Cortical spatiotemporal plasticity in visual category learning

Yang Xu

August 2013
CMU-ML-13-110

Carnegie Mellon

Cortical spatiotemporal plasticity in visual category learning

Yang Xu

August 2013
CMU-ML-13-110

School of Computer Science
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Robert E. Kass (co-chair)
Michael J. Tarr (co-chair)
Aarti Singh
Avniel S. Ghuman (University of Pittsburgh)

Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

This research was sponsored by the National Institutes of Health under grant numbers R01EB005847, R01MH064537, R01MH06453704, and R90DA023426. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Keywords: visual category learning, visually-similar categories, cortical spatiotemporal plasticity, ventral stream, prefrontal cortex, face perception network, hot spots, source localization, MEG

Abstract

Central to human intelligence, visual categorization is a skill that is both remarkably fast and accurate. Although there have been numerous studies in primates regarding how information flows in inferiortemporal (ITC) and prefrontal (PFC) cortices during online discrimination of visual categories, there has been little comparable research on the human cortex. To bridge this gap, this thesis explores how visual categories emerge in prefrontal cortex and the ventral stream, which is the human homologue of ITC. In particular, cortical spatiotemporal plasticity in visual category learning was investigated using behavioral experiments, magnetoencephalographic (MEG) imaging, and statistical machine learning methods.

From a theoretical perspective, scientists from work on non-human primates have posited that PFC plays a primary role in the encoding of visual categories. Much of the extant research in the cognitive neuroscience literature, however, emphasizes the role of the ventral stream. Despite their apparent incompatibility, no study has evaluated these theories in the human cortex by examining the roles of the ventral stream and PFC in online discrimination and acquisition of visual categories. To address this question, I conducted two learning experiments using visually-similar categories as stimuli and recorded cortical response using MEG—a neuroimaging technique that offers a millisecond temporal resolution. Across both experiments, categorical information was found to be available during the period of cortical activity. Moreover, late in the learning process, this information is supplied increasingly in the ventral stream but less so in prefrontal cortex. These findings extend previous theories by suggesting that the ventral stream is crucial to long-term encoding of visual categories when categorical perception is proficient, but that PFC jointly encodes visual categories early on during learning.

From a methodological perspective, MEG is limited as a technique because it can lead to false discoveries in a large number of spatiotemporal regions of interest (ROIs) and, typically, can only coarsely reconstruct the spatial locations of cortical responses. To address the first problem, I developed an excursion algorithm that identified ROIs contiguous in time and space. I then used a permutation test to measure the global statistical significance of the ROIs. To address the second problem, I developed a method that incorporates domain-specific and experimental knowledge in the modeling process. Utilizing faces as a model category, I used a predefined "face" network to constrain the estimation of cortical activities by applying differential shrinkages to regions within and outside this network. I proposed and implemented a trial-partitioning approach which uses trials in the midst of learning for model estimation. Importantly, this renders localizing trials more precise in both the initial and final phases of learning.

In summary, this thesis makes two significant contributions. First, it methodologically improves the way we can characterize the spatiotemporal properties of the human cortex using MEG. Second, it provides a combined theory of visual category learning by incorporating the large time scales that encompass the course of the learning.

Acknowledgments

Completion of this dissertation was made possible with guidance and support from many people at CMU and elsewhere.

Rob Kass has been my academic advisor ever since I came to CMU. Since the start of my PhD, Rob has highlighted MEG as a promising tool for cognitive neuroscience and directed me toward the central statistical problems in MEG imaging, which became a critical component of my dissertation. I thank Rob for sharing his thoughts and philosophy on statistics, for constantly reminding me not to be misguided by results that come from imprudent statistical practice, and for promoting me to do better science with methodological innovations. I am grateful that Rob has always met my setbacks with encouragement and support. I am most thankful that Rob has countered my complacency with criticality and given me clear reasons to improve.

Mike Tarr is my academic co-advisor who has guided me on the core scientific part of my dissertation. I thank Mike for emphasizing the importance of theory, for striving for brevity while retaining sharpness, and for being so generous and resourceful in sharing his ideas and knowledge about human vision and many other aspects of science and life. I am especially thankful for the time Mike has devoted to my projects, including his midnight email responses, despite his multiple responsibilities in the departments and to the school.

I would like to thank two other members of my thesis committee. In particular, I thank Avniel Ghuman for his many interesting ideas and suggestions about my PhD projects, for his shared passion in MEG and for his help with my experiments at UPMC. I thank Aarti Singh for critiquing the statistical issues in my analyses, for providing insight on my presentations and for being extremely efficient.

I thank Chris D'Lauro and John Pyles for collaborating with me on projects that became important parts of my dissertation. I thank Gus Sudre for sharing with me his knowledge and expertise of all aspects of MEG imaging, which made my data analysis an enjoyable process instead of a painful one. I thank Anna Hegedus for helping with the Sparrowhawk server that made my large-scale data analyses possible. I also thank Wei Wang and Doug Weber from Pitt for our early collaborations, during which I was acquainted with MEG imaging and experimentation.

I would like to give my special thanks to Charles Kemp for introducing me to cognitive science and computational modeling, which allowed me to consider my PhD research from

different perspectives. I am grateful that Charles has been such a creative, stimulating and considerate advisor and collaborator on projects that helped shape my interests and broaden my knowledge and skill set.

I thank all members of the Kass lab for contributing feedback on my research, especially Lucia Castellanos, with whom I shared many valuable discussions about my work, and Spencer Koerner for critiquing the statistical methodologies and helping me improve the titles of my chapters. I thank all members of the Tarrlab and VisCog group, especially Carol Jew, Deb Johnson and David Munoz for helping with my experiments and data collection.

I thank faculty and staff members at the UPMC, and particularly Theodore Amdurs, Anna Haridis and Erika Laing, who helped me with my MEG experiments and recordings. I thank friends and colleagues at MLD and CNBC and a vibrant community that promotes interdisciplinary research. I especially thank Diane Stidle for her efficiency and kindness in scheduling all matters related to my PhD dissertation and defense. I thank Tyler Rice for proofreading this dissertation.

Finally, I thank my parents, who withstood countless hours of discussion with me about my PhD research and life in general. I am extremely grateful for their love, guidance and encouragement. And I thank my wonderful wife for her love and support, which helped me to persevere in my scientific pursuits without hesitation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Humans constantly interpret and analyze their environment, and we do so largely by sorting perceptual inputs into meaningful categories. Among all, the ability to discriminate between objects in the visual environment, or *visual categorization*, is critical to human intelligence [4, 5]. For example, an infant is able to distinguish her mother from a stranger based on facial cues [6]. A toddler learns word meanings by grouping objects with similar visual shapes [7]. More generally, this visual ability is fundamental in guiding how we behave, reason about objects and their properties, and form complex concepts and knowledge [8].

Despite the enormous diversity of objects [9], human visual categorization is remarkably fast and accurate [10]. Essential to efficient visual categorization is *visual category learning*, the process of obtaining categorical perception of novel visual stimuli. Infants exhibit an innate preference for novel visual stimuli—they stare at them for longer periods than they do familiar stimuli [11]. However, their ability to perceive objects as categories matures over the course of development [12, 13], suggesting that learning refines visual categorization and extends the repertoire of things that can be visually discriminated. Importantly, previous experiments have indicated that the acquisition of categories is an *abstraction* process as opposed to memorization [5]. For example, using random dot patterns as visual stimuli, Posner and others [14–16] have shown that, according to central tendencies, or category prototypes [17], humans correctly categorize new exemplars in addition to the exposed ones.

However, the neural mechanisms that support abstraction of novel visual categories remain poorly specified. The critical question I aim to address in this thesis is this: how

does the human cortex support the acquisition and rapid discrimination of visual categories? Specifically, I focus on characterizing spatiotemporal changes in the cortex—*cortical spatiotemporal plasticity*—associated with learning of novel visual stimuli. Expanding upon previous research that emphasizes an understanding of either the cortical substrates (spatial aspect) or the time course (temporal aspect) of visual categorization, I have developed a thesis framework that helps to characterize the *spatiotemporal* properties of the cortex over the course of learning. Using this framework, I aim to address some key theoretical and methodological questions.

Theoretically, it remains unclear which part of the cortex is crucial to the acquisition and discrimination of visual categories, particularly when they appear perceptually similar. Cognitive neuroscientists have traditionally considered the *ventral stream* or the ventral visual pathway (VVP) [18] in the human cortex to be the dominant site for representing visual categories, but there is no consensus in its role in acquiring novel categories that differ subtly in feature space (e.g. [19, 20]). Moreover, we understand little about how the ventral stream encodes visual categories in the rapid discrimination time course, and how its role compares with other cortical regions. In particular, scientists working on non-human primate brains have demonstrated how information emerges in *prefrontal cortex* (PFC) [21–23] in the rapid discrimination time course of visually similar categories, hence positing its dominant role in category representation. Despite the apparent incompatibility of these theories, no study has compared fine-grained time course of the ventral stream and PFC during visual category learning. In this thesis, I investigate how category information flows in these regions during visual discrimination and over the course of learning. I propose the view that the ventral stream is crucial to the encoding of visual categories in the long term, but PFC encodes categories jointly with the ventral stream during the early stage of learning.

Methodologically, characterizing fine-grained cortical time course is challenging from two perspectives. From an experimental perspective, functional magnetic resonance imaging offers poor temporal resolution and cannot capture the sub-second cortical dynamics in visual discrimination. On the other hand, while multi-electrode arrays (or electrocortigography) offer precise spatiotemporal recordings, they are not sufficient to monitor neural activities at the scale of the human cortex. Instead, I used magnetoencephalography or *MEG*, a brain imaging technology with superior temporal resolution and a substantial coverage of the cortex, to record human cortical activities during visual category learning. To allow for sus-

tained learning, I designed two learning experiments with novel *visually-similar categories*. To study cortical changes concomitant with learning, I used a trial-to-trial, feedback-driven online learning paradigm that allows cortical activities to be recorded with minimal intervention and delay.

From a technical perspective, MEG imaging suffers from high dimensionality of data and coarse reconstruction of cortical activities—a process called *source localization*. To alleviate these problems, I developed two statistical methods that incorporate experimental and domain-specific knowledge—in contrast to generic approaches that do not explicitly leverage these information in the modeling process. To reduce false discoveries in high dimensional data, I used an excursion algorithm to find spatially and temporally contiguous regions of interest (or *hot spots*) in the cortex. To characterize the statistical significance, I applied a permutation test that measures a global p-value for the discovered hot spots. In the second method, using faces as an extensively studied visual category, I showed that incorporating knowledge about a cortical "face" network helps to increase the precision in reconstructing cortical source activities. Furthermore, given the unique structure of the learning experiment, I proposed a trial-partition approach. I showed that using trials in the middle of learning for model estimation helps to improve the accuracy of source localization in the initial and final stages of learning.

Figure 1.1 illustrates the outline of my thesis. In Chapter 2, I provide background information on the cortical basis of visual category learning. I then describe the thesis framework from both theoretical and methodological perspectives and summarize my contributions. The remainder of this thesis is comprised of two main parts: one scientific, the other methodological. In the scientific section, which comprises Chapters 3 and 4, I describe two main experiments and the accompanying data analyses. In Chapter 3, I present the first experiment where participants learned to discriminate two visually-similar object categories. In Chapter 4, I present the second experiment where participants learned two face categories of greater complexity. In the methodological section, which comprises Chapters 5 and 6, I describe two statistical methods that I developed for MEG data analysis. In Chapter 5, I present a generic method that characterizes cortical spatiotemporal hot spots. In Chapter 5, I present a more specific method for source localization tailored to the face network. In Chapter 7, I draw conclusions and discuss limitations and possible extensions of my thesis.

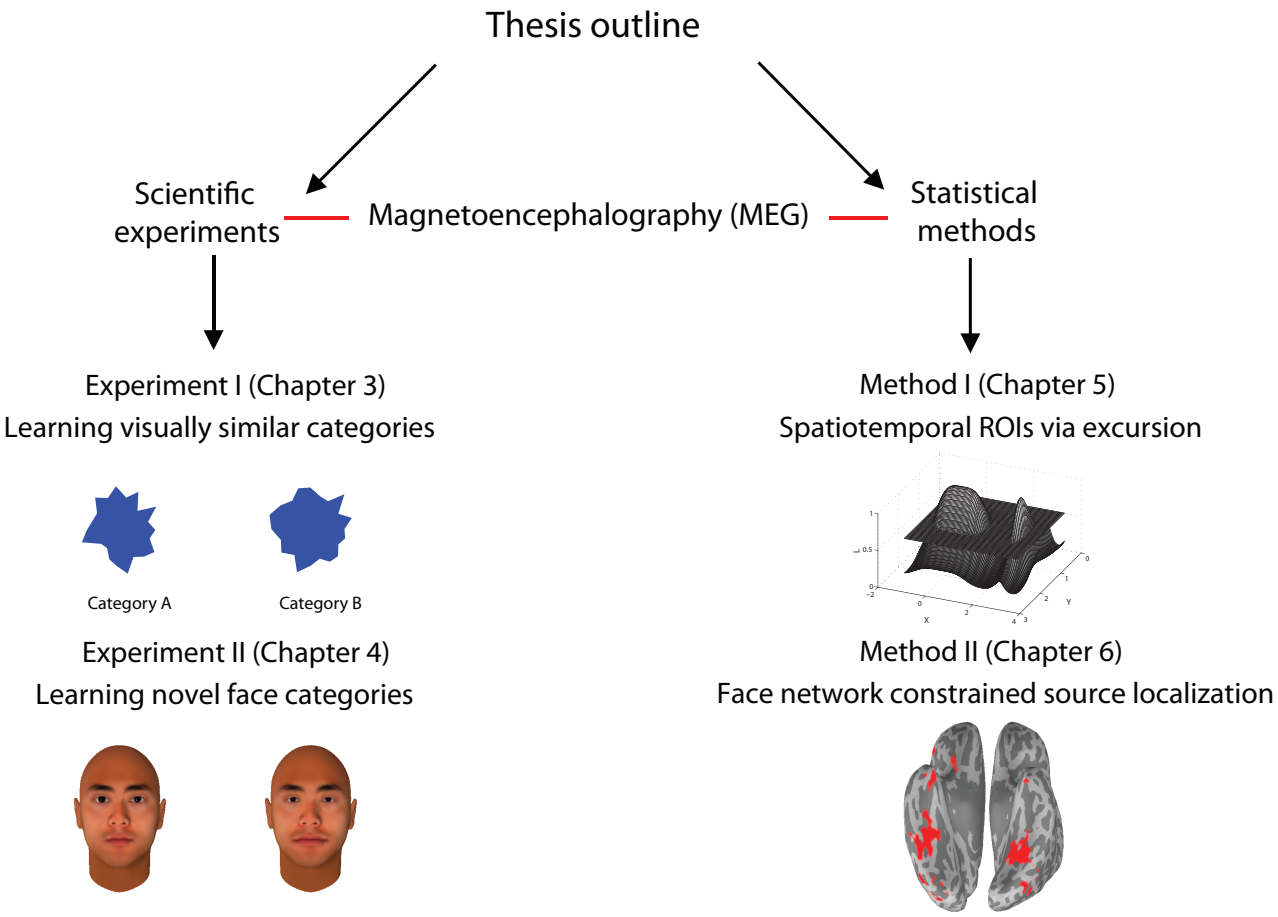

Figure 1.1: **An outline of the thesis.**

Chapter 2

A spatiotemporal framework for visual category learning

Previous research has suggested that the ventral stream, or the ventral visual pathway, in the human cortex serves as the main territory for encoding visual object and faces [18]. Evidence of this typically takes either a spatial perspective (cortical substrates) or a temporal perspective (time course), so it is uncertain how the cortex encodes visual categories spatiotemporally, and more specifically how categories emerge during the course of learning. Over the past decade, however, researchers have explored the potential of analyzing fine-grained time course in the primate cortex. These studies have proposed that prefrontal cortex plays a more dominant role in encoding visual categories than the inferiortemporal cortex (a homolog to the human ventral stream) [21, 22], hence incompatible with common beliefs about the ventral stream. Methodologically, it is desirable to characterize information flow in the human cortex during visual categorization with good temporal resolution. In this chapter, I present a thesis framework that addresses both of these issues. Scientifically, we can better evaluate theoretical proposals about the ventral stream and prefrontal cortex with a spatiotemporal approach to visual categorization. By characterizing spatiotemporal plasticity in visual category learning, we can produce a fuller picture of cortical functions as opposed to accounts that emphasize the end point of learning. Meanwhile, scientists are uncertain how to utilize existing brain imaging technologies to pursue these goals with precision. Statistics and machine learning provide powerful tools for this purpose.

2.1 The cortical basis of visual category learning

2.1.1 Encoding and acquisition of visual categories in the ventral stream

The ventral stream, traditionally referred to as the "what" pathway in the human brain, is also known as the ventral visual pathway (VVP) [18, 24, 25]. This lengthy region of the cortex begins with the primary visual cortex. It then passes the ventral fusiform and inferior temporal regions, and ends near the anterior pole of the temporal lobe. It is unquestionable that the visual cortex plays a role in the processing of visual inputs that are delivered from the retina and further relayed from the lateral geniculate nucleus, e.g. [26, 27]. Computational models have also proposed mechanisms in the ventral stream for visual recognition and categorization [28, 29]. Furthermore, neurophysiological studies have suggested that the ventral stream encodes visual stimulus categories—almost certainly with common, familiar objects [18]. But what is not entirely clear is whether such encoding is due to representation of categories *per se* or whether it is due to sensitivity to visual features.

There is strong evidence for object-based coding in the ventral stream from brain lesion studies. For example, lesions to the fusiform gyrus and the junction of occipital and temporal lobes caused deficits in visual recognition [30–32]. These observations suggest that territories in the ventral stream are directly associated with representation of objects used for visual recognition.

More relevant studies in functional magnetic resonance imaging (fMRI) have shown that the ventral stream supports encoding of common visual categories such as faces [33], places [34, 35], bodies [36], letters [37], tools [38] and animals [38]. Other studies using multivariate pattern analysis have revealed that common visual categories such as faces, houses and chairs can be "decoded" from the blood-oxygenation-level-dependent (BOLD) responses [39].

On separate lines, event-related potential (ERP) and magnetoencephalography (MEG) studies have shown that waveforms at early windows such as *N170* and *M170* (approximately $170msec$ post-stimulus) during the recognition time course are more selective for face cat-

egories than other objects (e.g. [40]). Although the sources of these waveforms cannot be precisely identified, they are typically reported to be in the vicinity of occipitotemporal regions [41], suggesting category coding may occur in the ventral stream during the visual recognition time course.

Despite the above evidence, other works have suggested that visual category learning has little impact on the ventral stream, e.g. [20]. Specifically, they challenge the view that the ventral stream is capable of obtaining discriminability toward novel visual categories that differ in relevant feature dimensions. For example, Jiang and colleagues [20] found that there is no enhancement in neural representations around the category boundary in visual regions near the fusiform and lateral occipital cortex (LOC). Others have also failed to attribute a significant increase in category discriminability in object-sensitive regions of the visual cortex to category learning [42, 43].

Folstein et al. [19], however, have proven acquired discriminability exists in the fusiform area and have theorized that the failure to observe category discriminability in the ventral stream may be attributed to insufficient category learning, such that category performance at the behavioral level is either inadequate or unconfirmed. In sum, there is no consensus on the role of the ventral stream in acquiring novel, visual categories that differ in subtle feature dimensions. A key factor, as pointed out by [19, 44], is that most studies in this debate fail to address whether the visual categories, which are often artificially designed, are in fact delineated by clear boundaries. Moreover, no study has comprehensively examined how fine-scale temporal properties of the ventral stream evolve during the learning of visually similar categories. By examining the time course of the ventral stream during visual categorization—particularly to determine whether such a time course embeds information about visual categories—one can provide direct means towards understanding the role of the ventral stream in visual category discrimination and acquisition.

2.1.2 The ventral stream and prefrontal cortex

Most studies that challenge the role of the ventral stream in visual categorization use morphspace category designs [20–22]. Morphspace is created using techniques from computer graphics. “Categories” are defined by first generating images via a blending of a number

of prototype images—thus creating a continuum of morphs—and the new images are then separated by being divided down the middle of the continuum, forming the category "boundaries". These resulting category boundaries are often difficult to validate. In particular, it is unclear whether they are perceptually separable due to the arbitrariness in the morph continuum, so the generated categories can be highly confusable [44].

Nevertheless, a major finding from these studies is that prefrontal cortex (PFC), in both primates and humans, is more dominant in coding the morphed visual categories than the ventral stream, hence portraying a *PFC-dominant view* in visual categorization [20–22, 45]. In particular, Freedman et al. [21] compared prefrontal cortex to inferiortemporal cortex (ITC), a primate homolog of VVP, in a delayed-match-to-sample task. In that task, a monkey is first given a sample image of a computer-generated cat or dog, and is asked after a delay period to judge whether a second, test image (either a cat or dog) belongs to the same category as the sample. There they found that PFC in well-trained monkeys carries as much information about visual categories as ITC during the recognition time course. But more so, by using images in the middle of the morphspace (or "category" boundaries), they created an experimental condition where a sample stimulus can be "conceptually" closer to one category yet "perceptually" closer to another. They used such a condition to test whether PFC is better than ITC at categorical abstraction, such that despite the samples appear visually distinct, they should still be considered as a member of the same category conceptually, or *vice versa*. With this paradigm, they found that PFC is much better than ITC in encoding these "abstract" categories, whereas ITC is much more confusable about the true category memberships of these samples. They concluded that PFC dominates in categorical representation, whereas ITC is sensitive to perceptual features in the visual stimuli but, as such, does not encode categories.

