robustly to images of faces, regardless of the species of the image subject (macaque or human). Nevertheless, a stronger case could be made for our results if behavioral data from a face detection task was collected from the subject in our study.

In summary, we argue that sites in the MFP do not act as face detectors given that the correlation between the response of neurons in the monkey MFP and human face detection behavior to our set of experimental images was low. It could be argued that our conclusions are limited given the use of a single animal, the use of human behavior in our correlation analysis, and because we did not look for sites that did respond to our image set in a manner that was well correlated with human behavior. We concede that addressing these concerns would provide the basis for a stronger conclusion from our data. However, we also argued that these claims should not seriously impact the conclusions we have drawn under the reasonable assumptions that (1) our subject is not an outlier among other animals, (2) makes use of a typical face processing system which (3) is a good model for the human face processing system, and (4) that our conclusions should be directed to the hierarchal stage that the MFP belongs to in the putative face network.

4.3.2 Extensions and future directions

Two obvious direct extensions of the research conducted in chapter three revolve around the methodological limitations described earlier. Replicating the same experimental results in a second animal would greatly enhance the generalizability of our results, while expanding our study to include human subjects and behavioral

ratings from monkeys would directly address some of the limitations discussed previously. We know from earlier work that our experimental image set does elicit significant responses from the FFA in human subjects (Meng et al., 2012). However, the magnitude of the response in the FFA to the experimental image set (as compared to conventional images) is not known. Given the suspected homology between the monkey MFP and the human FFA, we might expect to observe that the face images from our experimental image set elicits weak but reliable responses from the FFA in comparison to faces from our conventional image set. Additionally the large field of view afforded by fMRI could suggest where a face detection signal for our experimental image set might be localized. On the other hand, the human FFA may have a similar response to the experimental image set as to conventional images of faces, which would suggest a fundamental difference between the FFA and the MFP.

The experimental results from chapter three could further be expanded by examining face detection behavior, and the spiking responses of neurons recorded from the MFP, in the same subject. Ideally, neural recording would occur while the animal is actively engaged in a face detection task, though several issues could make the interpretation of such results problematic. First of all, it is unclear how a monkey naturally makes use of face information (i.e. in the wild), therefore the specific task employed in an experiment might not behaviorally engage the animal in the same way that an experimenter might have initially envisioned. While standard operant conditioning techniques could be used to train macaques to perform face detection in a behavioral paradigm, previous studies (reviewed earlier) have observed that macaques can use alternate behavioral strategies in performing certain kinds of tasks with images of faces. Indeed, Dahl and colleagues have argued that the conflicting results previously observed in the experimental literature with inversion tasks are due to methodology or learning effects, emphasizing the importance of natural unrestrained viewing methods for collecting

reliable behavioral data from monkeys (Dahl et al., 2009; Dahl, Logothetis, and Hoffman, 2007).

Another way of extending the work described in chapter three centers around the larger response we observed to conventional images of faces. This would suggest that there are key feature differences between the experimental and conventional face images that matter to MFP sites. Understanding these differences would constitute an important set of experiments that would directly extend the findings reported in chapter three, and potentially lead to a class of encoding hypothesis for MFP sites that could be directly examined with novel stimuli. A number of first order hypothesis could account for these differences. For example, one difference between the conventional and experimental image sets was the resolution of the images. The experimental images were lower in resolution than the conventional images. While face identification for human subjects is not disrupted for familiar faces with lower resolution (Sinha, Balas, Ostrovsky, and Russell, 2006), the images in our study were not "famous" for the monkey, so it is not clear that the image resolution did not play a part in the weaker responses we observed at sites in the MFP to the experimental images in our study. Previous studies have found that lower resolution images of faces can degrade the response of cells selective for faces (Perrett et al., 1984). Finally, extending our experimental protocol to sites outside the MFP would enhance the overall work produced in chapter three. Sampling putative face patches in downstream regions of the visual hierarchy might expose areas of cortex that correlate better to human behavior than MFP sites, under the assumption that these images are being detected as faces by the monkey in the experiment.