There are several issues with the described paradigm. Most notably, the fact that morphspace contains samples conceptually closer to one category yet perceptually closer to another implies the designed category boundary is arbitrary—meaning that the notion of visual categories here is ill-defined. For example in the real world, it is extremely unusual to find a cat that perceptually looks closer to dogs than to cats. If that is the case, it is more likely that one would categorize that cat as a dog. However, monkeys in the experiment by Freedman et al. [21] have been repetitively trained for the task, and it is possible that certain rules are employed in categorizing samples at the morph boundary; and previous

studies have shown that rule-based categorization does invoke PFC [46–48]. But this is not equivalent to saying that PFC dominates representation of categories—yet more to do with rule formation. Moreover, Freedman et al. [21] found no evidence that PFC is superior to ITC in coding sample stimuli that are far apart in the morphspace, which is supportive of the idea that ITC codes visual categories.

Contrary to the PFC-dominant view, a recent work by Krigolson et al. [49] has suggested that the ventral stream is sensitive to categories even when the visual stimuli are highly similar, e.g. at subordinate-level. Using artificial blob-like stimuli that differ only in the edge contours, Krigolson et al. [49] have found that ERP waveform at posterior-ventral regions in the human cortex show significant modulation at around $250msec$ post-stimulus ($N250$) for well-learned subjects. Because the visual stimuli were designed to be highly resemblant, their finding contradicts the study by Freedman et al. [21] that the ventral stream cannot distinguish subordinate-level categories. Still, however, no direct comparison has been made between the ventral stream and PFC in that study, so there is no direct evidence that PFC does not encode these blob categories equally as well.

In addition, researchers have proposed in a prominent MEG study of human visual recognition a competing *PFC-complimentary view* [24, 44, 50–52] of the PFC-dominant proposal. By investigating the rapid time course in PFC and the ventral stream (near fusiforms) during visual recognition, Bar et al. [52] have found that PFC precedes the ventral stream in the time course, hence proposing that PFC provides top-down guidance (as opposed to dominates) in inferring the object identity. Such findings do not deny the role of PFC in visual recognition, but they do suggest that PFC and the ventral stream coordinate rather than one dominating the other. However, their paradigm has focused on visual recognition as opposed to categorization, so it does not provide direct evidence for a complementary view of PFC in visual categorization.

As yet, no comprehensive study comparable to those in primates [21, 22] has been conducted to compare the respective roles of the human ventral stream and prefrontal cortex in visual category learning. It is particularly unclear how these cortical regions support the emergence of novel visual categories both in the time course of visual discrimination and during the course of learning, so we cannot fully evaluate theoretical proposals about whether PFC dominates or complements in visual categorization.

2.1.3 Faces as a model visual category

Even if it were confirmed that the ventral stream supports visual categorization, a related fundamental question is how visual categories are encoded in the ventral stream. One prominent proposal is that the ventral stream is hierarchically organized [24, 27, 53] such that it mediates a distributed representation of visual categories or identities with increasing complexity from low to high order areas. However, it remains unclear how such a hierarchical scheme is reflected in spatiotemporal dynamics of the ventral stream. An important goal of this thesis is to probe fine-grained spatiotemporal properties of the ventral stream during discrimination of faces—a relatively well-documented model visual category.

The cortical basis of face perception has been studied extensively in the literature [54, 55], but the nature of cortical mechanisms involved in face representation is far from clear. Kanwisher et al. [33] found in an early study that the ventral stream supports representation of faces. By comparing the BOLD response while participants viewed faces or objects, they found that the mid-fusiform (mFus) area (or the "fusiform face area" or FFA) is significantly more activated for faces than objects. With this result, they proposed that the middle fusiform area in the cortex is crucial for face processing [18, 33]. Later works by Tarr and colleagues [56] have suggested that the middle fusiform area also serves general recognition purposes and that the reason for its strong activation for faces is the unleveled sophistication in face recognition compared with other categories (e.g. faces are typically perceived and recognized at the individual level, whereas very few other object categories require discrimination at this fine level). In particular, they trained participants to recognize novel artificial objects called "greebles" and found that well-learned subjects exhibit increase activation in mFus. At the same time, theey found mFus activation toward faces decreases toward the end of learning, suggesting there is a competition for resources in mFus from greebles and faces. Haxby and colleagues [39] have further demonstrated that cortical regions that code for faces, chairs and houses are highly overlapped and distributed, and face-related ventral areas may serve general recognition purposes.

In more recent works, scientists have located a distributed cortical network for face processing. In particular, Nestor et al. [57] has shown that in addition to mid-fusiform, patches anterior to FFA and close to the temporal pole code for face identities. Avidan et al. [58]

also confirms such findings in primates and proposes that the connection between anterior inferiotemporal region (aIT) and posterior regions may be crucial for face recognition. In addition, Pitcher and colleagues also find that the occipital face area near the inferior occipital gyrus [59, 60] is crucially related to the processing of facial parts. This piece of work sheds light on the cortical substrates underlying face perception, although it is not yet understood how the discovered face network represents faces during a rapid discrimination time course (see [61]), or how novel face identities become represented in the first place. Elucidating these issues especially in the context of faces as a model category will provide new venues into the nature of category representation in the ventral stream.

2.2 The thesis framework

With the overall framework of this thesis, I aim to understand how the human cortex supports the acquisition and discrimination of visual categories, with a particular emphasis on characterizing cortical spatiotemporal properties for learning visually-similar categories. To do so, I designed and conducted two scientific experiments that involved supervised learning of novel visually similar object and face categories. I used magnetoencephalography (MEG)—an imaging tool with superior temporal resolution—to record cortical responses during these visual category learning experiments, and developed statistical machine learning methods that help to improve the precision of spatiotemporal MEG-based analysis. In this section, I describe each aspect of the thesis framework in turn.

2.2.1 Learning visually-similar object categories

As described in Section 2.1.1 and 2.1.2, despite the common belief that the ventral stream encodes visual categories, there is less consensus on whether it encodes categories that appear visually similar, particularly whether it acquires such categorical discriminability from category learning. It also remains to be determined whether prefrontal cortex plays a predominant role in visual categorization (and learning) in the human cortex. My primary goals for the first experiment (see Chapter 3 for the detailed analysis) are hence twofold: 1) to test the hypothesis that the ventral visual pathway is capable of acquiring novel visual

categories and 2) to evaluate the competing theories that the prefrontal cortex dominates or complements the ventral visual pathway in visual categorization.

To achieve these goals, I designed two novel, blob-like object categories that bear a high degree of perceptual similarity (see [49] for a similar design). Differed from [49], which does not specify a category boundary for the stimuli, I generated these blobs from pre-defined distributions centered around two prototypes and ensured the resulting categories have a distinct boundary in the design space. Within each category, I used a large number of non-repeated samples in order to ensure category learning would be due to an abstraction as opposed to memorization of the visual stimuli. To promote learning of this stimulus set, I used a trial-to-trial supervised learning paradigm where subjects are given a feedback of their response based on their judgement about the category affiliation of a given visual stimulus. The idea is that initially in learning, subjects would not be aware of the category memberships of the stimuli and simply guess about these in their responses. Over time, however, they should learn the mapping of the visual stimuli to the categories. I recorded cortical response with MEG throughout this learning process, and by comparing the initial and final stages of learning I tested the proposed hypotheses about the ventral stream and PFC.

To characterize the cortical encoding of visual categories, I used discriminability (i.e. a contrast between two categories) as an approximate measure and compared this discriminative information flow in ventral visual and prefrontal regions. I was then able to assess whether category-discriminative information becomes more prominent in the ventral visual pathway toward the end of learning (an indicator for category acquisition) and whether prefrontal cortex contains more or less information about stimulus categories in time course than the ventral visual pathway.

2.2.2 Learning novel face categories

In the second experiment (see Chapter 4 for the detailed analysis), I used a similar experimental paradigm in order to strengthen the findings from the first experiment. Critically, it involves the learning of two visually similar face categories instead of object-based stimuli. I created these face categories based on prototypes of two faces that differ only in designated features around the eyes and the mouth.

Because faces have been extensively studied in the literature, I was able to 1) use existing knowledge about cortical substrates of face perception to develop a method that improves the precision for spatiotemporal analysis in a cortical face processing network, and 2) test hypotheses about the spatiotemporal dynamics of face coding in this network at a fine scale. More specifically, by zooming into regions best known for face processing, I evaluated—in the rapid face discrimination time course—whether temporal encoding of faces is mediated hierarchically along the ventral visual pathway, with regions downstream (e.g. OFA or IOG) containing less information about face identities than those upstream (e.g. aIT). I was also able to investigate how this coding scheme evolves over the course of learning when subjects become more accurate in recognizing the face identities. Finally, based on pre-defined regions of interest, I conducted a time-frequency analysis that helped to elucidate how region-region interactions may contribute to successful discrimination of face identities.

2.2.3 Magnetoencephalography (MEG)

I used magnetoencephalography (MEG) to record cortical activities for the described visual category learning experiments. I chose MEG for several reasons. Firstly, MEG offers a *millisecond* temporal resolution that is far superior to that of existing technologies such as fMRI, which has a temporal resolution in the order of seconds. This precision is crucial to a fine-grained analysis for the rapid cortical time course during visual categorization. Second, MEG allows recordings to be conducted at the cortical level, whereas imaging technologies such as electrocortigography (ECoG) offer equally good temporal resolution but highly limited cortical coverage. And finally, MEG is less susceptible to electrical interference due to current conduction through the scalp (unlike electroencephalography, or EEG), and it is a non-invasive technology that is suitable for experiments involving human participants. In this section, I briefly discuss the biophysical principles of MEG.

Magnetoencephalography measures the magnetic fields generated from brain activities and was first invented by Cohen in 1968 [62]. Specifically, MEG records the temporal activities of populations of neurons from superconductive sensors called SQUIDS [63–65]. This is a sensitive equipment that allows extremely weak magnetic fields near the order of 10^{-18} $Tesla$ to be detected (neurons generate magnetic fields near the order of 10^{-13} $Tesla$). At the same time, they are also susceptible to ambient noise such as power supplies or heart

beats. Because the magnetic field strength decreases rapidly with an increasing distance from the source, MEG is better suited for detecting activities from the cortical surface than from deep, subcortical structures, hence it is ideal for investigating the cortical dynamics.

Neurons in the brain constantly generate electric currents. These currents are typically due to ionic flows across the membranes at the cellular level. Differing from action potentials, which are far more rapid, slower postsynaptic potentials contribute most to MEG signals. Depending on the type of neurotransmitters involved, these potentials can be either excitatory (EPSP) or inhibitory (IPSP). Due to the interaction between the EPSPs and IPSPS—EPSP constituting a sink (attracting electrons) and IPSP constituting an active source (electron flows)—the resulting current flows in and out of the membrane at synapses, which in turn creates a magnetic field. The source–sink configuration induces a coherent magnetic field that is referred to as a current dipole. These are the main components detected by the SQUID sensors.

Since no overall magnetic field is detected from radial sources, only magnetic fields at certain orientations can be picked up. These are typically tangential sources generated from pyramidal neurons, which have long dendrites and are situated perpendicularly to the cortical surface (hence why they generate magnetic fields that are tangential to the cortical surface). As each individual neuron generates an infinitesimal magnetic field, MEG can only detect synchronous activities from a large population of neurons (e.g. 10^5). Because magnetic fields travel much faster than the synaptic currents, activities in the order of milliseconds can be easily detected by the SQUID sensors, hence offering a temporal resolution close to the dynamics of neurons.

2.2.4 Methodological challenges and solutions

Despite the superior temporal resolution of MEG, this technology has its fundamental limitations. The two major challenges in quantifying cortical spatiotemporal properties with MEG are *high dimensionality of data* and *low precision in source localization*. One of my primary aims with this thesis is to develop statistical methodologies that alleviate these problems in order to make MEG-based scientific analysis more precise

The high temporal resolution of MEG produces thousands of data points in a given trial. Multiplying the temporal dimension with hundreds of sensor locations, the spatiotemporal

dimension of MEG data is immense, and the problem is exacerbated in the source space given there can be thousands of source locations. High dimensionality not only places computational burdens on the analysis, but more critically it also creates statistical issues in identifying regions of interest across time and space. A sea of spatiotemporal data is highly susceptible to spurious and accidental discoveries of regions of interest due to chance alone, so it is imperative to apply some statistical procedures to prevent false discoveries. Conventional methods such as multiple-comparison corrections can help alleviating the issue, but a caveat in these approaches toward spatiotemporal data is that they do not necessarily enforce contiguity. As a consequence, retrieved ROIs can be located sporadically over the cortical space and along the time axis, making it difficult to interpret the findings. Additionally, results from functional imaging and electroencephalographic studies have suggested that cortical neuronal populations generally activate in clusters instead of discretely, and that they generate continuously more so than in bursty waveforms, so we should incorporate some continuity constraints in the ROI identification process. Lastly, once these ROIs are located, there should be a principled way to quantify their statistical significance. If multiple ROIs are involved, a global p-value is required to summarize the overall significance.

The first method I have developed in this thesis tackles both problems of contiguous ROI discovery and quantification of global significance in the vehicle of a spatiotemporal *excursion* method (see Chapter 5 for details). An excursion can be conceived as a cut in space and time based on a certain threshold, e.g. a pre-defined statistic or α-level. To enforce contiguity, I use a clustering algorithm that groups cortical sources contiguous in space and time together. To reduce spurious possibilities, I rank these grouped regions based on a summary of their statistics, attending to ones that have large summary statistics and discarding those that have small negligible statistics. As a validation step, I repeat the excursion procedure many times with shuffled data (the null distribution), and characterize the ROIs from excursion with a global p-value based on a permutation test.

A more fundamental limitation to MEG imaging is that the brain signals are measured from sensors located above the scalp. Consequently, instead of being read directly off the recordings, cortical activities have to be reconstructed from the sensors through a process called *source localization*. Since there are many more possible cortical sources than sensors available, the reconstruction is mathematically highly under-constrained and is intrinsically ill-posed—solutions to this problem are non-unique which leave much uncertainty in the

estimated source activities. This is typically referred to as the *inverse problem*. To reduce the errors in reconstruction, one must impose additional constraints in the source localization procedure.

The second method I have developed in this thesis addresses the inverse problem. I take an alternative approach to most existing source models that use a generic algorithm (see Chapter 6 for details). Instead, I apply *domain-specific and experimental constraints* in a method specifically tailored to localize source activities for the face category learning experiment. Using faces as a model category, I design an independent localizer experiment commonly used in functional imaging to define a cortical network in face perception with MEG. This domain-based prior allows me to constrain the source localization model both spatially and temporally. And with the unique structure of the category learning experiment, I am able to use trials for different purposes. Specifically, I partition trials based on whether they occur in the initial, middle or final stages of learning. Utilizing trials in the midst of learning, I obtain a reliable estimate of a spatiotemporal source model constrained by the pre-defined face network. Using trials at both ends of learning, I verify the effectiveness of the source model on completely held-out data. This *trial-partition* approach allows source activities during earliest and latest stages of learning—parts that are most relevant to scientific hypothesis testing—to be reconstructed with better spatiotemporal precision.

2.3 Contributions

My overall contribution in this thesis is in developing a framework for characterizing spatiotemporal properties of the human cortex in visual category learning. In particular, I contribute to the theoretical findings, experimentation and statistical methodologies.

Theoretical findings

1. I confirmed, during rapid visual categorization time course, that the ventral stream or the ventral visual pathway (VVP), is capable of encoding visually-similar categories and acquiring category discriminability over the course of learning.

2. I evaluated two prominent yet competing theories about the functional roles of the ventral visual pathway and prefrontal cortex in visual category learning. I suggested that

these diverging theories may arise from discrepancies in the design of visual categories and in the nature of visual stimuli. I also extended these existing theories by suggesting the VVP and PFC play differential roles in visual category learning. Specifically, the ventral stream supports long-term encoding of visual categories, but PFC jointly encodes categories mostly during the initial stage of learning.

3. I verified that the encoding of identity-based visual categories is hierarchically organized with category information becoming increasingly prominent from the posterior to the anterior part of the ventral stream in the time course. This finding confirms previous theories on a hierarchical architecture in the ventral stream. However, I suggested that such hierarchical coding is not strictly enforced but it can be altered in adaptation to shifting strategies in visual categorization.

Experimentation

1. As an alternative to morphspace designs that do not yield clear boundaries, I designed novel and visually similar stimulus categories that have distinct category boundaries.

2. I implemented an online visual category learning paradigm that facilitates rapid trial-to-trial learning with feedback. This helped to reduce the overall experimentation time in comparison with traditional multi-session learning paradigms, hence allowing cortical responses to be compared across different stages of learning with minimal intervention and delay.

Statistical methodologies

1. I developed a novel statistical method for characterizing spatiotemporal regions of interest in the cortex. The method imposes spatiotemporal contiguity constraints in discovering regions of interest and allows their statistical significance to be globally quantified. This method provides an alternative to classical multiple comparison corrections that do not take into account contiguous structures across time and space.

2. I developed a new method for reconstructing cortical source activities in MEG that incorporates experimental specific and domain knowledge in the source model—contrary to most state-of-the-art methods that apply generic procedures. I further proposed a

trial-partitioning approach that splits data for model estimation and validation separately, offering a strong evaluation for source models in reconstructing source activities in the cortex.

Chapter 3

Experiment I: Learning visually-similar object categories

Humans are remarkably proficient at categorizing visually-similar objects. To better understand the cortical basis of this categorization process, we used magnetoencephalography (MEG) to record neural activity while participants learned–with feedback–to discriminate two highly-similar, novel visual categories. We hypothesized that although prefrontal regions would mediate early category learning, this role would diminish with increasing category familiarity and that regions within the ventral visual pathway would come to play a more prominent role in encoding category-relevant information as learning progressed. Early in learning we observed some degree of categorical discriminability and predictability in both the prefrontal cortex and the ventral visual pathway. Predictability improved significantly above chance in the ventral visual pathway over the course of learning with the left inferiortemporal and fusiform gyri showing the greatest improvement in predictability between 150-250$msec$ ($M200$) during category learning. In contrast, there was no comparable increase in discriminability in prefrontal cortex with the only significant post-learning effect being a decrease in predictability in inferior frontal gyrus between 250-350$msec$ ($M300$). Thus, the ventral visual pathway appears to encode learned visual categories over the long term. At the same time these results add to our understanding of the cortical origins of previously-reported signature temporal components associated with perceptual learning.

3.1 Theoretical background

Objects from visually-similar categories can be difficult to distinguish, but human observers can make accurate category judgments within a fraction of a second, a visual skill that is perfected by learning and experience [66]. Beyond the case of face individuation where each category is mapped to an identity, the more general ability to assign categories to visually-similar objects has important consequences in our natural environment. For example, distinguishing between ripe or poisonous berries, wet or icy roads, or Retrievers or Rottweilers, all necessitate placing one collection of visually-similar objects into a common category, yet keeping that category distinct from another collection of objects that are not only similar to one another, but to the objects in the first category. This sort of categorization is often referred to as "subordinate" to differentiate from "basic-level" categorization in which there are significant visual differences supporting placing objects into one category or another (e.g., pigs vs. airplanes). Moreover, it is often assumed that subordinate-level categorical decisions will incur a larger cost in response time as compared to basic-level categorical decisions–indeed, this functional definition is often used to ascertain whether a given category is considered basic or subordinate [17]. At the same time, this response time differential can be minimized through experience in that visual "experts" exhibit an entry-level shift whereby subordinate categorization for domains of interest becomes just as fast as basic-level categorization [67, 68]. For example, for bird experts, distinguishing between different species of birds–all nominally members of the same basic-level category—is likely to be just as fast as in telling a bird from a chair. Thus, we can view becoming proficient at categorizing visually-similar objects as an instance of perceptual expertise with subordinate category discriminations. While it is understood that both the ventral occipito-temporal visual cortex, in particular the ventral visual pathway (VVP), and the prefrontal cortex (PFC) are involved in such visual categorization tasks, there is no strong consensus on the relative roles of these neural substrates. Moreover, once specific subordinate-level categorization proficiency has been acquired, there is still a poor understanding of the precise timing of the contributions of the VVP and PFC during the on-line discrimination of visually-similar objects.

To better characterize the roles of the VVP and PFC in the categorization process, we use magnetoencephalography (MEG) to unravel the cortical time course in visual category

learning in order to evaluate two prominent, yet competing, theories. The first approach, which we refer to as "dominant PFC viewpoint", emphasizes the role of prefrontal cortex (PFC) in categorization and proposes the VVP to be sensitive to visual feature differences but agnostic as to category memberships [20, 23]. For example, Jiang and colleages [20] found that categorization training induces category-level changes in lateral PFC but only continuous shape-level changes in lateral occipital cortex (LOC). Related work in non-human primates likewise suggests a similar distinction between PFC and inferiortemporal cortical neurons [21, 22, 45]. These and other data paint a picture of PFC as the neural substrate supporting category learning and the VVP as the neural substrate providing the undifferentiated (with respect to category) perceptual input that the category-knowledgeable PFC utilizes.

An alternative approach, which we will refer to as "complementary PFC viewpoint", suggests that the VVP and PFC play complementary roles in categorization [24, 44, 50–52]. Under this view, the VVP exhibits category boundary sensitivity [69, 70] and the PFC provides early top-down categorical inferences that facilitate initial learning of category-relevant feature dimensions [71]. Learning and reinforcement progressively instantiate these stimulus dimensions within the VVP; that is, the VVP becomes increasingly sensitive to learned category boundaries as the high-dimensional stimulus space is mapped. This is clearly seen in fMRI for highly overlearned, "expert" domains in which the VVP shows spatially localized, differential responses to subordinate-level categories such faces [33], novel objects [72, 73], birds and cars [74]. Similarly, event-related potential (ERP) has consistently revealed category sensitivity in the VVP-sourced $N170$ component [75] and, in several studies of visual expertise, has been localized to posterior occipito-temporal areas [41]. Again, as with the fMRI results, this category sensitivity for domains of expertise has been found for both real world [75] and lab-trained experts [76].