4.4 General Conclusions

This thesis has attempted to characterize the macaque middle face patch by examining the spatial structure of category selective sites in the cortical tissue identified by fMRI. In addition, we examined the role that spiking responses localized to this region have in supporting face detection behavior. We found that the activation identified by fMRI consisted of an enriched zone of category selective sites that preferred conventional images of faces. This enriched zone extended ~6mm in diameter on the 2D inflated surface. The enrichment of sites selective for images of faces was not uniform throughout this region: the purity of face selective sites could range from a peak as high as 96% in the center to a value as low as ~3% outside the region, depending on how category selectivity was defined. We defined this physiologically identified cortical region as the physiological middle face patch (pMFP). This novel result reconciles conflicting claims from the literature, because we demonstrate that the enrichment for face selective cells gradually builds from the periphery. The variation in face selectivity within the pMFP is so strong that it is difficult to rule out a strict module type of organization (i.e. the pMFP has a hard boundary). While sites under a weak face selectivity criteria could be observed outside the pMFP, sites with very high category selectivity (d'>2) were almost 50x more likely to be observed inside the pMFP, suggesting that the region is instrumental in representing information about images of faces. Given the strong category selective signal observed in the pMFP, we examined face detection using a novel image set that was challenging for computational face detection algorithms, but trivial for human subjects. Surprisingly, we found that sites in the pMFP were only weakly correlated with face detection behavior by human subjects using the same images, suggesting that contrary to some hierarchal face system models, the MFP does not act as a gate on whether an image contains a face or not. Similarly,

this finding shows that the pMFP has not solved the face detection problem, in that the face detection performance observed in human subjects to our set of images, is not realized in the spiking responses of sites in the pMFP. Though there are a number of limitations to the studies described in this thesis, from this work we conclude that the pMFP is a region of cortex that contains a spatially graded selectivity profile that, at its center, is highly enriched for mid-level shape representations that likely contribute to face processing in the ventral visual pathway.

Bibliography

Abbott, L. F., Rolls, E. T. and Tovee, M. J. (1996). **Representational capacity of face coding in monkeys.** *Cerebral Cortex (New York, NY : 1991), 6*(3), 498–505.

Afraz, S.-R., Kiani, R. and Esteky, H. (2006). **Microstimulation of inferotemporal cortex influences face categorization.** *Nature, 442*(7103), 692–695.

Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R. and Zoccolan, D. (2013). **Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons.** *PLoS Computational Biology, 9*(8), e1003167.

Baylis, G. C., Rolls, E. T. and Leonard, C. M. (1985). **Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey.** *Brain Research, 342*(1), 91–102.

Baylis, G. C., Rolls, E. T. and Leonard, C. M. (1987). **Functional subdivisions of the temporal lobe neocortex.** *Journal of Neuroscience, 7*(2), 330–342.

Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B. H. and Ungerleider, L. G. (2009). **Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging.** *Journal of Neurophysiology, 101*(2), 688–700.

Bell, A. H., Malecek, N. J., Morin, E. L., Hadj-Bouziane, F., Tootell, R. B. H. and Ungerleider, L. G. (2011). **Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity.** *Journal of Neuroscience, 31*(34), 12229–12240.

Bouvier, S. E. and Engel, S. A. (2006). **Behavioral deficits and cortical damage loci in cerebral achromatopsia.** *Cerebral Cortex (New York, NY : 1991), 16*(2), 183–191.

Brincat, S. L. and Connor, C. E. (2004). **Underlying principles of visual shape selectivity in posterior inferotemporal cortex.** *Nature Neuroscience, 7*(8), 880–886.

Bruce, C., Desimone, R. and Gross, C. G. (1981). **Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque.** *Journal of Neurophysiology, 46*(2), 369–384.

Bruce, V. and Young, A. (1986). **Understanding face recognition.** *British Journal of Psychology (London, England : 1953), 77 (Pt 3)*, 305–327.

Buzsáki, G. (2004). **Large-scale recording of neuronal ensembles.** *Nature Neuroscience, 7*(5), 446–451.

Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F. and Rossion, B. (2005). **Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage.** *Journal of Cognitive Neuroscience, 17*(10), 1652–1666.

Calder, A. J., Rowland, D., Young, A. W., Nimmo-Smith, I., Keane, J. and Perrett, D. I. (2000). **Caricaturing facial expressions.** *Cognition, 76*(2), 105–146.

Clark, V. P., Keil, K., Maisog, J. M., Courtney, S., Ungerleider, L. G. and Haxby, J. V. (1996). **Functional magnetic resonance imaging of human visual cortex during face matching: a comparison with positron emission tomography.** *NeuroImage, 4*(1), 1–15.

Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009). **Power-Law Distributions in Empirical Data.** *Siam Review, 51*(4), 661–703.

Cohen, J. D. and Tong, F. (2001). **Neuroscience. The face of controversy.** *Science (New York, NY), 293*(5539), 2405–2407.

Connor, C. E., Brincat, S. L. and Pasupathy, A. (2007). **Transformation of shape information in the ventral pathway.** *Current Opinion in Neurobiology, 17*(2), 140–147.

Cox, D. D., Papanastassiou, A. M., Oreper, D., Andken, B. B. and DiCarlo, J. J. (2008). **High-resolution three-dimensional microelectrode brain mapping using stereo microfocal X-ray imaging.** *Journal of Neurophysiology, 100*(5), 2966–2976.

Dahl, C. D., Logothetis, N. K. and Hoffman, K. L. (2007). **Individuation and holistic processing of faces in rhesus monkeys.** *Proceedings of the Royal Society of London B: Biological sciences, 274*(1622), 2069–2076.

Dahl, C. D., Wallraven, C., Bülthoff, H. H. and Logothetis, N. K. (2009). **Humans and macaques employ similar face-processing strategies.** *Current Biology, 19*(6), 509–513.

Dale, A. M., Fischl, B. and Sereno, M. I. (1999). **Cortical surface-based analysis. I. Segmentation and surface reconstruction.** *NeuroImage, 9*(2), 179–194.

Damasio, A. R., Tranel, D. and Damasio, H. (1990). **Face agnosia and the neural substrates of memory.** *Annual Review of Neuroscience, 13*, 89–109.

De Souza, W. C., Eifuku, S., Tamura, R., Nishijo, H. and Ono, T. (2005). **Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques.** *Journal of Neurophysiology, 94*(2), 1252–1266.

Desimone, R., Albright, T. D., Gross, C. G. and Bruce, C. (1984). **Stimulus-selective properties of inferior temporal neurons in the macaque.** *Journal of Neuroscience, 4*(8), 2051–2062.

DiCarlo, J. J. and Cox, D. D. (2007). **Untangling invariant object recognition.** *Trends in Cognitive Sciences, 11*(8), 333–341.

DiCarlo, J. J., Zoccolan, D. and Rust, N. C. (2012). **How Does the Brain Solve Visual Object Recognition?** *Neuron, 73*(3), 415–434.

Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M. and Kanwisher, N. (2006). **Domain specificity in visual cortex.** *Cerebral Cortex (New York, NY : 1991), 16*(10), 1453–1461.

Downing, P. E., Jiang, Y., Shuman, M. and Kanwisher, N. (2001). **A cortical area selective for visual processing of the human body.** *Science (New York, NY), 293*(5539), 2470–2473.

Eifuku, S., De Souza, W. C., Tamura, R., Nishijo, H. and Ono, T. (2004). **Neuronal correlates of face identification in the monkey anterior temporal cortical areas.** *Journal of Neurophysiology, 91*(1), 358–371.

Epstein, R. and Kanwisher, N. (1998). **A cortical representation of the local visual environment.** *Nature, 392*(6676), 598–601.