Some of the discrepancies between results supporting these two approaches may be accounted for by differences in stimulus-morphing procedures used in different experiments. In particular, research supporting a dominant PFC view has typically used a more difficult-to-learn type of morph space (i.e., blended morphspace). In contrast, research supporting the complementary PFC view has typically relied on a grid morph space (for a thorough consideration of the topic, see [44]). This raises the possibility that morphing procedures are actually driving the apparent differences in the role of the PFC for these experiments:

the extremely difficult morphspaces require more PFC intervention for participants to map category boundaries, which in turn supports a dominant PFC viewpoint, while the more comprehensible morphspace experiments find that VVP areas are capable of instantiating category boundaries in and of themselves, supporting a complementary PFC viewpoint. As such, perceptually homogenous subordinate categories that have clear decision boundaries, may serve as an ideal test for comparing these views of the PFC's role in categorization. In our experimental paradigm, category membership is never as indeterminate as it would be in the blended morphspace seen in dominant PFC studies, but accurate categorization is still challenging, due to the subtle differences in category features. This design retains the difficulty of blended morphspaces with the predictability of grid morphspaces. Thus, this experiment has the potential to resolve some of the reported differences between the two approaches on the magnitude of the role that VVP plays in the context of subordinate categorization.

To evaluate both approaches, we studied human cortical activity while participants learned to discriminate between two novel and highly-similar visual categories. We hypothesized that although both the VVP and PFC would be involved in the categorization process, their roles would differ during different phases of learning—consistent with the VVP-PFC complementary viewpoint. More specifically, we predicted that the VVP would acquire categorical representation as learning progressed to the point where the category boundaries are better distinguished by participants. In contrast, we predicted that PFC would play a more significant role in category encoding in the initial phases of learning, during the early formation of categorical representations, but that this role would diminish later in learning. With respect to these predictions, the differential roles of the VVP and PFC have been explored by Bar et al. [52], who found, in a visual recognition task, that PFC responses both temporally preceded those in the VVP and were more sensitive to low spatial frequencies. They hypothesized that PFC may be involved in providing early inferences regarding object identities that are subsequently refined by further visual processing within the VVP. Our predictions are related to this hypothesis, but are critically different in two important aspects. First, we focused on categorization instead of individual object recognition—it remains unclear whether the VVP and PFC both play a role in discriminating between visually-similar categories. Second, we investigated the change in response for the VVP and PFC over the course of category learning as opposed to investigating only the end point of learning. In

particular, this latter manipulation allowed us to monitor how neural coding of categories change over time—hopefully offering a means for better elucidating the functional roles of both PFC and the VVP.

To pursue these goals, we created two novel, visually-similar shape categories inspired by the stimuli used by Krigolson and colleagues [49]—Figure 3.1A illustrates the stimulus image space, showing five samples from each of the two categories. Each of these blob-like exemplars is unique and represents a jittered version derived from one of the two prototypes located at the center of the space of samples forming each category. Although these exemplars are perceptually similar with small differences in the edge contours, a distinct category boundary is embedded in the overall space, as illustrated by two distinct clusters shown in Figure 3.1B. Participants were trained to discriminate between these two "blob" categories in a feedback-driven categorization task in which we monitored neural activity using MEG. At a fine temporal scale, MEG's millisecond temporal resolution afforded us the ability to investigate how different cortical regions embed discriminative information about the blob categories over time. At a coarser temporal scale, we were able to explore how the encoding of this category information evolves during the course of learning, particularly with respect to categorical representations in the ventral and occipito-temporal visual and prefrontal cortices.

3.2 Materials and methods

Ethics statement

All experimental procedures were approved by the Institutional Review Boards at Carnegie Mellon University and the University of Pittsburgh. All participants gave written informed consent and were compensated financially for their participation.

Participants

Ten right-handed participants (4 females and 6 males) aged between 17 and 35, recruited from the Pittsburgh area, were run in the experiment. Participants were financially compensated

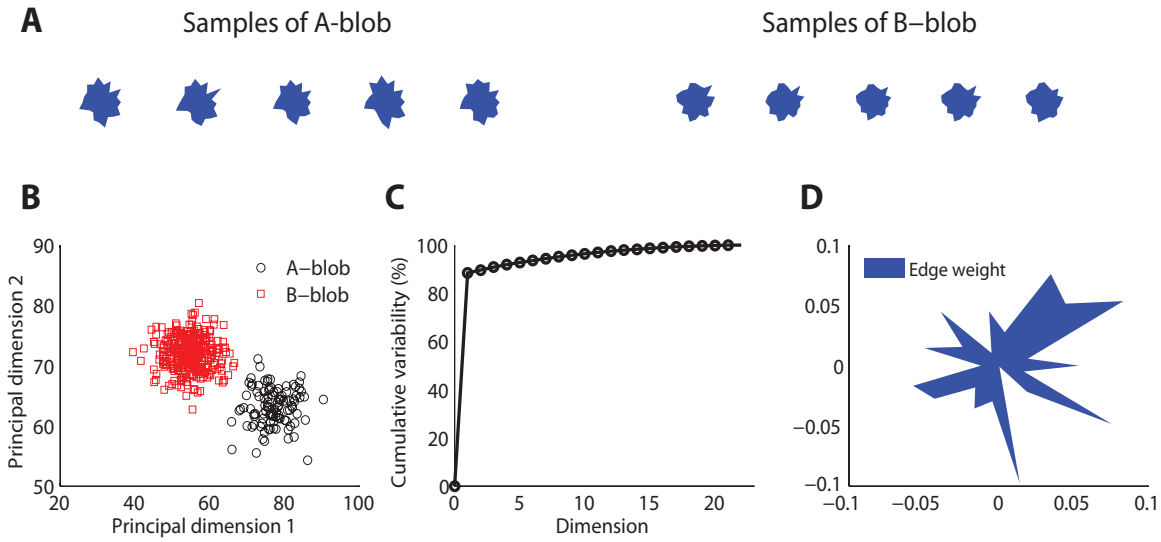

Figure 3.1: **Visual stimulus design.** **(A)** Blob samples from A and B categories. **(B)** Projection of A and B blobs in two principal dimensions via principal components analysis. **(C)** Cumulative variability accounted for in the principal dimensions. **(D)** Normalized edge weights from the first principal dimension.

for their participation. Two of the participants ran in experimental sessions in which trigger failures meant that the timing of individual trials could not be retrieved, so the data for these two participants was discarded. One participant was unable to correctly learn the blob category boundaries, exhibiting near-chance categorization accuracy throughout the experimental session, so the data for this participant was likewise discarded. Thus, the results reported here are based on the remaining seven participants.

3.2.1 Stimulus design

The visual stimuli were generated from two novel artificial categories, A and B. Each category was defined around a prototype "blob" that corresponded to the center of a space of blob exemplars (Figure 3.1A). Within each category, 300 unique blob exemplars were generated from a parameterized distribution. Each blob was the result of a random two-dimensional polygon with 20 edges (or dimensions) similar to the design used by Krigolson et al. [49]. The edges were defined as distances of proportion $30-70\%$ of the distance between an origin

and 20 vertices uniformly distributed around a unit-length circle. To control for statistical variability, blobs were generated from a multivariate Gaussian distribution specified for each category, where the mean of the distribution is the 20-dimensional vector of the prototype, and the covariance is a diagonal matrix with variance in each dimension proportional (20%) to the difference in edge distances across the two exemplars. This ensures samples within each category vary slightly from each other but remain distinct from the other stimulus category. Figure 3.1A shows several samples drawn from each of the two categories. This design yields a distinct category boundary—illustrated by the two separate blob distributions as projected into the space defined by the first two principal components of a PCA (Figure 3.1B). A comparison of the number of dimensions to cumulative variability establishes that the greatest variation ($\sim 90\%$) among the blob samples is captured in one to two dimensions (Figure 3.1C). Finally, Figure 3.1D illustrates the normalized weight that each edge shares in the first principal dimension—a lengthier edge accounts for more variability in this dimension and hence is more likely to be used as discriminative features for successful categorization.

3.2.2 Experimental procedures

The experiment involved a trial-by-trial feedback-driven visual category learning task where the participants' task was to discriminate between the two blob categories. Each experimental session consisted of 600 trials that included randomized presentations of 300 unique A-blobs and 300 unique B-blobs. The session was divided into five equal blocks of trials with brief self-paced breaks between each block to reduce fatigue. The sequence of A and B blobs was permuted for each participant and the number of presentations of stimuli from each category was balanced during each block.

Each trial began with a machine-synthesized random auditory label of "A" or "B" ($630msec$) transmitted via non-magnetic ear-plugs while the participant visually fixated on a centered cross. A projector was used to back-project stimuli on a non-magnetic screen ($58cm \times 81cm$) to display all visual stimuli. After an extended $120msec$ fixation, either an A-blob or B-blob exemplar was displayed at the center of the screen (subtending a visual angle of approximately 3.4 degrees both vertically and horizontally) for a brief interval of $750msec$. During the period while the blob was displayed, the participant responded with a

finger press to indicate whether the audio category label matched the blob category ("yes" or "no"). For example, if the participant heard the label "A" followed by a visually-presented "B" blob, the participant would press a button to indicate "yes", match, or "no", a mismatch. The "yes" and "no" labels were displayed along the left or right bottom corners of the screen with their positions counterbalanced for each experimental session. A glove response pad was used to allow participants to respond with finger presses with minimal wrist movement. Shortly after response, the participant would receive on-screen feedback after a jittered interval of $150 - 300msec$: "correct", "wrong", or "too slow" were displayed in the center of the screen for $750msec$ to indicate the correctness of their response. Participants had to respond within the $750msec$ window to avoid the "too slow" feedback. The inter-trial-interval was $500msec$ before the next trial began.

Our experimental procedure is similar to the study by Krigolson and colleages [49] with two important distinctions. First, we used an audio label as a prompt for each category to be matched to the subsequent visual presentation of a blob exemplar, whereas in Krigolson et al. (2009) each blob stimulus was simultaneously shown below a randomized written label showing either the letter "A" or "B". Their trial design made it difficult to determine whether the observed categorical visual responses were driven by the visual letter or the blob stimulus. Second, Krigolson et al. (2009) were equally interested in categorization and error-driven learning, so they continually shortened stimulus presentation to ensure an adequate number of errors for analysis. In contrast, our primary interest was in understanding visual category learning, therefore we maintained a stable visual presentation time throughout our experiment.

3.2.3 MEG data acquisition and preprocessing

Using magnetoencephalography (MEG), we recorded cortical activity while participants were trained to discriminate between the two blob categories. All experiments were conducted in an electromagnetically shielded room with participants seated comfortably and head-fixed throughout the session. Neural data were recorded using a 306-channel whole-head MEG system (Elekta Neuromag, Helsinki, Finland). The system has 102 channels where each is a triplet of a magnetometer and two perpendicular gradiometers.

MEG signals were sampled at $1000Hz$. Four head position indicator coils were placed on

the scalp to record relative head positions to the MEG machine at each session. Electrooculography and electrocardiography were recorded by additional electrodes placed above, below and lateral to the eyes and at the left chest respectively. The coil and electrode signals were used to correct for movement and artifacts throughout the experiments, the MEG signals were bandpass-filtered between 0.1 and $50Hz$ to prevent power-line interference at $60Hz$, and signal projection methods were used to remove artifacts such as heart beats. Any delay in the visual display of stimuli on the screen was measured by photodiodes and was corrected for in all reported results. For all of our analyses, we focused on the $400msec$ period after visual stimulus onset and prior to the participant's categorization responses. The baseline defined as $120msec$ prior to the onset of the blob stimulus was removed for each trial to account for signal drift.

Cortical source estimates were computed using the Minimum Norm Estimates (MNE) [2] in MNE Suite software (http://www.nmr.mgh.harvard.edu/martinos/userInfo/data/sofMNE.php). Source dipoles were evenly distributed ($5mm$ separation between neighboring sources) with orientations fixed normally to the cortical surface. Surface brain models for each individual participant were constructed by Freesurfer software (http://surfer.nmr.mgh.harvard.edu/) from structural magnetic resonance imaging scans acquired at the Scientific Imaging and Brain Research Center at Carnegie Mellon University (Siemens Verio 3T, T1-weighted MPRAGE sequence, 1×1×1mm, 176 sagittal slices, TR = $1870msec$, TI = $1100msec$, FA = 8 degrees, GRAPPA = 2). Based on the neural anatomy of each individual participant, 24 ventral visual and prefrontal cortical regions containing multiple source dipoles were identified from Freesurfer segmentation using the Desikan-Killiany Atlas [77].

3.2.4 MEG sensor-space analysis

A multivariate Hotelling's t-test was applied across the MEG time series data to evaluate whether MEG sensor signals carry information capable of discriminating between categories A and B. At each time point, a multi-dimensional vector was defined as the ensemble signal from 102 scalp magnetometers averaged within a $10msec$ window (the time-averaging was performed by taking the mean within a moving window of $20msec$ in step of $10msec$ along the time course). This vector was then collected for each single trial where a blob exemplar was presented. All trials were divided into two groups based on the category membership of the

presented blob stimulus in each trial for the t-test. To assess whether the multivariate sensor signal is identical under A and B groups (null hypothesis), the high-dimensional vectors were first mapped into a lower-dimensional space via principal components that preserved at least 99% signal variability prior to the test. This ensures a non-singular inversion in estimating the covariance matrices in the t-tests. The resulting projected vectors from all trials were subsequently evaluated with the Hotelling's t-test. The computed value was expressed in terms of a χ^2 statistic at each time point, and it was repeatedly applied through the entire time course between $0 - 400ms$ after the visual onset.

3.2.5 MEG source-space analysis

Similar procedures were applied to the MEG source space. Anatomically bounded regions in the ventral visual pathway and prefrontal cortex were first defined by the segmentation result from Freesurfer. Because each region contained multiple dipoles, a multivariate Hotelling's t-test was performed over time to evaluate whether dipoles within each cortical region discriminated trials containing A or B blobs. At each time point, a multidimensional vector was constructed by the ensemble of cortical dipole amplitudes averaged in $10msec$ windows. This vector was then reduced via principal components analysis to lower dimensions that capture 99% variability (again to ensure non-singular inversion in the covariance estimation). The resulting projected vectors from all trials were evaluated with the Hotelling's t-test at each available time point. The analysis was repeated among first 100 trials and final 100 trials separately to compare the neural of visual categories at different stages in the learning process.

An excursion test [3] was used to evaluate the significance of the discriminative time course in source space. This followed a number of steps. First, discriminative time course was thresholded and only contiguous time points that exceeded the threshold were proposed as potential regions of interest. Contiguity was satisfied if any of the immediate neighbors of a given point in time also passed the threshold—this procedure helped to prune isolated events that are likely to occur due to chance. This same procedure was then applied to the same data multiple times (100-fold permutations), but in each case, category labels were shuffled—this provided a baseline measure, or a null distribution. A p-value was then computed using a standard permutation test by comparing the discriminability statistics

within the proposed regions of interest to those in the permuted data following procedures described in [3].

Logistic regression was used to predict blob categories from cortical time course activities at predefined time windows. Within each of 24 anatomically defined cortical regions, time courses of all available cortical dipoles were averaged across time windows $50-150$, $150-250msec$, $250-350msec$ and $0-50msec$ (baseline) post-stimulus respectively. The predictive decoding analysis was then performed within each of these windows. First, ensembles of cortical dipole amplitudes were collected for 100 trials in earliest and final phases of the learning session separately. For each phase, a leave-one-trial-out cross-validation was used to predict the category membership of blob presented at a single held-out trial. Specifically, the multidimensional ensemble of dipole amplitudes for each anatomical region were projected to a low-dimensional space via principal components that captured 99% variability. Then, at each round of cross validation, a logistic regression classifier was used to predict the blob category in an unseen held-out trial given logistic weights estimated from all remaining trials. This procedure was repeated for all trials until every trial was predicted, and the overall accuracy was reported based on the percentage of trials where the classifier correctly predicted the blob category.

3.3 Results

3.3.1 Behavioral category learning performance

Seven participants successfully learned the blob categorization task. Figure 3.2A shows the individual categorization accuracies in the first and final 100 trials (error bars indicate standard errors of the means), representing behavioral performance during early and late stages of learning. All but one participant improved significantly ($p < 0.05$ from binomial tests with Bonferroni corrections) over the course of learning. The remaining participant also improved, although the improvement was only marginally reliable ($p = 0.07$). However, all participants were able to categorize the blobs significantly above chance rate 50% ($p < 0.01$ from t-tests) with an average terminal accuracy of 83% for the late stage of learning. Figure 3.2B shows the mean reaction times for the early and late stages of learning (error bars indicate standard errors of the means). Only three subjects showed significant reduction

in the reaction time ($p < 0.005$ from t-tests with Bonferroni corrections)—this was expected because the $750msec$-deadline period was sufficiently short for a combined perceptual and motor response for some participants.

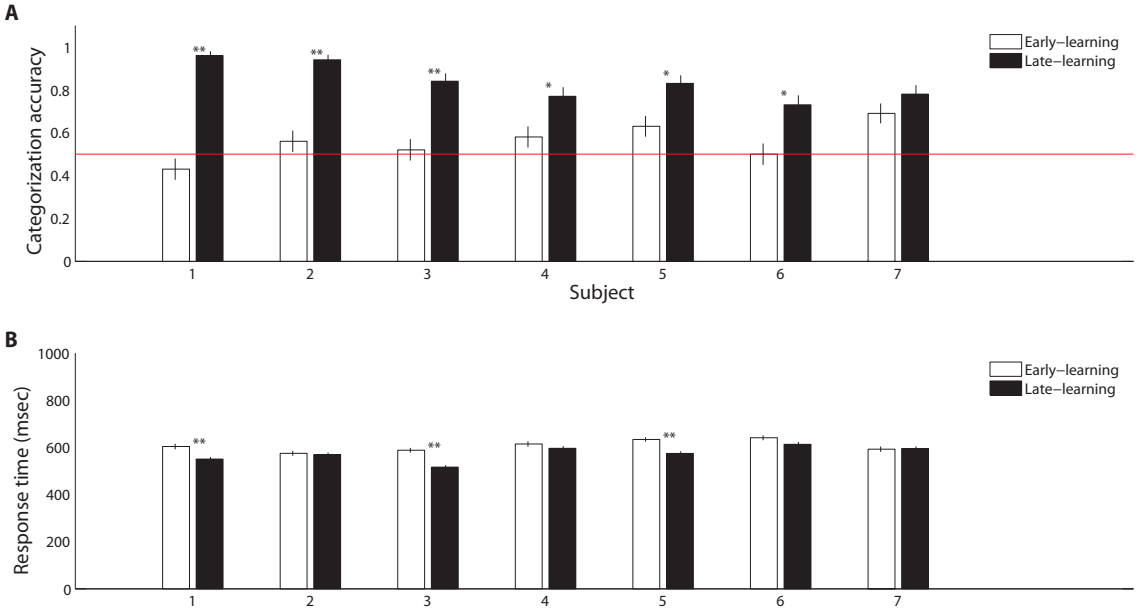

Figure 3.2: **Summary of behavioral category learning performance. (A)** Categorization accuracies during the early (first 100 trials) and late (final 100 trials) periods of the learning experiment. **(B)** Reaction times during trials from the same periods. "*" and "**" indicate significance at $p < 0.05$ and $p < 0.005$ respectively with Bonferroni corrections.

3.3.2 Category coding in MEG sensor space

Given that our participants successfully learned the two visual categories, our next step was to assess whether category memberships can be reliably discriminated from MEG sensor data. We expected the recorded sensor data to differentiate trials in which participants recognized blobs from category A as compared to category B. To evaluate this proposal, we performed Hotelling's t-tests with dimension-reduced magnetometer signals and computed category discriminability (χ^2 statistic) over time using all available trials partitioned into A and B categories. To obtain a chance-level distribution for comparison, we also applied this

procedure to trials with shuffled category labels (100-fold permutations) for each individual subject.

Figure 3.3 shows the group-level statistics. We were able to reliably discriminate the A and B blob categories within the half-second period after visual onset in a single trial. In particular, the mean category discriminability rises post-$50msec$ and is highly separable from the chance-level after $100msec$. To assess the significance of these results, we applied an excursion procedure similar to [3] that compares the temporal statistics from the original data (without permutation) with the permuted statistics. We found that category discriminability is statistically significant post-$100msec$ for all subjects (combined $p < 1.8\times10^{-8}$ from Fisher's method; $p < 0.01$ from individual-based excursion tests). Figure 3.9 in the Appendix of this chapter shows the time course for each individual subject.

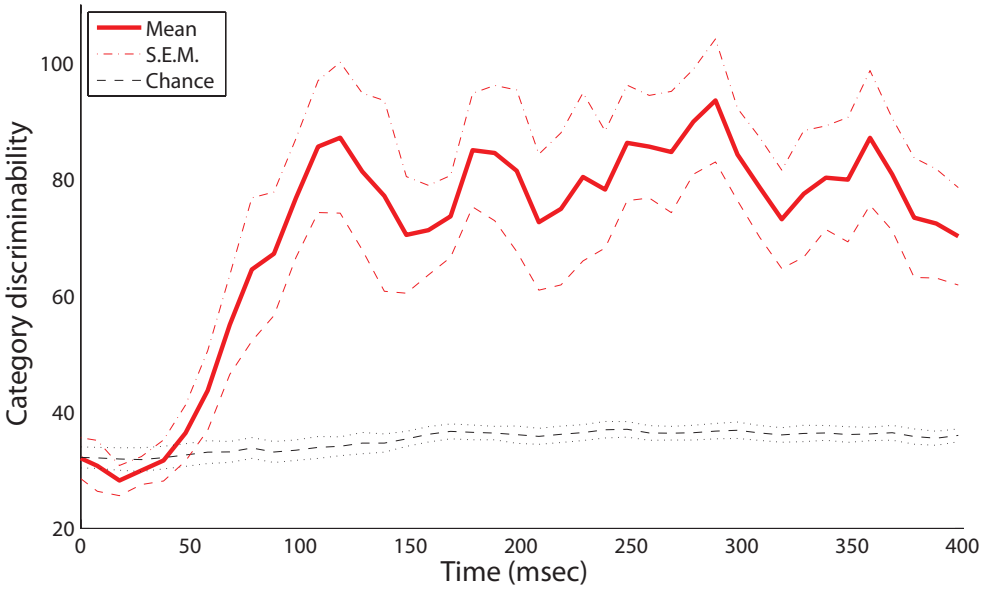

Figure 3.3: **Category-discriminative time course in MEG magnetometers.** Group-level category-discriminabilitive time course (visual stimulus onset at $0msec$) compared against pooled chance-level time course computed from trials with shuffled category labels.