Fabre-Thorpe, M., Richard, G. and Thorpe, S. J. (1998). **Rapid categorization of natural images by rhesus monkeys.** *Neuroreport, 9*(2), 303–308.

Felleman, D. J., and Van Essen, D. C. (1991). **Distributed hierarchical processing in the primate cerebral cortex.** *Cerebral cortex* (New York, NY : 1991), 1(1), 1–47.

Fischl, B., Sereno, M. I. and Dale, A. M. (1999). **Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system.** *NeuroImage, 9*(2), 195–207.

Freiwald, W. A. and Tsao, D. Y. (2010). **Functional compartmentalization and viewpoint generalization within the macaque face-processing system.** *Science (New York, NY), 330*(6005), 845–851.

Freiwald, W. A., Tsao, D. Y. and Livingstone, M. S. (2009). **A face feature space in the macaque temporal lobe.** *Nature Neuroscience, 12*(9), 1187–1196.

Fujita, I., Tanaka, K., Ito, M. and Cheng, K. (1992). **Columns for visual features of objects in monkey inferotemporal cortex.** *Nature, 360*(6402), 343–346.

Fukushima, K. (1980). **Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.** *Biological Cybernetics, 36*(4), 193–202.

Gauthier, I., Behrmann, M. and Tarr, M. J. (1999). **Can face recognition really be dissociated from object recognition?** *Journal of Cognitive Neuroscience, 11*(4), 349–370.

Gauthier, I., Skudlarski, P., Gore, J. C. and Anderson, A. W. (2000). **Expertise for cars and birds recruits brain areas involved in face recognition.** *Nature Neuroscience, 3*(2), 191–197.

Grill-Spector, K. and Malach, R. (2004). **The human visual cortex.** *Annual Review of Neuroscience, 27*, 649–677.

Gross, C. G., Rocha-Miranda, C. E. and Bender, D. B. (1972). **Visual properties of neurons in inferotemporal cortex of the Macaque.** *Journal of Neurophysiology, 35*(1), 96–111.

Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y. and Young, M. P. (2003). **How do monkeys view faces?--A study of eye movements.** *Experimental Brain Research, 150*(3), 363–374.

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K. and Rosen, B. R. (1999). **Location of human face-selective cortex with respect to retinotopic areas.** *Human Brain Mapping, 7*(1), 29–37.

Harries, M. H. and Perrett, D. I. (1991). **Visual Processing of Faces in Temporal Cortex: Physiological Evidence for a Modular Organization and Possible Anatomical Correlates.** *Journal of Cognitive Neuroscience, 3(1), 9-24.*

Hasselmo, M. E., Rolls, E. T. and Baylis, G. C. (1989a). **The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey.** *Behavioral Brain Research, 32*(3), 203–218.

Hasselmo, M. E., Rolls, E. T., Baylis, G. C. and Nalwa, V. (1989b). **Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey.** *Experimental Brain Research, 75*(2), 417–429.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L. and Pietrini, P. (2001). **Distributed and overlapping representations of faces and objects in ventral temporal cortex.** *Science (New York, NY), 293*(5539), 2425–2430.

Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch, P., Schapiro, M.B. and Rapoport, S. I. (1991). **Dissociation of object and spatial visual processing pathways in human extrastriate cortex.** *Proceedings of the National Academy of Sciences of the United States of America, 88*(5), 1621–1625.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P. and Grady, C. L. (1994). **The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations.** *Journal of Neuroscience, 14*(11 Pt 1), 6336–6353.

Haxby, J., Hoffman, E. and Gobbini, M. (2000). **The distributed human neural system for face perception.** *Trends in Cognitive Sciences,* 4(6), 223–233.

Hietanen, J. K., Perrett, D. I., Oram, M. W., Benson, P. J. and Dittrich, W. H. (1992). **The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex.** *Experimental Brain Research, 89*(1), 157–171.