3.3.3 Category coding in the ventral visual pathway and prefrontal cortex

Our previous analysis demonstrates that the time course in MEG sensors contains significant category information in aggregate, but it does not address the question of localizing which brain regions are the sources of this information or how these sources may change with learning. To evaluate our hypotheses regarding the relative roles of the ventral visual pathway and the prefrontal cortex, we used similar methods to compute category-discriminative time series in MEG *source* space. In particular, we focused on anatomically-defined regions in ventral occipito-temporal visual and prefrontal cortices.

To test whether the ventral visual pathway is capable of learning and discriminating exemplars from visually-similar categories, we compared time courses in related cortical regions during both the early and late stages of learning. Similar to our sensor-space analysis, a category-discrimination time course in source space was computed by performing multivariate Hotelling's t-tests from cortical dipole activities across time. To distinguish trials in the early and late stages of learning, tests were performed for the 100 earliest and the 100 latest trials separately with equal numbers of A and B blobs presented.

Figure 3.4 summarizes the results for 12 visual cortical regions and 12 prefrontal regions in both left and right hemispheres. During early learning as illustrated in Figure 3.4 A-B, we observed that category discriminability rises at approximately $100msec$ post-stimulus in both hemispheres. During late learning as illustrated in Figure 3.4 C-D, we observed that category discriminability also rises at approximately $100msec$, but discriminability peaks post-$200msec$ in the lingual, lateral-occipital and fusiform gyri in the left hemisphere. This time window agrees roughly with $N250$ as previously reported in [49], except here we provided better localization of its sources in the cortex. In comparison, we observed relatively scarce discriminability in prefrontal cortex throughout time course and learning as illustrated in Figure 3.4 E-H.

To assess the significance of the category-discriminative time course, we performed an excursion test following [3]. Specifically, for each subject, we obtained regions of interest by thresholding the time course at 20 and kept contiguous time points that passed the threshold. We evaluated the significance for each subject by comparing the discriminability statistics within the proposed regions of interest against the statistics within regions found

from the permutated data (100 folds)—this yielded a global p-value. Figure 3.5 shows the temporal regions of interest pooled across subjects (combined $p < 1.8 \times 10^{-8}$ from Fisher's method; $p < 0.01$ from individual-based excursion tests). These results show that category information flows primarily in the bi-lateral occipital, lingual, pericalcarine, fusiform and inferior-temporal gyri during both early and late learning, suggesting that the VVP acquires discriminability of novel, visually similar categories during learning.

Figure 3.5 also shows that regions of interest in prefrontal cortex are more sparse in comparison with those in the VVP. In particular, whereas temporal coding appears in prefrontal cortex during early learning, it decreases in late learning, suggestive of a diminished role of prefrontal cortex. Figure 3.10 in the Appendix of this chapter shows that such a pattern is consistent across all subjects. Our current set of results, however, does not rule out the possibility that coding in PFC becomes more sparse over time (e.g. [22]) or that it could be generated from a deep source which is difficult to detect with MEG.

3.3.4 Predicting categories from cortical activity

To this point, our analyses have explored category discriminability across a continuous time course. These analyses also help identify time windows that appear to offer availability of category-discriminative cortical information. Thus, one question we can ask is how temporal windows differ from one another with respect to what information they carry regarding visual category learning. A similar question may be asked with respect to spatially localized activity—does the ventral visual pathway carry more information regarding subordinate-level visual categories relative to prefrontal cortices?

To address these questions, this next analysis evaluates to what extent the ventral visual pathway and prefrontal cortex are *predictive* of blob categories at the discrete temporal windows of $M100(50-150msec)$, $M200(150-250msec)$ and $M300(250-350msec)$, as well as, critically, how category predictability within these temporal windows changes over the course of learning. We predict that the ventral visual pathway will play a significant role in category learning and representation. In particular, VVP is expected to acquire an increasing degree of category predictability—more than PFC—during learning.

To test this prediction, we performed a decoding analysis to assess category predictability in the same 24 anatomically-defined regions in the ventral visual pathway and prefrontal

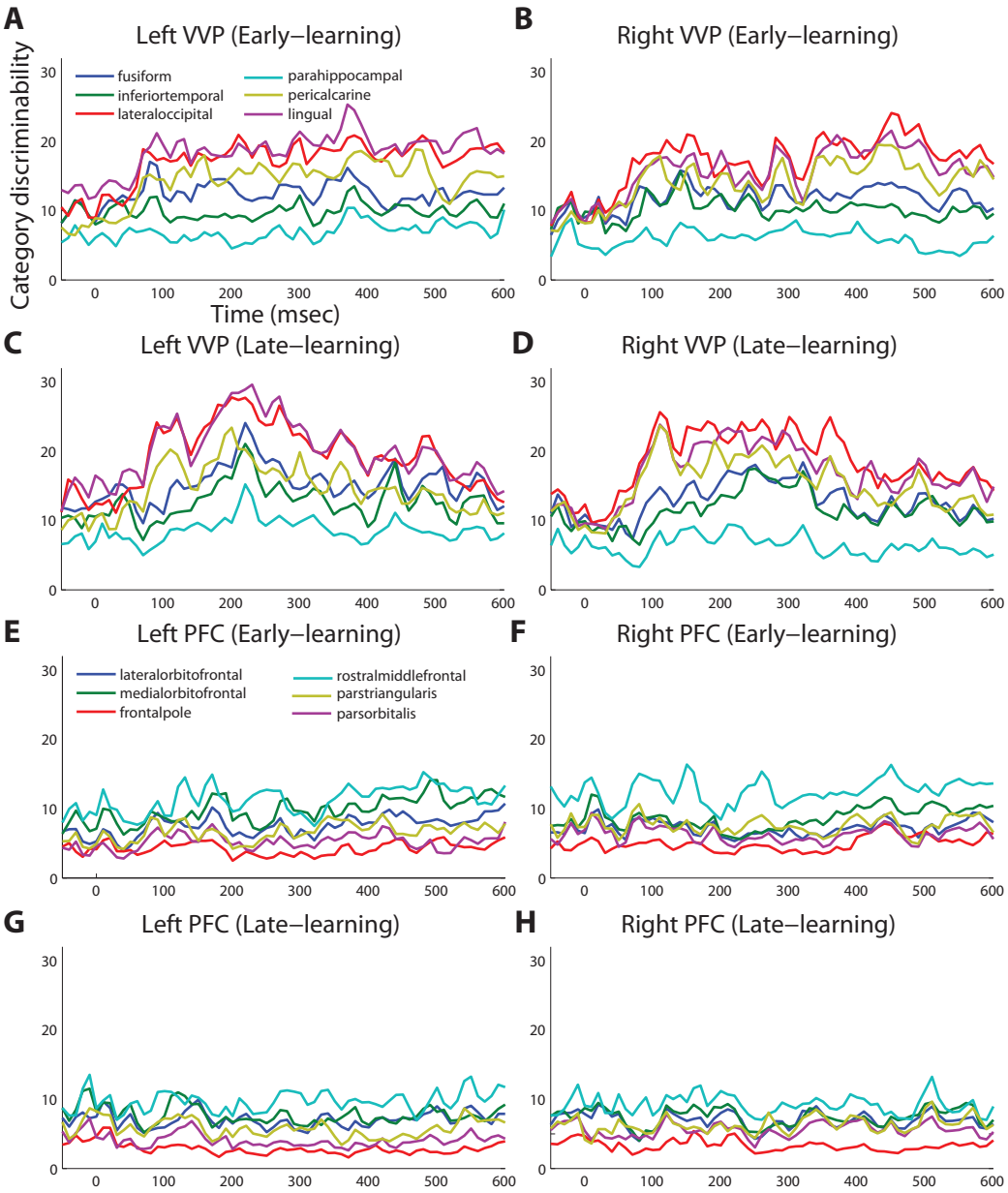

Figure 3.4: **Category-discriminative time courses in ventral visual and prefrontal cortices. (A)** Group-level discriminative time courses in right-hemispheric VVP contrasting dipole responses in trials containing A and B blob categories during early learning. **(B)** Discriminative time courses in left-hemispheric VVP regions during early learning. **(C-D)** P-value time courses in VVP regions from left and right hemispheres during late learning. **(E-H)** Discriminative time courses in PFC regions under similar conditions.

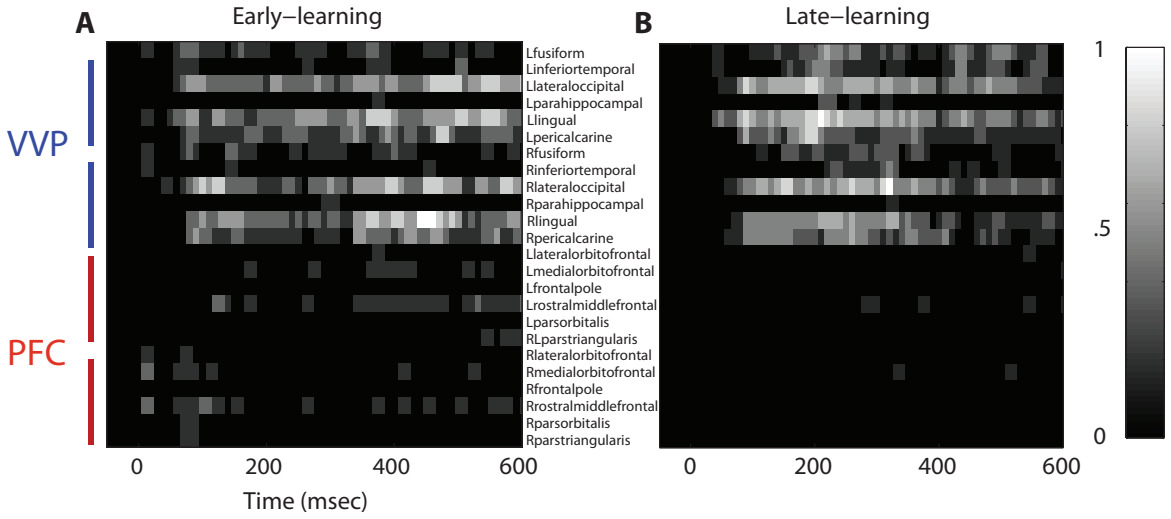

Figure 3.5: **Regions of interest in ventral visual and prefrontal cortices after excursion tests. (A)** Group-aggregated regions of interest during early learning. **(B)** Group-aggregated regions of interest during late learning. The color bar indicates the tally (normalized across subjects) where a specific cortical region at a time point passes the excursion test.

cortex used in our earlier analyses. Within each of these regions, we ran held-out predictions regarding blob categories on a trial-by-trial basis using multidimensional cortical dipole activities averaged within the following time windows: $M100$ $(50-150msec)$, $M200(150-250msec)$, and $M300(250-350msec)$, as well as the baseline of $0-50msec$, post-stimulus. This was implemented using a standard leave-one-out cross validation technique which evaluated to what degree category membership of a blob presented in a single trial not part of the training set can be predicted based on region-bounded dipole responses and blob category labels from the remaining trials in the training set. To compare predictability during initial and end-stage learning, as in the previous analysis, this decoding analysis was conducted separately for the first and final 100 trials.

Figure 3.6 summarizes blob category-predictive accuracies across all 24 cortical regions and time windows in the early and late stages of learning. At $M100$, pericalcarine gyri, right lingual and left LOC gyri become highly predictive with respect to blob categories ($p < 0.005$ under t-tests), but no significant difference was observed in predictability between early and late learning ($p > 0.05$ under t-tests)—suggesting that category predictability in this early

time window may not be shaped by category learning *perse*. Within $M200$ and $M300$ windows, across most of visual cortex, predictive accuracies in the late learning stage are considerably better than they were in the initial learning stage.

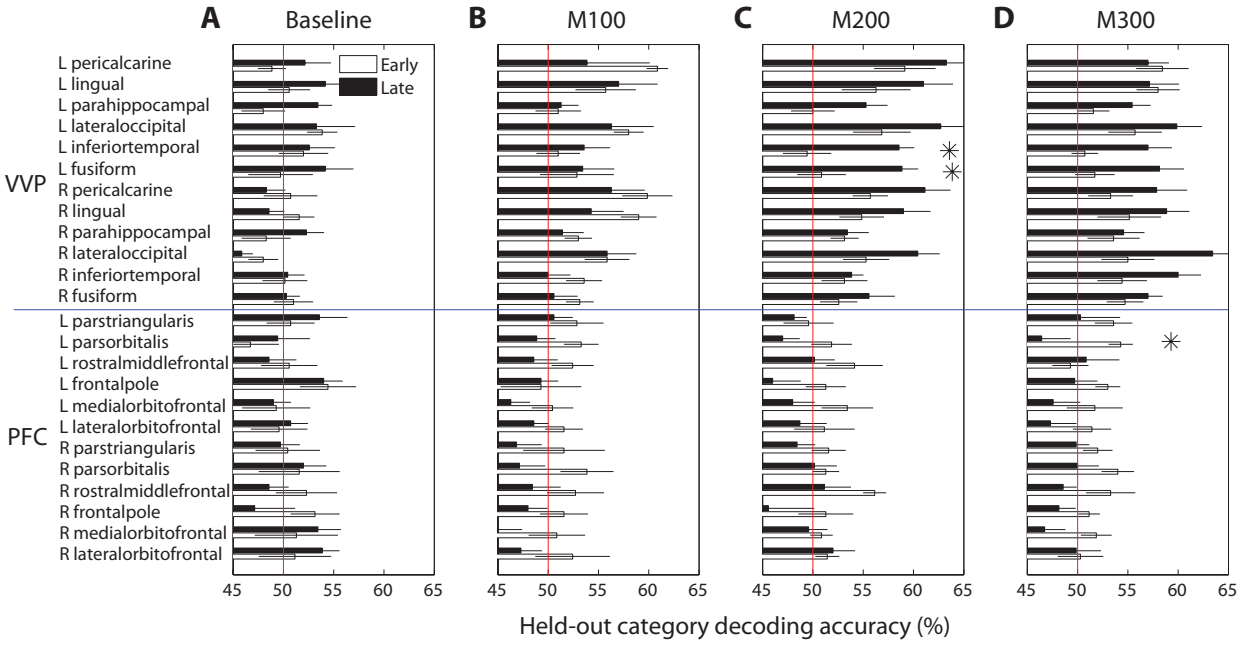

Figure 3.6: **Category predictive accuracies in ventral visual and prefrontal cortices.** **(A)** Group-average blob category predictive accuracies in 24 ventral visual and prefrontal cortical regions based on dipole activities in $0-50msec$ after onset during early and late learning. **(B)** Decoding accuracies in $M100$ $(50-150msec)$ window. **(C)** Decoding accuracies in $M200$ $(150-250msec)$ window. **(D)** Decoding accuracies in $M300$ $(250-350msec)$ window. Asterisks indicate significant difference ($p < 0.05$) in predictive accuracy between early and late learning.

In particular, left inferiortemporal (ITG) ($p < 0.024$) and fusiform (FG) ($p < 0.025$) gyri show significant increases in category-predictive accuracy. This pattern suggests that learning plays a greater role in shaping cortical responses at these later temporal stages of processing—confirming our hypothesis that visual cortex encodes and represents subordinate visual categories. To visualize these cortical learning effects, we extracted dipoles that showed reliable differential response ($p < 0.001$) across the A and B blob categories within the $M200$ window. Figure 3.7 illustrates the significant discriminability in source dipoles that appeared

in the left ITG, the left FG, and bilaterally in the LOC later in learning—effects that were absent during the initial learning phase of the experiment.

Unlike visual cortex, regions in prefrontal cortex are generally less predictive about blob categories (bottom panels of Figure 3.6). In addition, these regions are marginally more predictive earlier in learning relative to later in learning, with left pars orbitalis (or inferior frontal gyrus) showing a marginally significant ($p < 0.05$) decrease in predictive accuracy at $M300$. These observations are suggestive that prefrontal cortex plays a greater role in category encoding during learning, but they do not exclude the possibility that learning induces sparse coding in PFC or a more complementary role of PFC that jointly participates category coding with the VVP.

Figure 3.8 compares the ventral visual pathway and prefrontal cortex at $M100$, $M200$ and $M300$ by pooling predictive accuracies across dipoles within each of these cortical regions. The result suggests that both VVP and PFC are near chance in predicting the blob categories during initial learning. However, later in learning, the ventral visual pathway becomes significantly more category-predictive than prefrontal cortex at $M200$ and $M300$ ($p < 0.005$ under t-tests) but not at $M100$ ($p > 0.5$). Interestingly, we found significant interaction between the VVP and PFC during the three time windows during late learning ($p < 0.005$ under 2×3 ANOVA) but not initially during learning ($p > 0.1$). Together, these results support the hypothesis that VVP and PFC function as complements to one another, suggesting that improved categorization performance over the course of learning is associated with increased predictability post-$100msec$ for VVP, but decreased predictability across the time course for PFC.

3.4 Discussion

Models addressing the neural basis of visual category learning have focused on the interplay between the ventral visual pathway (VVP) and prefrontal cortex (PFC). However, there has been no clear consensus on the respective roles of these two neural substrates, with some theories taking a dominant PFC view in which category membership is encoded within PFC, while the VVP is sensitive only to visual feature differences (albeit correlated with category membership) [21, 23]. In contrast, the complementary PFC view holds that the

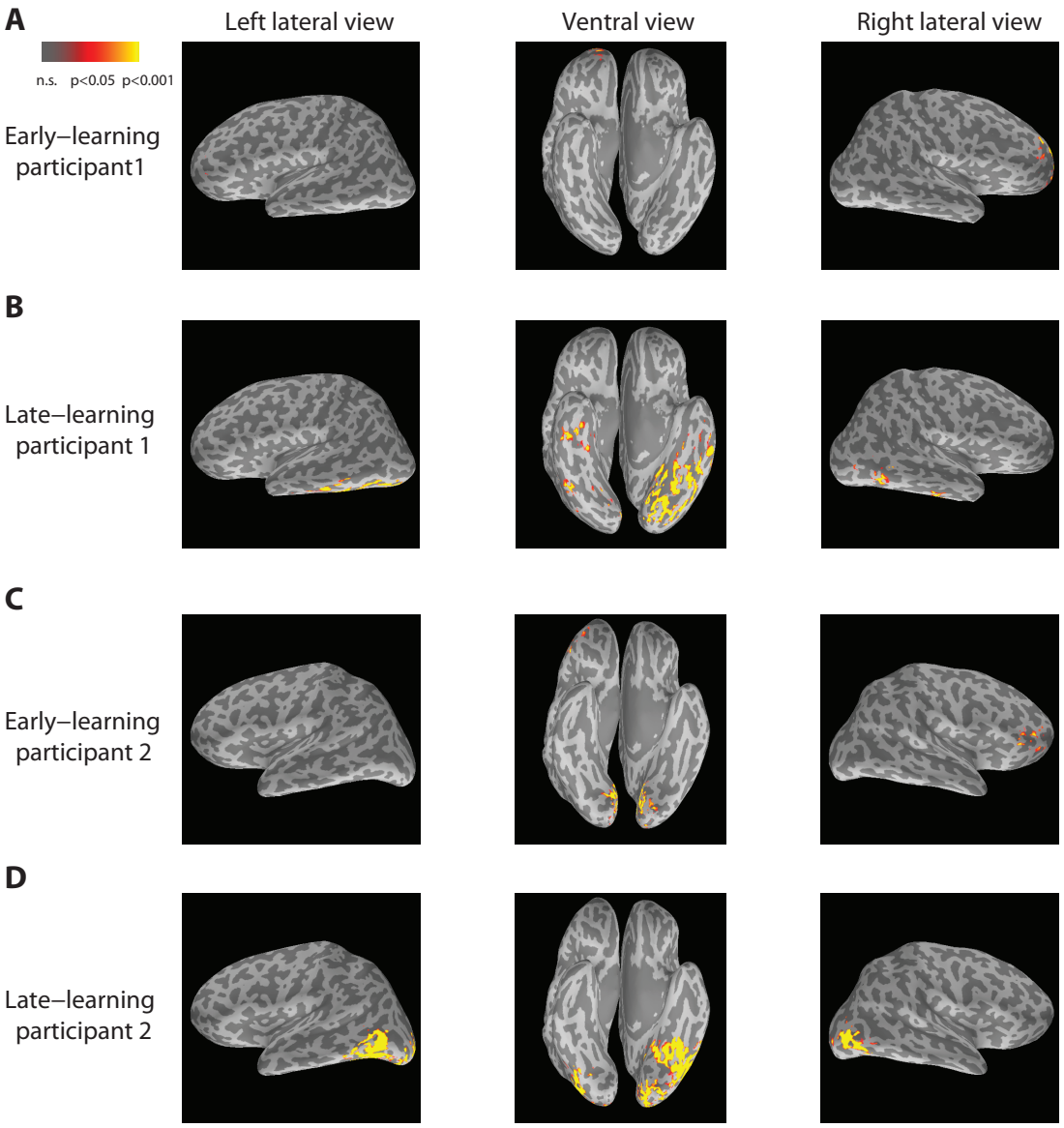

Figure 3.7: **Visualization of category-discriminative cortical dipoles at M200. (A)** Category-discriminative clusters of cortical dipoles from a representative participant during $150 - 250msec$ earlier on in learning. **(B)** Category-discriminative dipoles under similar conditions during late learning. **(C-D)** Category-discriminative dipoles extracted under similar conditions from a second participant.