Hoffman, E. A. and Haxby, J. V. (2000). **Distinct representations of eye gaze and identity in the distributed human neural system for face perception.** *Nature Neuroscience, 3*(1), 80–84.

Issa, E. B. and DiCarlo, J. J. (2012). **Precedence of the eye region in neural processing of faces.** *Journal of Neuroscience, 32*(47), 16666–16682.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. and Smith, S. M. (2012). **FSL.** *NeuroImage, 62*(2), 782–790.

Kanwisher, N. (2000). **Domain specificity in face perception.** *Nature Neuroscience, 3*(8), 759–763.

Kanwisher, N. and Yovel, G. (2006). **The fusiform face area: a cortical region specialized for the perception of faces.** *Philosophical transactions of the Royal Society of London Series B, Biological sciences, 361*(1476), 2109–2128.

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*(11), 4302–4311.

Keating, C. F. and Keating, E. G. (1993). **Monkeys and mug shots: cues used by rhesus monkeys (Macaca mulatta) to recognize a human face.** *Journal of Comparative Psychology (Washington, DC : 1983), 107*(2), 131–139.

Kobatake, E. and Tanaka, K. (1994). **Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex.** *Journal of Neurophysiology, 71*(3), 856–867.

Kreiman, G., Hung, C. P., Kraskov, A., Quiroga, R. Q., Poggio, T. and DiCarlo, J. J. (2006). **Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex.** *Neuron, 49*(3), 433–445.

Kriegeskorte, N., Formisano, E., Sorger, B. and Goebel, R. (2007). **Individual faces elicit distinct response patterns in human anterior temporal cortex.** *Proceedings of the National Academy of Sciences of the United States of America, 104*(51), 20600–20605.

Ku, S.-P., Tolias, A. S., Logothetis, N. K. and Goense, J. (2011). **fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques.** *Neuron, 70*(2), 352–362.

Liu, J., Harris, A. and Kanwisher, N. (2010). **Perception of face parts and face configurations: an FMRI study.** *Journal of Cognitive Neuroscience, 22*(1), 203–211.

Logothetis, N. K. (2003). **The underpinnings of the BOLD functional magnetic resonance imaging signal.** *Journal of Neuroscience, 23*(10), 3963–3971.

Logothetis, N. K., Guggenberger, H., Peled, S. and Pauls, J. (1999). **Functional imaging of the monkey brain.** *Nature Neuroscience*, 2(6), 555–562.

Mack, M. L., Gauthier, I., Sadr, J. and Palmeri, T. J. (2008). **Object detection and basic-level categorization: sometimes you know it is there before you know what it is.** *Psychonomic Bulletin and Review*, *15*(1), 28–35.

Mayer, E. and Rossion, B. (2013). Prosopagnosia. In O. Godefroy (Ed.), *The Behavioral and Cognitive Neurology of Stroke* (pp. 316–335). Cambridge University Press.

McCarthy, G., Puce, A., Gore, J. C. and Allison, T. (1997). **Face-specific processing in the human fusiform gyrus.** *Journal of Cognitive Neuroscience*, *9*(5), 605–610.

Meng, M., Cherian, T., Singal, G. and Sinha, P. (2012). **Lateralization of face processing in the human brain.** *Proceedings of the Royal Society of London B: Biological Sciences*, *279* (1735), 2052–2061.

Mineault P. (2011). **Auto gaussian and gabor fits.** http://www.mathworks.com/matlabcentral/fileexchange/31485. MATLAB Central File Exchange. Retrieved August 12, 2012.

Miyake, S. and Fukushima, K. (1984). **A neural network model for the mechanism of feature-extraction. A self-organizing network with feedback inhibition.** *Biological Cybernetics*, *50*(5), 377–384.

Moeller, S., Freiwald, W. A. and Tsao, D. Y. (2008). **Patches with links: a unified system for processing faces in the macaque temporal lobe.** *Science (New York, NY), 320*(5881), 1355–1359.

Moscovitch, M., Winocur, G. and Behrmann, M. (1997). **What Is Special about Face Recognition? Nineteen Experiments on a Person with Visual Object Agnosia and Dyslexia but Normal Face Recognition.** *Journal of Cognitive Neuroscience, 9*(5), 555–604.

Nahm, F. K., Perret, A., Amaral, D. G. and Albright, T. D. (1997). **How do monkeys look at faces?** *Journal of Cognitive Neuroscience, 9*(5), 611–623.

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G. and Tootell, R. B. H. (2011). **Scene-selective cortical regions in human and nonhuman primates.** *The Journal of Neuroscience, 31*(39), 13771–13785.

Ohayon, S., Freiwald, W. A. and Tsao, D. Y. (2012). **What makes a cell face selective? The importance of contrast.** *Neuron, 74*(3), 567–581.

Op de Beeck, H. P., Deutsch, J. A., Vanduffel, W., Kanwisher, N. G. and DiCarlo, J. J. (2008). **A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex.** *Cerebral cortex (New York, NY : 1991), 18*(7), 1676–1694.

Parr, L. A. (2011). **The inversion effect reveals species differences in face processing.** *Acta psychologica, 138*(1), 204–210.

Parr, L. A., Winslow, J. T., Hopkins, W. D. and de Waal, F. B. (2000). **Recognizing facial cues: individual discrimination by chimpanzees (Pan troglodytes) and rhesus monkeys (Macaca mulatta).** *Journal of Comparative Psychology (Washington, DC : 1983), 114*(1), 47–60.

Perrett, D. I., Hietanen, J. K., Oram, M. W., Benson, P. J. and Rolls, E. T. (1992). **Organization and Functions of Cells Responsive to Faces in the Temporal Cortex.** *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 335* (1273), 23–30.

Perrett, D. I., Mistlin, A. J. and Chitty, A. J. (1987). **Visual Neurons Responsive to Faces.** *Trends in Neurosciences, 9*(10), 358–364.

Perrett, D. I., Mistlin, A. J., Chitty, A. J., Smith, P. A., Potter, D. D., Broennimann, R. and Harries, M. (1988). **Specialized face processing and hemispheric asymmetry in man and monkey: evidence from single unit and reaction time studies.** *Behavioral Brain Research, 29*(3), 245–258.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J. and Thomas, S. (1991). **Viewer-centred and object-centred coding of heads in the macaque temporal cortex.** *Experimental Brain Research, 86*(1), 159–173.

Perrett, D. I., Rolls, E. T. and Caan, W. (1982). **Visual neurones responsive to faces in the monkey temporal cortex.** *Experimental Brain Research, 47*(3), 329–342.

Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. and Jeeves, M. A. (1984). **Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception.** *Human Neurobiology, 3*(4), 197–208.

Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. and Jeeves, M. A. (1985). **Visual cells in the temporal cortex sensitive to face view and gaze direction.** *Proceedings of the Royal Society of London Series B, Biological Sciences, 223* (1232), 293–317.

Peyre G (2009). **Toolbox wavelets on meshes.** http://www.mathworks.com/matlabcentral/fileexchange/17577-toolbox-wavelets-on-meshes. MATLAB Central File Exchange. Retrieved March 10, 2012.

Pinsk, M. A., DeSimone, K., Moore, T., Gross, C. G. and Kastner, S. (2005). **Representations of faces and body parts in macaque temporal cortex: a functional MRI study.** *Proceedings of the National Academy of Sciences of the United States of America, 102*(19), 6996–7001.

Pinsk, M., Arcaro, M., Weiner, K., Kalkus, J., Inati, S., Gross, C. and Kastner, S. (2009). **Neural Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study.** *Journal of Neurophysiology,* 101, 2581-2600.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C. and Kanwisher, N. (2011). **Differential selectivity for dynamic versus static information in face-selective cortical regions.** *NeuroImage, 56*(4), 2356–2363.