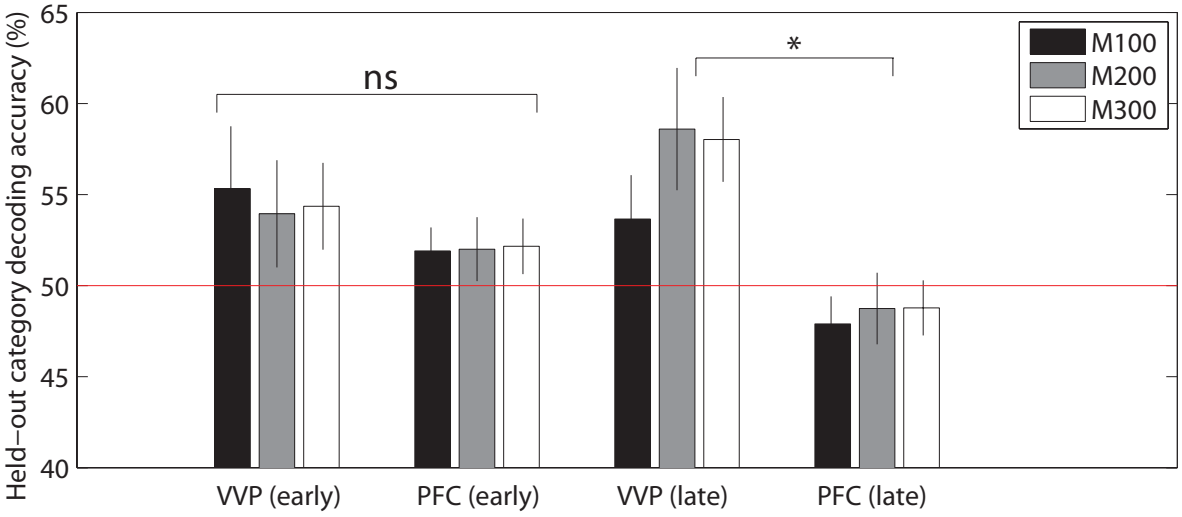

Figure 3.8: **Cortical category-predictive accuracies.** Pooled held-out category predictive accuracies from ventral visual and prefrontal cortices based on the first and final 100 trials during $M100$ $(50 - 150msec)$, $M200$ $(150 - 250msec)$ and $M300$ $(250 - 350msec)$ after visual stimulus onset. Asterisk indicates significant difference ($p < 0.005$) in predictive accuracy between VVC and PFC at $M200$ and $M300$.

VVP and the PFC play different functional roles at different points in category acquisition—the PFC facilitating the learning of category-relevant features during the initial stages of learning, but with the VVP ultimately encoding these featural dimensions so as to become progressively more sensitive to category boundaries (as opposed to purely visual feature differences) [69, 71].

Using magnetoencephalography (MEG)—which provides superb temporal resolution and good spatial resolution—we conducted a decoding analysis to show that the ventral visual pathway contained the neural information to accurately categorize stimuli with in the first $400msec$ after stimulus presentation during a subordinate categorization judgment. Our analysis is unique in showing not only that ventral visual regions are category sensitive, but also that category specific information can be decoded from ventral visual cortical regions.

We obtained these findings by using multivariate discriminative and predictive analyses to assess the role of the VVP and PFC during visual category learning. Overall, our data suggested that category-discriminative information is available from the VVP responses in the $M200$ and $M300$ time windows and that responses originating from the left fusiform and inferiortemporal gyri acquire a higher degree of discriminability and predictability concomitant with increasing categorization performance. In comparison, we found little evidence that PFC carries significant information about visual categories, but the small sample size encourages a cautious interpretation of this fact.

The functional roles of the ventral visual pathway and prefrontal cortex

As already discussed, our study is in large part based on previous research on visual categorization and category learning using both single and multi-array neural recordings in primates [21, 22], and fMRI [19, 73], ERP [49, 76, 78, 79], and MEG [40, 80] in humans. However, to this point, subordinate-level category discrimination at fine-scale temporal resolution with good spatial resolution has primarily been studied at the physiological-level in primates [21, 22]. Critically, for the majority of these primate-based studies, the stimuli were created in a morphspace where the category boundary could not be clearly specified, an issue that places some constraints on what can be concluded from their results [44]. It is unsurprising that the complicated morphspace studies find more PFC activity than the

simpler grid-based design spaces, given the relative difficulty of these two categorization tasks. Meanwhile, Folstein et al. demonstrate that the VVP can instantiate newly-learned category boundary sensitivity when people can focus on diagnostic stimulus dimensions and, essentially, ignore non-diagnostic ones - and that these boundary sensitivites are retained even when task is no longer relevant.

To explore category discrimination in humans, we used a visual stimulus space in which we clustered exemplars to form a distinct category boundary. Moreover, these stimuli were novel to our participants, as such we were able to monitor how the categories became differentiated in the cortex from early to late stages of learning. Our analyses indicated that the measured neural data obtained through MEG tracked the qualitative changes seen in behavioral categorization performance. Our results are consistent with studies that find the VVP to acquire information about stimulus categories, e.g. [19]. More specifically, we found that the lateral occipital complex and the inferiotemporal cortex, possible homologs to the ITC in primates, became significantly more informative with respect to category membership over the course of learning. Contrary to previous findings that support the PFC-dominant theory [20, 21], we found that categorical representation is encoded in the human ventral visual pathway even when categories are comprised of perceptually similar items, supporting the idea that visual cortex plays an predominant role in category learning.

Of note, our study is somewhat different methodologically from many other prior category training studies [73, 76] in that training in our experiment occurred over a single session in which participants are received a training signal in the form of correctness feedback. In contrast, other studies have typically involved a pre-test, a set of training sessions to learn the categories, and a post-test, often including neuroimaging pre- and post- to assess training effects [72, 73]. For example, in Op de Beeck et. al. (2006), participants completed 10 training sessions in order to learn novel object categories, then performed a color change detection task while fMRI data was collected. Consistent with our present results, they observed a wide range of category-selective responses across the VVP. Interestingly, in this study they observed a change in the spatial distribution of the category-selective responses across training, suggesting that the neural representation of categories changes dynamically with experience. In that our study relied on a single training session, our data cannot address the question as to whether the pattern we observe in VVP would remain stable over further training. Finally, we note that although our single session protocol cannot eliminate

the possibility that some of our observed effects are due to attention—in that participants necessarily use attentional resources during learning—our results largely converge with these and other studies showing widespread VVP activation with category learning.

Overall, our work suggests that the VVP plays a central role in discriminating visually-similar object categories. However, our results do not rule out the possibility that prefrontal cortex also plays a role in shaping categories—exerting, possibly based on the nature of the categorization task, some top-down influence on visual cortex during learning [52]. At the same time, our results do not provide evidence for explicit coding of subordinate categories in prefrontal cortex. Beyond our arguments, it is also possible that the coding of categories in PFC is relatively sparse and therefore cannot be detected using the coarse spatial resolution of MEG. Thus, future work is needed to investigate whether sparse codes exist in prefrontal cortex and to address how prefrontal cortex coordinates with visual cortex in representing visual categories during different phases of learning.

The time course of cortical processing during visual category learning

The ERP and MEG literatures contain many proposals about signature waveforms that relate to visual categorization and recognition, the most common ones being time windows at $M100$ [40], $N170$ [78] or $M170$ [40], and $N250$ [49]—negative deflecting MEG or ERP components that peak around $100msec$, $170msec$ and $250msec$ post-stimulus. Unresolved is how these waveform components relate to coding of visual categories and to what extent they are shaped by learning. To the extent there is any consensus, within the literature the $N170$ has been found to exhibit a greater negative amplitude with increased perceptual experience with a particular stimulus category (e.g. wading birds). Similarly, the $N250$ component has been found to increase with increasing proficiency at identifying individual exemplars within a category. For example, work by Krigolson and colleagues [49] found increased negativity at $N250$ after participants learned to discriminate blob stimuli similar to those used here. However, these and related studies focused on negativity as measured by sensor-averaged signals and did not show whether components such as $N170$ and $N250$ actually carry sufficient information to discriminate or predict the learned visual categories.

In contrast, in our study we went beyond finding raw amplitude differences between

categories and asked whether neural signals support category discrimination. In particular, we demonstrated post-stimulus MEG data can both discriminate and predict subordinate visual categories. Moreover, we were able to identify critical time windows by comparing their respective roles in category learning, finding that the $M100$ component is minimally sensitive to learning and seems to be driven largely by low-order visual processes, while the $M200$ and $M300$ components both become more predictive of visual categories by the end of learning. These results are largely consistent with Krigolson et. al.'s (2009) results and support their claim that the $N250$ is a crucial component in characterizing perceptual learning. We further suggest that the $N250$ component is particularly prominent in visual processing and increased category predictability in the ventral visual pathway, possibly due to an interaction between inferiortemporal and fusiform cortices. More generally, these findings are consistent with previous proposals that place the source of $N170$ in posterior inferior temporal cortex [75, 78, 79, 81] and $N250$ in fusiform areas [76]—a claim that might be further resolved by simultaneous MEG and EEG recordings to establish a better correspondence between the ERP and MEG time components.

In interpreting these results, we would like to note that although we posit specific temporal windows at $M100$, $M200$ and $M300$ as playing important roles in category learning, these markers should not be taken as a strict classification or as markers of mechanisms arising from isolated cortical areas. On the contrary, these components are more likely to arise from functional networks driven by a combination of bottom-up and top-down interactions among cortical and subcortical structures [82, 83], where the measured waveforms are manifestations of cortical systems that exhibit the most robust responses. Future work should examine how visual category learning is communicated interactively among cortical and subcortical areas to achieve efficient categorization, as well as how such communication emerges in category learning.

Finally, we should note that although our study focused on cortical dynamics—the domain in which feedforward visual category coding most plausibly occurs—a separate, yet important, aspect of visual categorization involves feedback learning, often propagated through deeper structures such as the basal ganglia and anterior cingulate cortex. While extensive research [83–87] indicates that basal ganglia and anterior cingulate cortex are crucial in trial-and-error learning and decision making processes such as that employed in our category learning task, detecting neural signals from deep cortical and subcortical structures is

typically not feasible using MEG [2]. For this reason, some category information may also be contained in these neural substrates, but would not be revealed by our analyses due to the depth of these structures and the limitations of the MEG signal.

In sum, our findings support a complementary PFC view of visual category learning. This view is supported by previous work showing both early PFC influences in object recognition processes [52] and category boundary sensitivity within VVP areas [19]. Critically, not only does the VVP carries category-predictive information, but it does so in a time frame that agrees with the predictions of the complementary PFC viewpoint: the VVP increases in its category predictiveness as learning increases. More generally, our work offers an account that uniquely considers *combined* spatiotemporal properties associated with the encoding of subordinate categories, and further, how these properties change over learning. As such, we consider this study to be a starting point for a better understanding of the complex and interactive neural mechanisms underlying visual category learning.

3.5 Appendix: Additional results

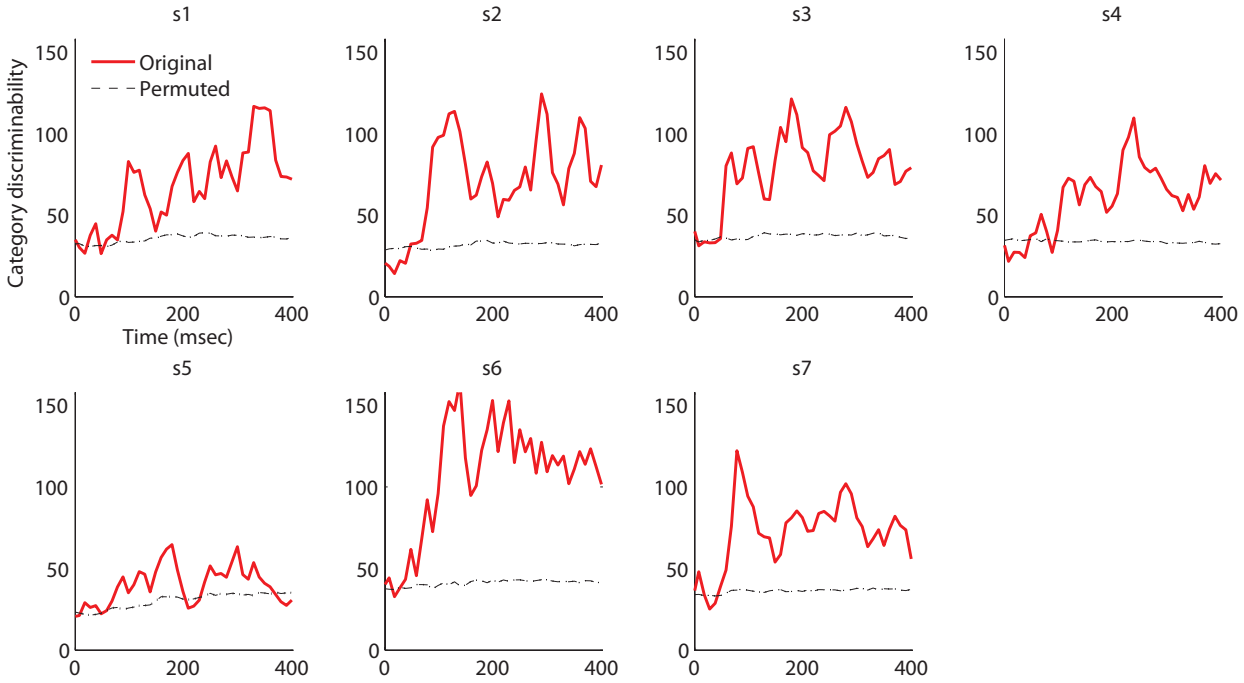

Figure 3.9: **Category-discriminative time course in MEG magnetometers.** Individual category-discriminative time course (visual stimulus onset at $0msec$) compared against discriminative time course computed from trials with shuffled category labels (100 permutations). The 95% confidence intervals of the permuted time course (almost) overlap with the mean.

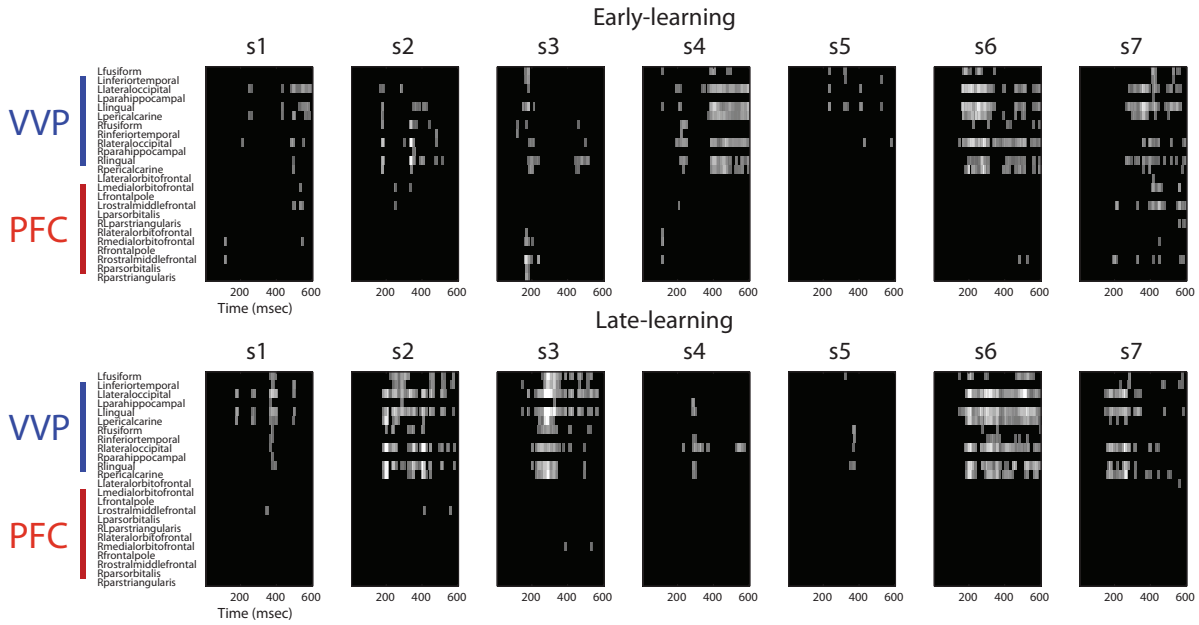

Figure 3.10: **Regions of interest in ventral visual and prefrontal cortices after excursion tests.** The upper row shows regions of interest during early learning for each individual subject. **(B)** The bottom row shows regions of interest during late learning. Regions of interest (in bright color) for each subject was validated using an excursion test that yielded a $p < 0.01$ with 100-fold permutations.

Chapter 4

Experiment II: Learning novel face categories

One of the most common visual categorization tasks is face individuation. Previous research has suggested that a cortical network is implicated in face perception, but little is known about how this network encodes faces during the rapid discrimination time course, and critically, how the temporal code evolves during the learning of new faces. Using the *millisecond* resolution of magnetoencephalography (MEG), we record from the cortex when participants learn to distinguish between two face categories based on prototypes that differ only in facial components. To accurately reconstruct the cortical activities recorded by MEG sensors, we developed a novel method, customized for the face network, that will help make source localization substantially more precise. We use our methodology to test several hypotheses about the face network. First, we show that regions in the ventral visual pathway and prefrontal cortex both encode information about face categories in the time course initially during learning. However, face discriminability in prefrontal cortex reduces over the course of learning, supporting the view that prefrontal cortex is complementary as opposed to dominant in representing visual categories. Next, we demonstrate that temporal discriminability of faces becomes increasingly more prominent from posterior to anterior areas of the ventral stream in the right hemisphere, which is in line with the long proposed hierarchical visual architecture. Interestingly, we find that the coding pattern reverses over time with significant improvement in the inferior occipital gyrus (IOG) and reduced discriminability in the ante-

rior inferiortemporal gyrus (aIT), possibly because of an increased attention to facial parts during the individuation process. Finally, we postulate that the ventral cortex facilitates this shifted coding pattern by better analyzing part-based features for face individuation, instantiated in increased local synchrony between the inferior occipital and middle fusiform (mFus) gyri in the gamma band (30-50Hz) at approximately 100$msec$ post-stimulus toward the end of learning. Overall, we use methodological and theoretical insight to unravel the spatiotemporal learning dynamics of the face perception network, thus providing a starting point for understanding the complex cortical mechanisms involved in visual category learning.

4.1 Theoretical background

Crucial to our conception of objects and beings in the world, the human brain constantly learns new visual categories. Visual categorization can occur at different levels [9, 17], yet one of the most demanding visual tasks is face individuation [88], where a category is mapped to a single identity, and between-category differences are often very small. However, humans recognize faces under a sub-second time course [89], raising the critical question of how the cortex encodes faces so rapidly. And there is an even more fundamental question: how do humans learn novel faces that are often homogenous (e.g. identification of twins in the extreme case [90]), whereby the cortex must support the analysis of fine-level features diagnostic of facial identities. Although we already know much about the neural substrates that correlate with recognition of faces [91–93], we do not completely understand how they encode faces at a fine temporal scale or which cortical changes concomitant with learning facilitate faster discrimination of faces. Both of these questions depend on a characterization of spatiotemporal properties of the cortical face network.

Traditionally, scientists have used either a spatially or temporally driven approach to pursue neural code for faces. A key finding from this research is that a distributed cortical network, commonly known as the "face perception network" or the "face network" [92, 94], meditates the perception and recognition of faces. For example, studies from functional imaging have suggested both a core and an extended cortical networks are involved in face processing [54, 55, 92, 94]. Critical regions including inferior occipital gyrus (IOG, also known as the occipital face area [95]), middle fusiform gyrus (mFus, or the fusiform face area [33, 96])

and superiortemporal sulcus (STS, and posterior STS in particular [41, 97]) are considered the "core" face network, whereas prefrontal cortex or PFC (e.g. inferior frontal gyrus, or IFG, and orbitofrontal cortex, or OFC) [92, 94] and amygdala constitute the "extended" network [91, 97, 98]. More recent works have also found that the anterior inferiortemporal gyrus (aIT)-an area anterior to the fusiform (near the temporal pole)-actively codes face identities [57, 58, 99, 100] and is anatomically linked to the mFus and IOG [101], which makes it an essential member of the existing face network. However, the core and extended networks may not be as straightforward as portrayed, because experiments with face category learning tasks have shown that regions in the ventral visual pathway (VVP) and prefrontal cortex (PFC) both participate in the learning of novel, perceptually similar face categories [102, 103]. But these studies use a learning paradigm that spans days with multiple sessions, so it is possible that cortical changes from such consolidated learning are different from those concomitant with online learning. This paradigm is less susceptible to cortical changes being diluted over time. Overall, due to limited temporal resolution in functional imaging, researchers have not fully explored how face categories are encoded in the face network at a fine temporal resolution.

A separate line of research using electroencephalography (EEG) and magnetoencephalography (MEG) has, with much better temporal precision, discovered neural waveforms that are idiosyncratically evoked in face perception (e.g. $M100$ [40]), individuation (e.g. $M170$ [40] and $N170$ [104, 105]) and familiarization (e.g. $N250$ [106]) tasks. Although previous works have sourced these temporal components to occipital-temporal cortical regions (e.g. [41]), researchers have not addressed how they relate to the face network or, more precisely, how different components in the face network code faces in the recognition time course. In sum, both spatial and temporal approaches have shed light on cortical components related to face processing, but neither provides an integrated account for spatiotemporal dynamics during learning and discrimination of faces in the face network.