Puce, A., Allison, T., Gore, J. C. and McCarthy, G. (1995). **Face-sensitive regions in human extrastriate cortex studied by functional MRI.** *Journal of Neurophysiology, 74*(3), 1192–1199.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C. and Tootell, R. B. H. (2011). **The "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in Humans and Monkeys.** *PLoS Biology, 9*(4), e1000608.

Rajimehr, R., Young, J. C. and Tootell, R. B. H. (2009). **An anterior temporal face patch in human cortex, predicted by macaque maps.** *Proceedings of the National Academy of Sciences of the United States of America, 106*(6), 1995–2000.

Riesenhuber, M. and Poggio, T. (1999). **Hierarchical models of object recognition in cortex.** *Nature Neuroscience,* 2(11), 1019–1025.

Riesenhuber, M. and Poggio, T. (2000). **Models of object recognition.** *Nature Neuroscience, 3,* 1199–1204.

Rolls, E. T. and Baylis, G. C. (1986). **Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey.** *Experimental Brain Research, 65*(1), 38–48.

Rolls, E. T., Baylis, G. C. and Leonard, C. M. (1985). **Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey.** *Vision Research, 25*(8), 1021–1035.

Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L. and Azzopardi, P. (1994). **The responses of neurons in the temporal cortex of primates, and face identification and detection.** *Experimental Brain Research, 101*(3), 473–484.

Rousselet, G. A., Macé, M. J.-M. and Fabre-Thorpe, M. (2003). **Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes.** *Journal of Vision,* 3(6), 5.

Sato, T., Uchida, G. and Tanifuji, M. (2009). **Cortical columnar organization is reconsidered in inferior temporal cortex.** *Cerebral cortex (New York, NY : 1991), 19*(8), 1870–1888.

Schneidermann H (2004). **Feature-centric evaluation for efficient cascaded object detection.** *In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (Los Alamitos, CA)*: IEEE Press.

Sergent, J. and Signoret, J. L. (1992). **Functional and anatomical decomposition of face processing: evidence from prosopagnosia and PET study of normal subjects.** *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 335* (1273), 55–61.

Sergent, J., Ohta, S. and& MacDonald, B. (1992). **Functional neuroanatomy of face and object processing. A positron emission tomography study.** *Brain, 115 Pt 1,* 15–36.

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U. and Poggio, T. (2007a). **A quantitative theory of immediate visual recognition.** *Progress in Brain Research, 165,* 33–56.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. and Poggio, T. (2007b). **Robust object recognition with cortex-like mechanisms.** *IEEE transactions on pattern analysis and machine intelligence*, 29(3), 411–426.

Sinha, P., Balas, B., Ostrovsky, Y. and Russell, R. (2006). **Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About.** *Proceedings of the IEEE*, 94(11), 1948–1962.

Spearman, C. (1904). **The proof and measurement of association between two things.** *The American Journal of Psychology*, 15(1), 72-101.

Spearman, C. (1910). **Correlation calculated from faulty data.** *British Journal of Psychology*, 3(3), 271-295.

Spiridon, M. and Kanwisher, N. (2002). **How distributed is visual category information in human occipito-temporal cortex? An fMRI study.** *Neuron*, 35(6), 1157–1165.

Spiridon, M., Fischl, B. and Kanwisher, N. (2006). **Location and spatial profile of category-specific regions in human extrastriate cortex.** *Human Brain Mapping*, 27(1), 77–89.

Tanaka, J. W. and Farah, M. J. (1993). **Parts and wholes in face recognition.** *The Quarterly Journal of Experimental Psychology. A: Human Experimental Psychology*, 46(2), 225–245.

Tanaka, K. (1992). **Inferotemporal cortex and higher visual functions.** *Current Opinion in Neurobiology, 2*(4), 502–505.

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

Tanaka, K. (2003). **Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities.** *Cerebral Cortex (New York, NY : 1991), 13*(1), 90–99.