Whereas lack of such an account is partially attributed to technological limitations in high-resolution imaging across space and time, it is also due to methodological challenges in designing online learning experiments, and even more crucially, technical difficulties in reconstructing fine-grained temporal cortical activities using existing imaging technologies. Recent advance in MEG has already made possible cortical recording with *millisecond* precision, but only at the cost of a coarser spatial resolution due to the inverse problem [2]

in reconstructing cortical source activities from MEG sensor data, a process referred to as source localization. Because this operation is mathematically highly under-constrained, accuracy of source localization suffers from uncertainty in the estimation and time course of cortical sources. Despite the fact that a large body of work has developed methods for MEG source localization (e.g. [1, 107–110]), the majority of this research seeks generic approaches, and none has designed methods that specifically target the cortical face perception network. Therefore, we are still uncertain to what extent we can utilize MEG to elucidate the fine temporal dynamics of the face network.

With respect to these open issues, we develop a framework that addresses both the methodological and scientific questions. Methodologically, we design a learning experiment in which subjects learn to discriminate two novel face categories based on two prototypes (identities) in an online, feedback-driven paradigm. The two category prototypes are designed to differ only in facial features around the eyes and nose with everything else being identical (see Figure 4.1A), and each category is made up of hundreds of unique faces that are variants from the prototypes (Figure 4.1B). As a consequence, we expect learning to be gradual, and well-learned subjects to increasingly leverage the component-based features for face individuation.

We use MEG to record cortical activities at millisecond resolution throughout the course of learning. Importantly, unlike typical MEG-based studies that often resort to off-the-shelf methods to localize source activities—accuracies of which lack validation—we develop a novel source localization method that uses the face network to provide spatiotemporal constraints in order to improve accuracy in reconstructing cortical activities. Using what we call a *trial partitioning* approach, our method partitions the trial-by-trial face learning data into *early*, *late* and *middle* portions. More specifically, by using trials rather than learning for parameter estimation (*model training*), we build a source model that constrains regions and signature waveforms from a pre-defined cortical face network. We then apply this estimated model to trials from early and late learning (*model validation*). With this step, we can both validate the accuracy of our proposed model and better estimate cortical activities for testing hypotheses about learning. We show that our method reduces source localization error substantially on held-out trials in comparison with a robust classical method.

Scientifically, we use the developed methodologies to test three main hypotheses about the face network. At a broad scale, we evaluate two prominent yet competing theories on

the roles of the ventral visual pathway and prefrontal cortex in visual category learning. The first *dominant-PFC view* [20–23, 45] exerts that PFC plays an exclusive role in representing visual categories with little contribution from VVP. On the other hand, the competing *complementary-PFC view* [24, 44, 50–52] holds that both the PFC and VVP partake in coding visual categories, and the PFC possibly plays a more complementary role by providing top-down guidance to the VVP in forming categories. Our paradigm provides a good platform for evaluating these theories because the two face categories are designed to be perceptually similar (tight boundary), so it provides a strong test for category coding in VVP and PFC. If the dominant-PFC view is correct, we should expect little codability for faces in VVP yet exclusive coding from PFC. On the contrary, if the complemntary-PFC view is right, we should predict VVP and PFC to jointly code for faces in the time course. In addition, we predict that as learning proceeds, the PFC will become less informative about face category memberships if it serves a complimentary role in visual categorization instead of a dominant one.

At a finer scale, we investigate whether cortical time course during face discrimination reflects the *hierarchical architecture in the ventral stream*, primarily focusing on three key regions along the ventral visual pathway: IOG, mFus and aIT. There is a long-standing view that the visual system implements a hierarchical processing scheme [24, 27, 53] such that representation of object categories or face identities becomes increasingly more complex and category-based down the ventral stream. Recent works have challenged this view, and proposed that face processing does not follow this hierarchy, but rather starts from a higher order visual area such as the middle fusiform [111]. However, such observations were made from functional imaging which is suboptimal to account for the rapid time course in face perception. Thus it is questionable whether faces are special in this respect, but more generally, it is unclear whether the hierarchical scheme is necessarily implemented in the online discriminative time course for faces. Our paradigm is ideally situated to evaluate the ventral hierarchical hypothesis at a fine temporal scale. We predict that if the proposed hierarchical scheme holds for faces, we should expect a temporal coding pattern in VVP to roughly reflect a hierarchical structure. More specifically, the aIT and IOG being at the downstream and upstream of the ventral pathway should be best and worst, respectively, at coding face categories, while the mFus (positioned in between) should have mid-level codability for faces.

Finally, as a closely related hypothesis, we examine how face category learning influences *cortical synchrony*. Since our experiment encourages face individuation via components by design, we expect cortical mechanisms most related to processing of facial parts to be increasingly involved in coding faces toward the end of learning—possibly due to supervision or feedback from high-order areas. Several works have suggested that both the IOG and mFus partake in the processing of part-based facial features, but the mFus is more substantially involved in the processing of whole faces. For example, Pitcher et al. [59] used transcranial magnetic stimulation to show that the IOG may be causally related to the processing of facial parts rather than whole faces. We are unsure, however, whether the IOG merely provides an entry-level processing for faces or it contains identity-based information [60, 112]. On the other hand, researchers have found that the mFus is associated with the holistic processing of faces [112, 113], although Nestor and Tarr recently demonstrated that the mFus responds strongly to informative facial parts (or fragments) [114]. Given these findings, it is natural to ask how IOG and mFus should behave and coordinate in our face learning task. We hypothesize that given their differential roles, mFus as a higher order visual area may direct IOG in processing informative facial features during learning. We predict that coding should increase significantly in the IOG over the course of learning if its primary role is to process facial parts diagnostic of face identities, but such a change is unlikely because the IOG acts alone. Instead, mFus should provide some supervision to IOG (perhaps again under the influence of upstream aIT) during the learning process. If, as reported, the mFus is responsible for sifting informative parts for face individuation [114], we expect communication between the IOG and mFus to increase as an indicator of increased guidance from the mFus to the IOG to process category-sensitive facial features.

To preview our results, we find evidence for the complementary-PFC theory such that coding for face categories is distributed across the VVP and PFC, but as we have predicted, PFC carries less information about face identities over the course of learning. We also discover that, initially in learning, temporal coding of faces is indeed hierarchically organized in the ventral visual pathway, with the aIT providing the most information about face categories in the time course, and the IOG providing the least. However, this pattern almost reverses as learning progresses, where the IOG becomes much better at coding the face identities and the aIT becomes less involved. We show that learning increases local synchrony between the IOG and the mFus at approximately $100msec$ post-stimulus, which

suggests that improved face discrimination in the IOG may be a result of supervision or communication from the mFus, whose major role is to help identify informative facial parts in service of efficient face individuation.

4.2 Materials and methods

Participants and ethics statements

Ten right-handed participants (6 females and 4 males) aged between 18 and 35, recruited from the Pittsburgh area, were run in the experiment. Participants were financially compensated for their participation. All experimental procedures were approved by the Institutional Review Boards at Carnegie Mellon University and the University of Pittsburgh. All participants gave written informed consent and were compensated financially for their participation.

4.2.1 Stimulus design

Two novel face categories were created in a fully parameterized space. In particular, each category included a unique set of 364 face images (728 non-repetitive face samples in total) that are slight variations of a category prototype face. The two prototype faces were identical except for the eye size and mouth width (see Figure 4.1A). These two dimensions were systematically varied in a grid-based design space to yield a distinct category boundary as shown in Figure 4.1B, hence successful identification of faces relies on discriminating these facial parts diagnostic of category memberships. In general, faces in category A had larger eyes and a smaller mouth than faces in category B. Figure 4.1C further shows that the two categories are perceptually close but separable in the two-dimensional space projected via principal components analysis. All face images were rendered in 3D and generated using the FaceGen software (http://www.facegen.com/index.htm).

4.2.2 Experimental procedures

The experiment involved a continuous, online learning task where participants were asked to distinguish the two face categories with trial-and-error using feedbacks. The experimental

session consisted of 728 trials and was divided into four equal blocks (182 trials in each) with self-paced breaks in between to reduce fatigue. One additional 30-second break was introduced in the middle of a block to allow for eye blinking.

During each trial, a unique A or B face image (subtending a visual angle of less than 6 degrees vertically and horizontally) was presented for a brief duration (900 $msec$) for the participant to respond. The sequence of A and B faces was randomized for each experiment, and the occurrence of each category was balanced such that there were equal numbers of A and B faces shown within every 20 trials. The presentation of a face image was preceded by a fixation cross in the center of the screen. Meanwhile, a machine-generated sound (630$msec$) that reads either the letter "A" or "B" was played, followed by a slight jitter (120-150$msec$). This sound label was randomly assigned only to prompt the categorical decision, but it did not necessarily indicate the true category membership of a shown face. This randomized scheme helps to decouple the sound categories with the categories of face images, which are our primary interest. The numbers of "A" and "B" sounds were maintained to be equal within every 20 trials.

During the period when the face was shown in a given trial, the participant was instructed to respond with "yes" or "no" to indicate whether the sound label reflects the true category of a face image. The "yes" and "no" signs appeared in the left or right bottoms of the screen with their positions counterbalanced for each session. A glove pad was used for the participant to respond with finger tapping and to prevent arm or wrist movement. After a face image was shown, the fixation cross was shown again during a random jitter of 100-120$msec$. Then a feedback sign of "correct", "wrong" or "too slow" appeared in the screen center for 750$msec$ to inform the participant whether the response is correct, incorrect or missing within the given deadline. The inter-trial-interval was 400$msec$ before the beginning of the subsequent trial. To encourage learning, a small incremental reward scheme was used in the final three blocks, where participants received 3, 5 and 7 US dollars if their average categorization accuracy in block 2, 3 and 4 exceeded 70%, 80% and 90% respectively.

4.2.3 MEG data acquisition and preprocessing

Cortical activities were recorded using MEG while participants performed the face category learning and localizer tasks. All experiments were conducted in an electromagnetically

shielded room with participants seated comfortably and head-fixed throughout. Neural data were recorded using a 306-channel whole-head MEG system (Elekta Neuromag, Helsinki, Finland). The system has 102 channels, each a triplet consisting of a magnetometer and two perpendicular gradiometers. The data were acquired at 1 kHz, high-pass filtered at 0.1 Hz and low-pass filtered at 330 Hz. Eye movements (EOG) were monitored by recording differential activity of muscles above, below, and lateral to the eyes. These signals captured vertical and horizontal eye movements as well as eye blinks. Electrocardiography (ECG) was recorded by placing two additional electrodes above the chests. Four head position indicator coils were placed on the subject's scalp to record the position of the head in relation to the MEG helmet at the beginning of each session. These coils, along with three cardinal points (nasion, left and right pre-auricular), were digitized into the system and were later used for source localization.

The data were preprocessed using the signal space projection method [115] to remove artifacts such as eye blinks or movements. The MEG signals were bandpass-filtered between 0.1 and $50Hz$ to prevent power-line interference at $60Hz$. We discarded any trials that showed EOG or ECG activities that were three standard deviations away from the trial mean at any point. And for each trial, we removed any baseline defined as 120-0$msec$ prior to the onset of the visual face stimulus by subtracting its mean off every point along time.

Cortical source estimates using the Minimum Norm Estimates (MNE) [1, 2] were computed using the MNE Suite software (http://www.nmr.mgh.harvard.edu/martinos/userInfo/data/sofMNE.php). Source dipoles were evenly distributed ($7mm$ separation between neighboring sources) with orientations fixed normally to the cortical surface. Surface brain models for each individual participant were constructed by Freesurfer software (http://surfer.nmr.mgh.harvard.edu/) from structural magnetic resonance imaging scans acquired at the Scientific Imaging and Brain Research Center at Carnegie Mellon University (Siemens Verio 3T, T1-weighted MPRAGE sequence, 1×1×1mm, 176 sagittal slices, TR = 2300$msec$, TI = 900$msec$, FA = 9 degrees, GRAPPA = 2). Based on the neural anatomy of each individual participant, anatomically constrained cortical regions were identified from Freesurfer segmentation using the Desikan-Killiany Atlas [77].

4.2.4 Modeling cortical activities during face category learning

Our source localization method follows three main steps: 1) Registration - defining regions of interest (ROIs) in the face network, 2) Training - estimating source model parameters from trials in midst of the face category learning task, and 3) Testing (and application) - validating and applying the model on trials during the initial and final phases of learning. The Appendix of this chapter discusses step 1. For each individual subject, trials from the face learning experiment were divided into early (160 trials from the first block), middle (>300 trials during the middle) and late (160 trials from the final block) portions. The middle portion was then used for step 2, and early and late portions were used for step 3. In Chapter 6, we describe the method in full detail and show that it yields better localization accuracies for trials in the early and late parts of learning.

4.2.5 Time domain discriminability analysis

To measure category discriminability from each ROI in the face network, t-tests were applied over time (for $600msec$ duration after onset of face stimulus in a trial) to contrast the mean source activities from trials containing A or B faces—this was computed for trials during early and late learning separately. More specifically, within each ROI, the mean source activities were first obtained. These time series data (per trial) were further divided according to whether a trial belongs to the A or B category. Then, for every $10msec$ window, a t-test was performed based on ROI source currents averaged within that time bin, which helps to evaluate the null hypothesis that the mean responses are equal under A and B conditions. A t-statistic as an indicator for category-discriminability was then obtained for each available time and face ROI.

Additionally to the discriminability time course, an excursion test [3] was used to evaluate the significance of "hot spots" across time. This involved a number of steps. First, discriminative time course was thresholded and only contiguous time points that exceeded the threshold were proposed as potential regions of interest. Contiguity was satisfied if any of the immediate neighbors of a given point in time also passed the threshold—this procedure helped to prune isolated events that are likely to occur due to chance. This same procedure was then applied to the same data multiple times (100-fold permutations), but in each case,

category labels were shuffled—this provided a baseline measure, or a null distribution. A p-value was then computed for each individual subject using a standard permutation test by comparing the discriminability statistics within the proposed regions of interest to those in the permuted data following procedures described in [3].

4.2.6 Time-frequency domain phase-locking analysis

To quantify phase locking between two ROIs in the face network, a standard Hilbert transform (see [116]) was used to obtain phase information over time ($600msec$ duration in step of $1msec$ after visual onset) and frequency (0 to $50Hz$ in step of $2Hz$ to avoid power band oscillation at $60Hz$). Differing from the time domain analysis, cortical source activities were first detrended trial-by-trial with a $50msec$ moving average smoother and baseline removal. This step ensures that the time-frequency phase locking analysis only operates on the residuals of the time series. The detrended data were then Hilbert-transformed with a finite-impulse-response filter with a filter order of 100—edge effects were minimized by taking the transform in a much longer time window ($>300msec$ extra at both ends) with edges removed. For every source dipole in the ROI, the phase angle at each time and frequency was then computed. Finally, for every possible pair of source dipoles from two ROIs, their phase angles (ψ_i and ψ_j) were used to compute the phase-locking value (PLV; a value between 0 and 1) at each time-frequency entry:

$$PLV = \frac{1}{N}[|\sum_{n=1}^{N} e^{j\psi_i^n - \psi_j^n}|] \tag{4.1}$$

where $n = 1, ..., N$ indexes trials. This measure was used to quantify synchrony between two sources across trials, with a value of 1 indicating perfect synchrony. The mean PLV value was then calculated based on the average of PLVs across all pairs. To take into account baseline phase-locking activities, mean PLVs from 120-0$msec$ prior to visual onset were subtracted off at each time-frequency entry.

4.3 Results

4.3.1 Behavioral face learning performance

To evaluate whether participants learned the face categories, we compared their performance in the initial (*early-learning*) and final (*late-learning*) blocks of the experiment, each of which contains 160 trials. We kept the numbers of trials with A and B face stimuli identical across early and late learning to reduce bias. We assigned a score of 1 to trials in which participants correctly identified the face category and 0 to trials in which they were incorrect or failed to respond within the 900-$msec$ deadline. We then aggregated and normalized these scores into a 0-1 range (1 for 100% accurate) for the early and late stages of learning. Figure 4.1D shows that all subjects learned the task with their mean terminal accuracies exceeding 80% (five subjects reached 90%). All subjects showed significant improvement in the categorization accuracy from early to late learning ($p < 0.05$ from binomial tests with Bonferroni corrections), thus suggesting effective learning. Figure 4.1E summarizes the trial-averaged response times in the first and final blocks of the experiment. Seven subjects showed significant reduction in response time ($p < 0.05$ from t-tests with Bonferroni corrections) over the course of learning.

4.3.2 Cortical source localization

To improve the precision in localizing cortical source activities with MEG, we used a novel source localization method (see Chapter 6) following three steps. Firstly, we defined cortical regions of interest in the face network for each subject. We found eleven regions of interest by contrasting face and object conditions in a one-back MEG localizer task (see Appendix of this Chapter for detailed procedures), which include bi-hemispheric regions in the vicinity of the reported "core" network [92, 94] in the occipito-temporal ventral cortex, e.g. IOG, mFus and STS, regions near the temporal pole-aIT, and the "extended" network [92, 94] in prefrontal cortex, particularly IFG and OFC (we did not find OFC in the left hemisphere). Secondly, we estimated a cortical source model (Chapter 6) by applying differential shrinkages to regions within and outside the identified face network. In particular, we estimated the model parameters from trial data collected in the middle portion of the learning experiment.

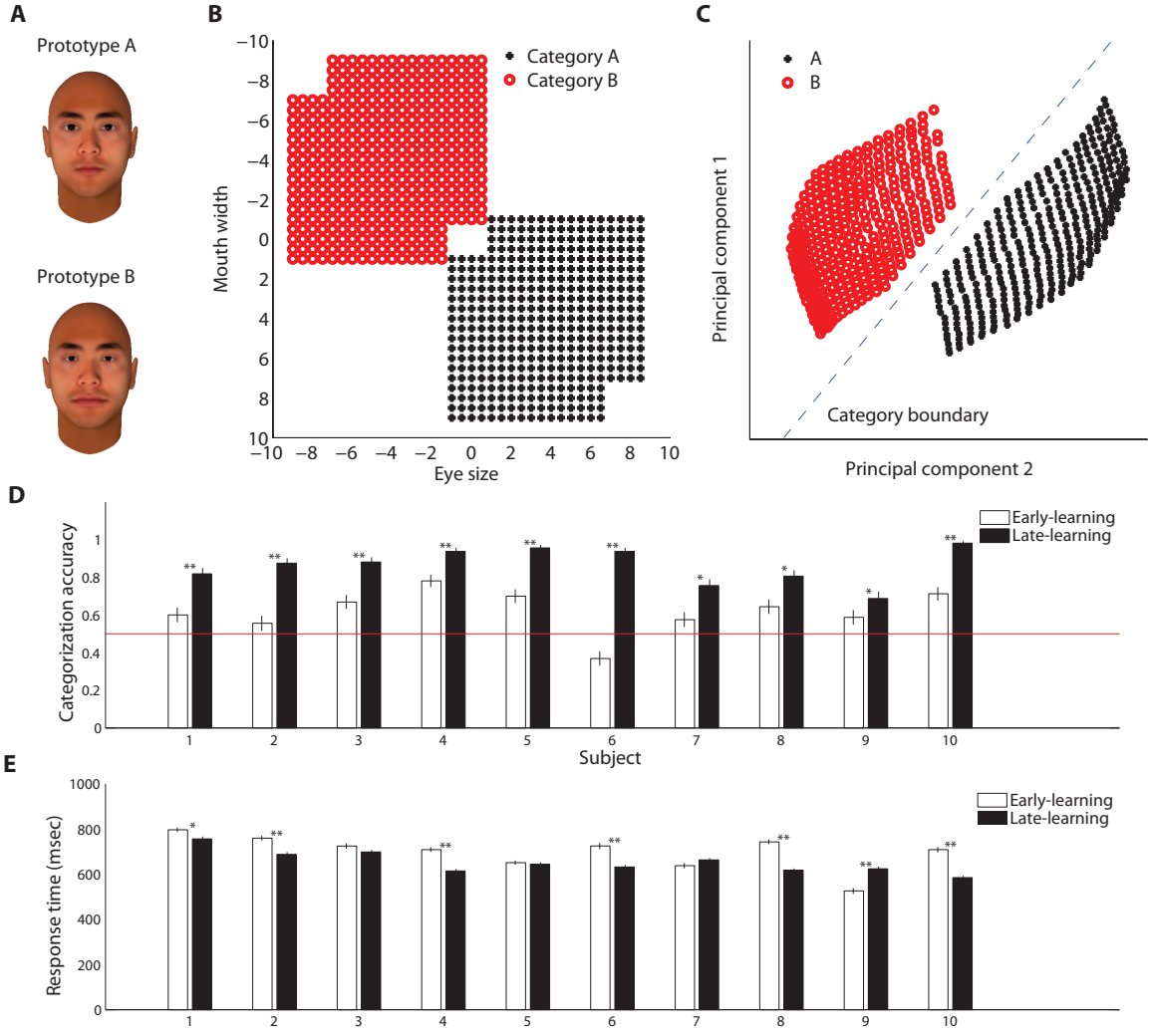

Figure 4.1: **Summary of stimulus design and behavioral performance.** **(A)** Prototype images for the face categories. **(B)** Two categories of faces parameterized along the eye and mouth dimensions. **(C)** A low-dimensional representation of face categories via the principal components analysis. **(D)** Mean categorization accuracies in the first (early-learning) and final (late-learning) blocks of the experiment. **(E)** Mean response times during early and late learning. "*" and "**" indicate significance at $p < 0.05$ and $p < 0.005$ with Bonferroni corrections across subjects.

Finally, we applied the estimated source model to localize trials in the early and late stages of learning.

We compared the accuracy of source localization between our method and a popular method called the minimum-norm estimates (MNE) [1, 2]. Figure 4.2A show the mean-squared error (MSE) in reconstructing sensor signals from estimated source activities for trials in early and late learning. Across both stages of learning, our method performed significantly better than MNE ($p < 0.022$ from binomial tests across subjects) in reducing the MSE. Figure 4.2B shows the result for 10 individual subjects. In all cases, our method outperformed MNE by producing smaller MSEs, hence improving the quality of localized cortical activities for the subsequent analyses.

4.3.3 Category-discriminative time course in the face network

Our first analysis explores whether category-discriminative information is present in the face network during online face discrimination. Specifically, we test the hypothesis whether such information flows predominantly in prefrontal cortex in comparison with the ventral visual pathway. To do this, we computed the degree to which face categories can be discriminated in time course of each region of interest in the face network. We partitioned the trials according to whether they contained A or B faces during early and late learning respectively. We then computed discriminability in steps of $10msec$ for a 600-$msec$ duration after the visual stimulus onset.

Figure 4.3 summarizes the result. The temporal traces in each panel represent the degree to which A or B faces can be discriminated from time course of a cortical region at the group level. Initially in learning (black dashed lines), prefrontal regions IFG and OFC, together with the aIT and the mFus in the right hemisphere show better discriminability $200msec$ post-stimulus than elsewhere. In particular, discriminability in the aIT is no weaker than the IFG in the time course. This provides evidence against the proposal that the PFC dominates in coding face categories. This observation is also consistent with recent proposals that aIT is involved in identity-based coding (e.g. see [57, 58, 99, 100]).

To evaluate the PFC-complementary view, we examined temporal coding in the final stage of learning. We predicted that if PFC plays a complementary role, it should exhibit a decrease in category-discriminability over the course of learning as behavioral face categorization

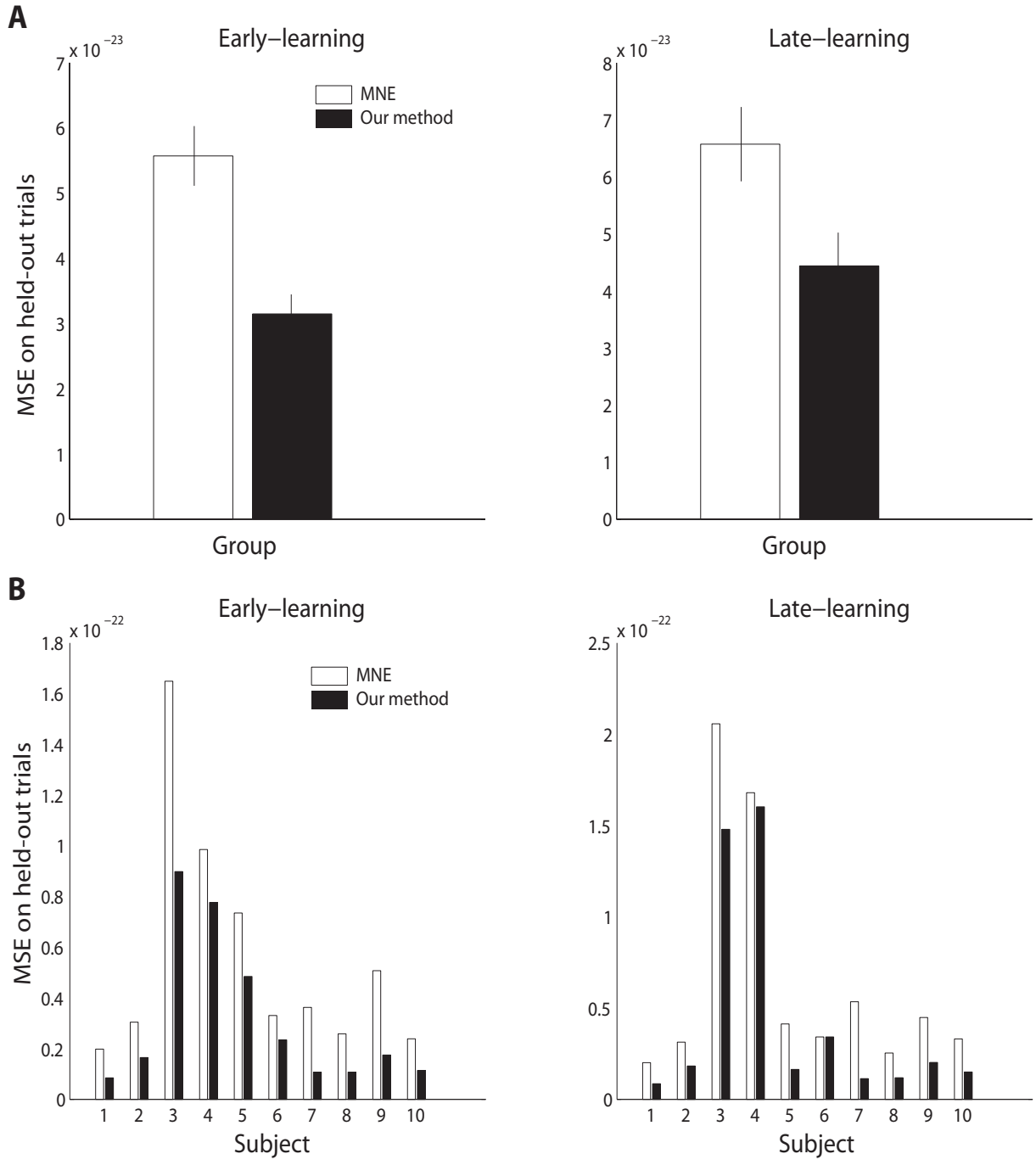

Figure 4.2: **Comparison of mean-squared error (MSE) in source localization for held-out trials. (A)** Group-level comparison of MSE in reconstructing sensor signals from from source activities during early and late learning. The vertical bars represent standard errors of the mean. **(B)** Comparison of MSE across 10 individual subjects. MNE stands for the minimum-norm estimates [1, 2].

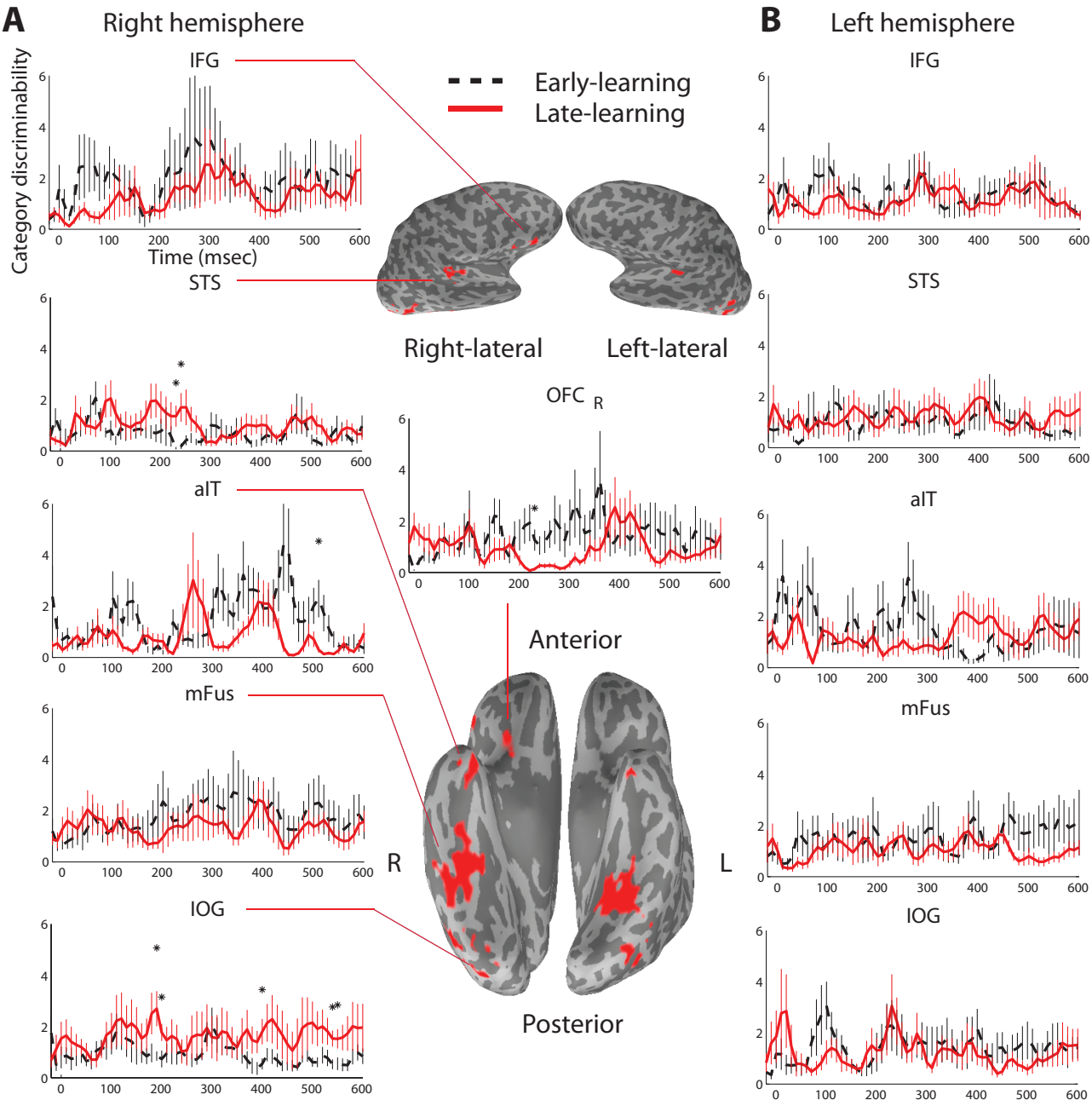

Figure 4.3: **Category-discriminative time course in the cortical face network during early and late learning. (A)** Group-average discriminative time course from the face network in the right hemisphere. **(B)** Discriminative time course in the left hemisphere. Inflated cortical surface is taken from a single subject with the face network identified in the inferior frontal gyrus (IFG), orbitofrontal cortex (OFC), superior temporal sulcus (STS), anterior inferiortemporal gyrus (aIT), middle fusiform gyrus (mFus) and inferior occipital gyrus (IOG). "0" on the time axis marks the visual onset. Vertical bars represent standard errors of the means. Asterisks indicate significant difference in discriminability between early and late learning at $p < 0.05$.

becomes more proficient. Figure 4.3 (solid red lines) shows that discriminability in both right-hemispheric IFG and OFC exhibits a diminishing trend toward the end of learning. In particular, right OFC exhibits significantly reduced discriminability ($p = 0.047$ from t-test) post-$200msec$ in the time course. This suggests that PFC is more actively involved in encoding categories initially, but it carries less information about faces as learning progresses. Similar regions from the left hemisphere play a relatively minor role in face coding.

Figure 4.3A shows, however, that along with PFC, the right-hemispheric aIT also exhibits less category discrminability post-$400msec$ ($p = 0.038$ at $500msec$ from t-test). Instead, posterior regions of the VVP, i.e. the right-hemispheric IOG, shows significant increase in discriminability for a sustained period ($p < 0.035$ between 170 and $230msec$; $p < 0.01$ post-$400msec$) during late learning. This set of findings suggests a possible shift in strategy in face discrimination. It is plausible whereas participants initially categorized faces based on the whole face (reflected by identity-based coding in aIT), they gradually shifted to a part-based strategy that focused around the eyes and mouth (part-based coding primarily in IOG). This result confirms previous findings about the functional role of the IOG in processing parts during face recognition [59, 60].

Along with the increased discriminability in IOG, we also observed significant increase ($p < 0.009$ from t-test) in discriminability in the right-hemispheric STS near $200msec$ post-stimulus. This is likely to be associated with expression-based strategies in face individuation. For example, a number of subjects reported they had perceived B faces as conveying sadness (or A faces as conveying happiness). However, such expression-based strategies were less consistently reported than part-based strategies that attend to the eyes and mouth.

To further assess whether the cortical time course contains category information better than chance, we conducted an excursion test following [3]. This procedure allowed us to evaluate the statistical significance of "hot spots" in the discriminative time course on a subject-by-subject basis. To do this, we obtained category-informative time points (in each region) by thresholding the time course at 3 and kept contiguous time points that passed the threshold. We evaluated the significance of these hot spots by comparing the discriminability statistics within the proposed time points against the statistics within regions found from a permutation test (100 folds, each of which category labels were permuted across trials—forming a null distribution). This procedure yielded a combined $p < 8 \times 10^{-10}$ across subjects (Fisher's method) for the hot spots identified during early and late learning (4 subjects had

$p = 0.01$ from the permutation test; 6 subjects had $p = 0.02$).

Figure 4.4 compares category discriminability between early and late learning in the face network. Each bar in the figure indicates the normalized tally (aggregated across subject) of hot spots after the excursion test for a particular region. The most notable differences are a substantial reduction in prefrontal regions and a 5-fold increase in hot spots in the right-hemispheric IOG. (Figure 4.8 also shows the tallies for each individual subject in selected prefrontal and ventral regions. Note that the pattern for IOG is consistent across 7 of 10 subjects.) Additionally, we also observed a general reduction across regions in the left hemisphere, suggesting face individuation becomes more biased toward the right hemisphere as learning progressed, consistent with the view that face perception is more dominant in the right hemisphere [54, 55]. In sum, this set of findings supports a distributed temporal coding scheme and the complementary-PFC view, where PFC and VVP both participate in encoding face categories initially, yet the VVP, particularly the right IOG, becomes better at face discrimination toward the late stage of learning.

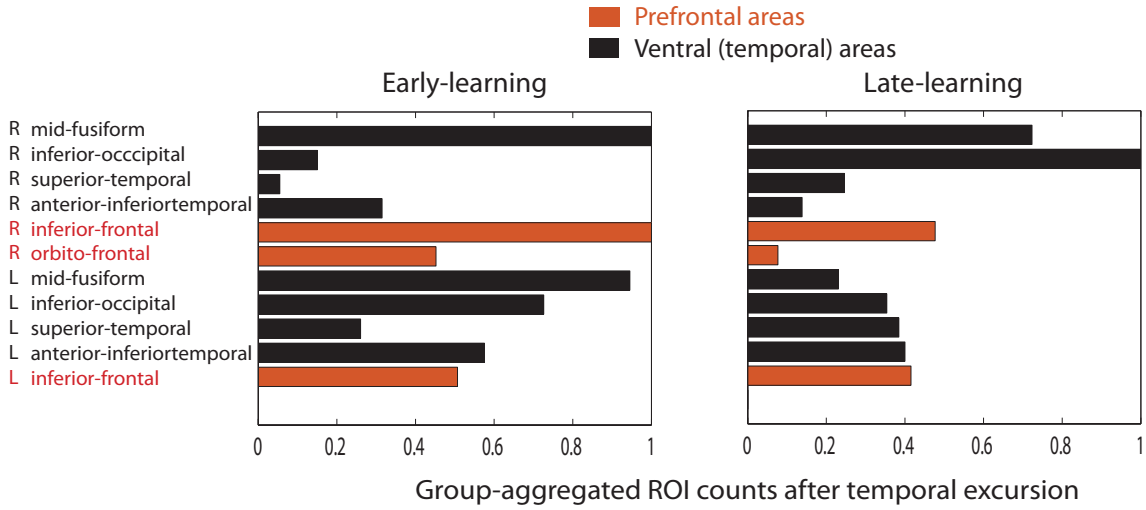

Figure 4.4: **Category-discriminative information in the face network with excursion.** Bar lengths correspond to tallies of hot spots in the discriminative time course pooled across subjects and normalized over 11 regions of interest. The hot spots were validated using the excursion procedures in [3]. The comparison between early and late learning shows a reduction of hot spots in prefrontal regions and left hemisphere but an increase in the right infeior-occipital gyrus.

4.3.4 Hierarchical face coding in the ventral visual pathway

Our second analysis focuses on face coding in three key regions positioned along the ventral visual pathway: the IOG, mFus and aIT in the right hemisphere. Figure 4.5A illustrates our hypothesis. We predict that if face identities are represented in a hierarchical scheme along the ventral stream, the degree to which they can be discriminated in the time course should roughly reflect the hierarchical organization of the IOG, mFus and aIT in the ventral pathway. More specifically, the aIT in the vicinity of upstream VVP should be best at coding face identities, followed by the mFus, and then the IOG.

Using the same discriminability measure as in Section 4.3.3, we computed and overlaid the time course of the IOG, mFus and aIT. Figure 4.5B (left panel) shows that initially in learning, the temporal face coding pattern corresponds to our prediction: IOG being upstream is least discriminative about face identities; aIT being at the top of hierarchy shows best discriminability post-400$msec$; mFus being in the middle has an intermediate level of discriminability risen after 200$msec$. With this result we confirm that temporal coding for faces is organized hierarchically along the ventral stream. Our finding also suggests that early in learning, participants most likely used a holistic strategy in categorizing the faces, where each category is represented by an identity, as encoded in the hierarchically organized ventral stream where aIT holds most of the identity information.

Interestingly, we found this hierarchical coding scheme to be reversed toward the end of learning. Figure 4.5B (right panel) shows that IOG becomes increasingly discriminative post-100$msec$ in the time course, but aIT is less discriminative on average and mFus maintains an intermediate-level discriminability. This shift in the temporal coding pattern matches our hypothesis that the strategic shift to part-based discrimination is associated with an increased involvement from the IOG. With our finding, we confirm previous proposals about the role of the IOG in the processing of facial parts [60], but we also show that processing at the entry level of the VVP can be informative about face category memberships. To evaluate the significance of the shifted coding pattern in VVP from early to late learning, we conducted multivariate Hotelling's t-tests point-wise along time. Specifically, we tested the null hypothesis of whether the mean discriminability in the aIT, mFus and IOG (as a three-dimensional variable) is unchanged from early to late learning—we ran this test across the time axis, and a χ^2 statistic at each available time point indicates deviation

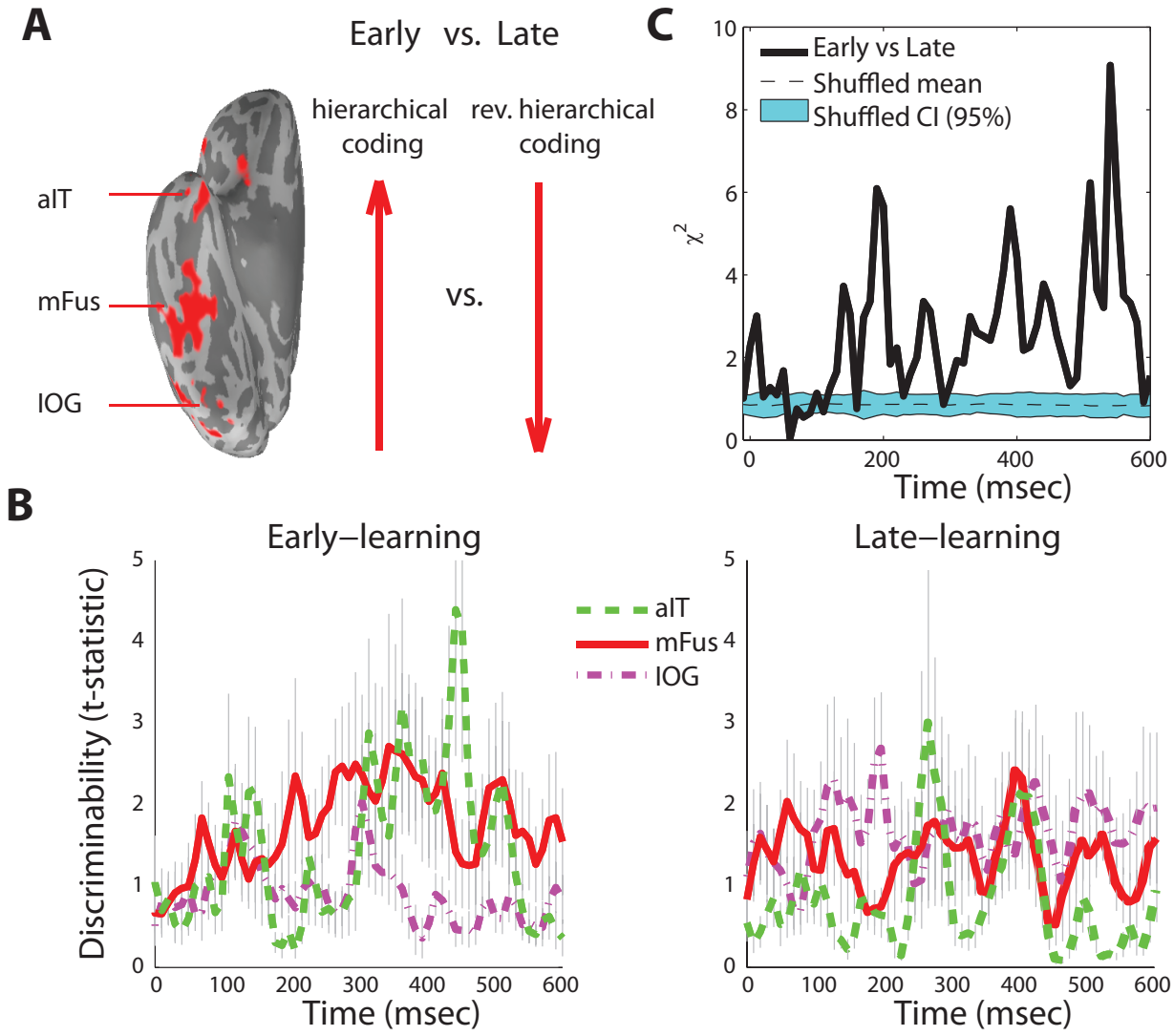

Figure 4.5: **Temporal coding of face categories in the ventral visual pathway.** **(A)** Illustration of the hierarchical coding hypothesis. **(B)** Temporal face codes along the right-hemispheric ventral visual pathway during early and late learning. Discriminability in the anterior inferiortemporal (aIT) and inferior occipital gyrus (IOG) reversed during late-learning possibly due to increased reliance on diagnostic facial parts. mFus represents middle fusiform gyrus. **(C)** MANOVA χ^2 statistic time course evaluating difference in temporal codes across early and late learning. Cyan-shaded regions correspond to the baseline measure from shuffled data. "0" on the time axis marks the visual onset.

from the null. The result appears in Figure 4.5C. We observed that the most appreciable difference in coding patterns occurs approximately post-100$msec$, which reflects the increased discriminability in the IOG and decreased discriminability in the aIT during learning. To ensure these statistics are not due to chance, we used a permutation test that repeats the entire procedure 100 times using shuffled trials. For each shuffled data set, we permuted trials across early and late learning stages, keeping the number of A and B trials identical—this establishes the null hypothesis that the temporal coding pattern does not exhibit change over the course of learning. Figure 4.5C shows bands containing probability 0.95 under the null hypothesis. Note that the statistics from the original trials (without permutation) are significantly higher than chance-level most notably post-100$msec$ with a global significance at $p < 0.01$. Thus we confirmed that temporal codes in the ventral stream changed possibly due to adaptations to strategic shifts in face categorization during learning.