Tanaka, K., Saito, H., Fukada, Y. and Moriya, M. (1991). **Coding visual images of objects in the inferotemporal cortex of the macaque monkey.** *Journal of Neurophysiology, 66*(1), 170–189.

Thorpe, S., Fize, D. and Marlot, C. (1996). **Speed of processing in the human visual system.** *Nature, 381*(6582), 520–522.

Tovee, M. J., Rolls, E. T. and Azzopardi, P. (1994). **Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque.** *Journal of Neurophysiology, 72*(3), 1049–1060.

Tsao, D. Y. and Livingstone, M. S. (2008). **Mechanisms of face perception.** *Annual Review of Neuroscience, 31*, 411–437.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B. and Tootell, R. B. H. (2003). **Faces and objects in macaque cerebral cortex.** *Nature Neuroscience, 6*(9), 989–995.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H. and Livingstone, M. S. (2006). **A cortical region consisting entirely of face-selective cells.** *Science (New York, NY), 311*(5761), 670–674.

Tsao, D. Y., Moeller, S. and Freiwald, W. A. (2008). **Comparing face patch systems in macaques and humans.** *Proceedings of the National Academy of Sciences of the United States of America, 105*(49), 19514–19519.

Ungerleider, L. G. and Mishkin, M. (1983). **Two Cortical Visual Systems.** In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D. and Anderson, C. H. (2001). **An integrated software suite for surface-based analyses of cerebral cortex.** *Journal of the American Medical Informatics Association, 8*(5), 443–459.

Vanduffel, W., Fize, D., Mandeville, J. B., Nelissen, K., Van Hecke, P., Rosen, B. R., Tootell, R. B. H. and Orban. G. A. (2001). **Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys.** *Neuron, 32*(4), 565–577.

Viola, P. and Jones, M. (2001). **Rapid object detection using a boosted cascade of simple features.** *IEEE Conference on Computer Vision and Pattern Recognition, Proceedings, 1*, I–1.

Viola, P. and Jones, M. (2004). **Robust real-time face detection.** *International Journal of Computer Vision. 57(2), 137-154.*

Vogels, R., Biederman, I., Bar, M. and Lorincz, A. (2001). **Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences.** *Journal of Cognitive Neuroscience, 13*(4), 444–453.

Wallis, G. and Rolls, E. T. (1997). **Invariant face and object recognition in the visual system.** *Progress in Neurobiology, 51*(2), 167–194.

Wang, G., Tanaka, K. and Tanifuji, M. (1996). **Optical imaging of functional organization in the monkey inferotemporal cortex.** *Science (New York, NY), 272*(5268), 1665–1668.

Wang, G., Tanifuji, M. and Tanaka, K. (1998). **Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging.** *Neuroscience research, 32*(1), 33–46.

Weiner, K. S. and Grill-Spector, K. (2010). **Sparsely-distributed organization of face and limb activations in human ventral temporal cortex.** *NeuroImage, 52*(4), 1559–1573.

Williams, M. A., Dang, S. and Kanwisher, N. G. (2007). **Only some spatial patterns of fMRI response are read out in task performance.** *Nature neuroscience, 10*(6), 685–686.

Yamane, S., Kaji, S. and Kawano, K. (1988). **What facial features activate face neurons in the inferotemporal cortex of the monkey?** *Experimental brain research, 73*(1), 209–214.

Yamane, S., Komatsu, H., Kaji, S. and Kawano, K. (1990). **Neural Activity in the Inferotemporal Cortex of Monkeys During a ace Discrimination Task.** In E. Iwai & M. Mortimer (Eds.), *Vision, Memory, and the Temporal Lobe* (pp. 89–100). New York: Elsevier.

Yin, R. K. (1969). **Looking at upside-down faces.** *Journal of experimental psychology, 81* (1), 141–145.

Young, M. P. and Yamane, S. (1992). **Sparse population coding of faces in the inferotemporal cortex.** *Science (New York, NY), 256*(5061), 1327–1331.