4.3.5 Cortical synchrony during face category learning

Our final analysis explores whether the improved face discrimination in IOG is due to synchronous activities from mFus as a possible form of supervised learning. Since we cannot directly assess the causal relationship between these two regions, we used cortical phase-locking (or synchrony) as a proxy. We postulate that if mFus plays a role in supervising IOG to process features informative for face individuation, there should be an intensified local communication between mFus and IOG as a result of learning.

To test this prediction, we conducted a time-frequency analysis that calculates phase-locking between IOG and mFus in time course during early and late stages of learning. In particular, we computed the phase-locking value between these regions within the same window used for the time domain analysis and investigated a frequency range from 0 to 50Hz. We then computed a statistical map based on phase locking value (PLV; a value between 0 and 1) maps across subjects, where a p-value was obtained for each time-frequency entry on this map to test the null hypothesis of whether the group-averaged PLV is equal to 0 (for no phase locking)—a significant PLV should yield a small p-value.

Figure 4.6A shows the mFus–IOG phase locking for early and late learning. Both maps show significant (e.g. $p < 0.001$) synchrony between the mFus and IOG in the *alpha* and *beta* bands ($\leq 20Hz$) during 100-200$msec$ post-stimulus. This time window is expected, and it

corresponds with previous accounts of $M100$ and $M170$ windows for face processing [40, 117]. But critically, we found a significant increase (e.g. $p < 0.001$) in synchrony at a much higher frequency band, the gamma-band ($30 - 50Hz$), post-$100msec$ during late-learning—this synchronous activity between mFus and IOG is sustained until about $500msec$ in the time course. As we have predicted, this high-frequency synchrony is most likely due to learning-induced communication between mFus and IOG, where mFus directs IOG in processing informative facial features for face discrimination.

We also repeated the phase-locking analysis between the mFus and aIT and regions in PFC such as the OFC and IFG. We found there is an increase in mFus-aIT (Figure 4.6B), mFus-IFG (Figure 4.6D) synchrony near $150msec$ (e.g. $p < 0.001$) post-learning in the *beta* band (15-$25Hz$), and a separate increase near $350msec$ in mFus-aIT in the *gamma* band (35-$45Hz$), possibly due to a a sharpened directive influence from these regions on mFus in face category learning, but the scale of these increases is much smaller than the mFus-IOG interaction in the *gamma* band post-learning. Finally, we observe increased synchrony in mFus-OFC during late-learning (Figure 4.6C), where such interaction is almost not present at all during early-learning. The bursty synchronous activities in the *gamma*-band seem to resonate with the *gamma*-band synchronous pattern observed between the mFus-IOG (during late learning), which suggests a PFC-VVC feedback loop during the face individuation process. Overall, these results show that an increasingly coordinated communication between the mFus and IOG is a prominent indicator for efficient face individuation with attention to facial components.

4.4 Discussion

We presented a framework for exploring fine-grained temporal dynamics of the face perception network during online learning of novel face categories. Methodologically, we proposed an alternative source localization method for MEG tailored to a learning paradigm and the face network. By partitioning trials into different stages of learning, we utilized data in the middle of learning—otherwise often under-used or discarded in testing hypothesis about learning—to estimate a source model constrained on spatiotemporal properties of a pre-defined face network. In turn, this model offered better estimation on cortical activities during initial and final learning that are more relevant to scientific evaluation of the learning

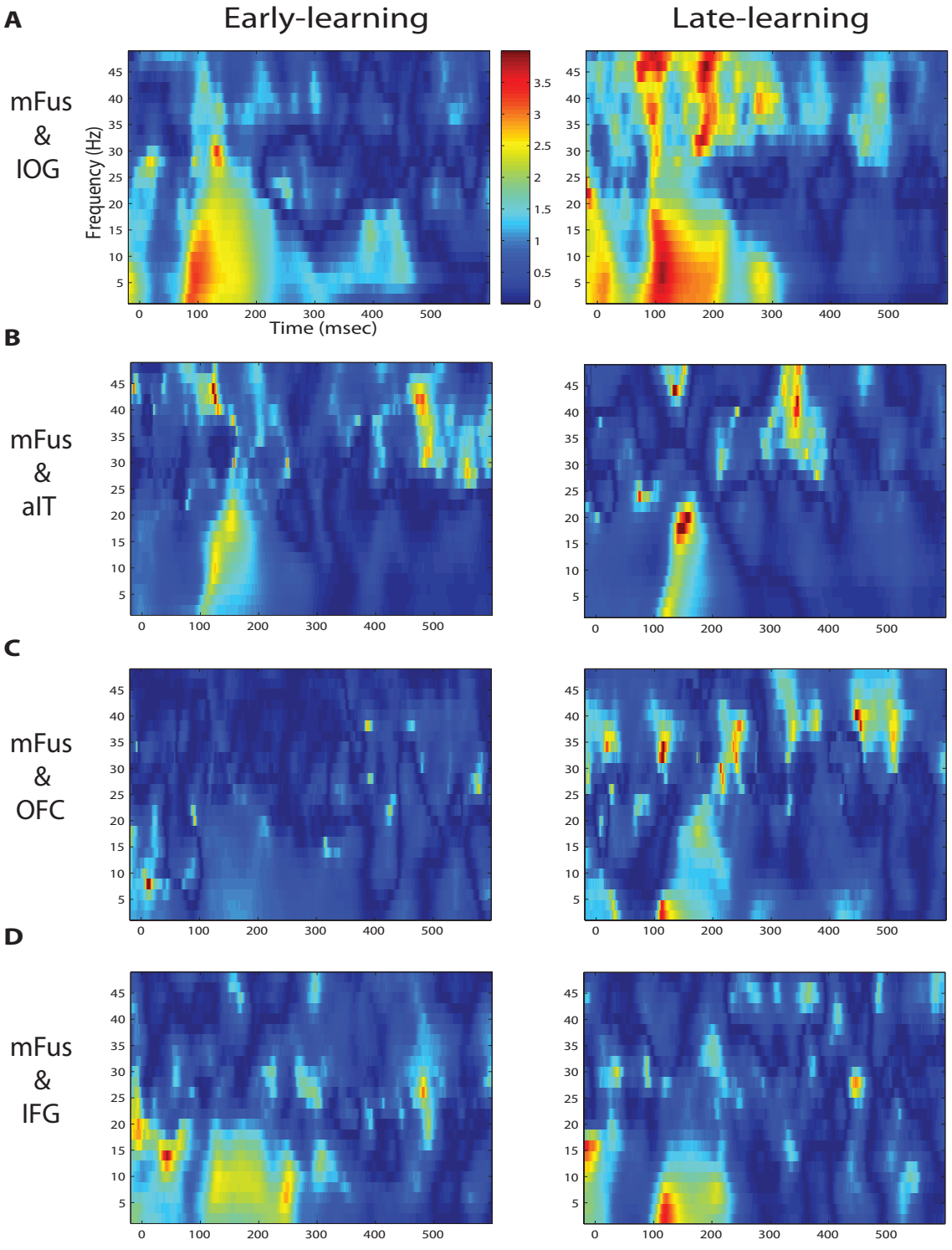

Figure 4.6: **Synchrony in the cortical face network during face learning. (A)** Statistical map of time-frequency phase-locking between the right-hemispheric middle fusiform (mFus) and inferior occipital gyrus (IOG) during early and late learning. **(B,C,D)** Statistical maps of phase-locking between the mFus and anterior inferior temporal gyrus (aIT), orbitofrontal cortex (PFC) and inferior frontal gyrus (IFG) respectively. "0" on the time axis marks the visual onset. Color bar indicates negative logarithmic p-value with base 10, e.g. "3" corresponds to group-level significance at $p < 0.001$ and "2" corresponds to $p < 0.01$.

dynamics. Theoretically, our spatiotemporal approach also allowed us to evaluate a number of hypotheses about the face network during a rapid, online discrimination time course.

At the cortex level, we showed that both the ventral visual pathway and prefrontal cortex are involved in encoding information relevant to face category memberships. But we found that, over time, prefrontal cortex become less discriminative about face categories, which suggests that PFC plays a greater role initially in helping map out the face identities and a lesser role as categorization becomes more efficient. Such a transition supports the complementary-PFC view [24, 44, 50–52], which states that the VVP and PFC together code for faces and, as previously proposed, but PFC does not necessarily dominate in representing visual categories [20–23, 45]. In Chapter 3, we have found similar support for the complementary role of PFC, where in learning highly similar visual categories, the VVP becomes better at encoding the category memberships.

In more confined regions within the ventral visual pathway, we confirmed previous theories that proposed hierarchical object (and identity) representation along the ventral stream [53, 113]. We demonstrated a hierarchical coding pattern for faces evident in time course of IOG, mFus and aIT. Importantly, we showed that this coding scheme is reorganized in adaptation to face individuation strategies relying more heavily on facial components. This suggests the hierarchical visual architecture is not strictly enforced, but instead it flexibly changes to suit the demand in visual categorization. Our result also indicated that a cortical basis for configural and component-based processing may be implemented in separate circuitries, with the anterior and posterior ventral cortices oriented more toward whole-face and part-based processing, respectively.

Finally, to investigate cortical interactions, we conducted a time-frequency phase locking analysis. We found synchronous activities between mFus and IOG in *alpha* and *beta* bands ($<25Hz$) during 100-200$msec$ post-stimulus in both early and late learning. This time window overlaps with signature waveforms (e.g. $M100$ and $M170$) reported in face perception [118], suggestive of a low-frequency interactive processing mechanism possibly due to long-range cortical feedback loops. Critically, we found that learning increases IOG-mFus synchrony in *gamma*-band (30-60Hz) starting at roughly the same time period. This high-frequency phase locking, concomitant with increased face discriminability observed in the IOG, suggests that feedback and feedforward mechanisms between the two regions may be key to part-based face coding. In particular, we hypothesize that the mFus provides top-down

supervision in directing the IOG to process informative facial fragments (e.g. [114]). In return, the IOG passes along component details that are processed at the entry level (e.g. [60]). A recent computational model of the mFus has suggested that the mFus serves to magnify differences in subordinate categorization (e.g. individuation of face identities) [119], which explains its functional role in the categorization of general objects other than faces [56]. Our result provides additional evidence for this proposal. In our case, because category boundary is defined by differences in facial parts, in order for mFus to magnify such differences, it should retrieve part-based information from IOG—this could explain the increased local synchrony between mFus and IOG during learning. A question for the future is to examine is whether learning necessarily mediates sparse coding in the mFus (e.g. [120]). We speculate that neural populations in the mFus should be active initially, but many of them may be suppressed as learning progresses to shape a sparse code in favor of a smaller number of neurons that are more relevant to the encoding of diagnostic facial features only.

Our work used parameterized face categories instead of real-world faces as visual stimuli, but we do not suggest that face recognition necessarily relies on simplistic strategies using part-based features. In fact, real-world faces can be highly irregular and asymmetric, as previous research has found a mixture of configural (holistic) and componential (part-based) processing in face perception [121, 122]. However, our design helped to reduce the amount of variability in the processing of real faces, thus allowing us to explore face individuation in a much more controlled environment. This allowed us to make more accurate predictions about the cortical regions associated the learning of these face categories. Given that cortical regions known for face processing are also recruited under artificial faces and other face-like stimuli [123], cortical principles found here should be generally applicable to the individuation of real faces.

Over the past decade, it has been suggested that a cortical network is implicated in face processing [92, 94]. Our framework adds to the existing literature on spatiotemporal dynamics in face recognition [57, 61, 124] by elucidating cortical properties associated with learning and online face discrimination at a fine temporal scale. Our findings suggest that successful individuation of novel faces is achieved by highly adaptive, flexible transformation in collective temporal codes from a cortical network including ventral visual and prefrontal cortices. The methodologies we developed should provide a starting point towards unveiling the complex spatiotemporal properties of the human cortex in visual category learning.

4.5 Appendix I: Defining regions of interest in the face network with MEG

This section discusses methods for defining cortical regions of interest (ROIs) in the face perception network using MEG. In particular, we used three main procedures: 1) A face localizer task in MEG, 2) Sensor-space excursion for defining the temporal region (window) of interest and 3) Source-space searchlight for defining face ROIs.

The face localizer task

In addition to the main face category learning experiment, a separate, short MEG localizer task was run for each participant. The experiment involved a 1-back task designed similarly to common localizer paradigms in functional imaging that help to locate category-selective cortical regions for faces and other object categories. The visual stimuli in the experiment included color images of four categories: faces, everyday objects, houses and scrambled objects. For our purpose, trials from face *versus* object conditions were used to define face ROIs. During the experiment, the images were presented (subtending a visual angle of less than 6 degrees vertically and horizontally) one at a time and each category was presented in blocked trials, with each block comprised of 16 trials (16 images). There were 12 blocks in each run with 3 blocks of each condition, and each participant had 4 runs, yielding 192 images for each category. Each image stimulus was presented for $800ms$ with a $200ms$ inter-trial interval and $6s$ fixations between each block. A 1-back identity task was used to maintain attention throughout the experiment. Participants pushed a button on the glove pad to indicate whether the current stimulus was the same as the preceding stimulus.

Temporal ROI excursion

Before locating cortical ROIs for faces, an excursion algorithm was first used to define a time window of interest. Trials in face and object conditions from the localizer task were used for all of the analysis. In particular, MEG sensor data (trials) were first divided into faces or objects based on the stimulus category of a given trial—this yields approximately 192 trials for each category by design. Then, a multivariate Hotelling's t-test was applied across time to

evaluate the null hypothesis of whether the mean sensor signals recorded from face condition are equal to those recorded from object condition. To do this, magnetometer signals at 102 distinct sensor locations were first collected to form a 102-element multidimensional vector for each trial. Then for every $10msec$ (binned) for the $400msec$ duration after stimulus onset, these high-dimensional vectors were mapped to a low-dimensional space via principal components analysis, with at least 99% variability preserved at each point—this step prevents singular matrix inversion in the t-test. Following this step, a Hotelling's t-test was then performed to compute a single p-value that reflects deviation from the null hypothesis at a certain time.

Figure 4.7A shows the group-level p-value time course from the t-tests. A robust and strongly discriminative peak was found in the vicinity at $170msec$ post-stimulus (e.g. $p < 10^{-10}$), or $M170$ [40] as also reported by [125]. This was followed by a second less prominent peak at approximately $300msec$ ($p < 10^{-8}$). But Figure 4.7B shows that at the individual level, $M170$ is a much more consistent response than at 300msec (Table 4.5 summarizes the peak times identified for all 10 subjects), so the excursion procedure was only applied at $M170$. Specifically, a peak-detection procedure was used to identify the smallest p-value across time within 100-300$msec$, and a small half-window of $20msec$ flanked around the peak was used as an excursion for temporal region of interest.

Subject	$M170$ peak ($msec$ after onset)
1	180
2	170
3	190
4	200
5	220
6	190
7	150
8	160
9	180
10	180

Table 4.1: **Identified peak instances for M170 across subjects.**

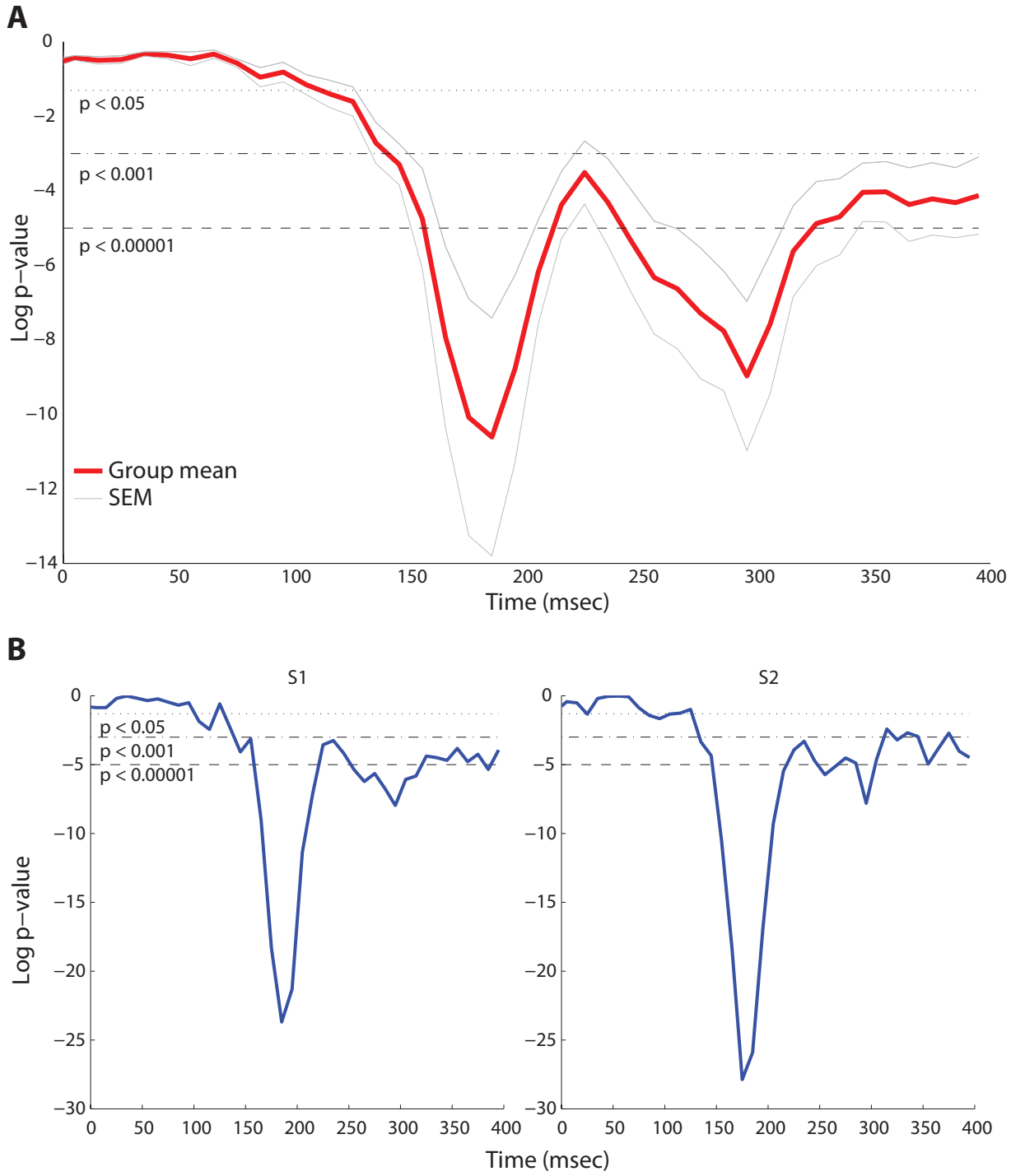

Figure 4.7: **MANOVA *p*-value time course.** **(A)** Group-level MANOVA *p*-value time course in contrasting face *vs.* object conditions. **(B)** *p*-value time course from two representative subjects illustrating peak discriminability at approximately 170*msec* after visual onset (0*msec*). The *p*-value is logarithmic with base 10.

Spatial ROI with MEG searchlight

To further register ROIs for faces, a MEG searchlight algorithm was used to find spatially nearby dipole clusters within the time window of interest defined in the previous section. Specifically, source activities were first estimated using the minimum-norm estimates (MNE) [1, 2] trial by trial. For each source dipole, the time course within the temporal window of interest was then averaged for each trial per subject. This step yields a single data point for a dipole in each trial.

To encourage spatially contiguous ROIs, a searchlight algorithm that groups nearby dipoles in cortex was used. In particular, for any dipole, discriminablity was determined by considering its 2 closest neighbors in space (group of 3). A multivariate t-test was then performed to test the null that grouped dipole response is equal under face *versus* object conditions. To focus on the core and extended face network as reported previously in [92, 94], the searchlight procedure was anatomically bounded in regions of fusiform, lateraloccipital, superiotemporal, inferior frontal and orbitofrontal gyri as defined using the Freesurfer software (http://surfer.nmr.mgh.harvard.edu/). Finally, a threshold of $p < 0.001$ was applied to retrieve clusters that show significant discriminability, and isolated small dipole groups were manually removed as possible. This procedure yields 11 total ROIs in the face network (summarized in Table 4.5, including 8 (4×2) in bi-hemispheric posterior-ventral-temporal cortex: the inferior occipital gyrus (IOG), the middle fusiform (mFus), the superiortemporal sulcus (STS), the anterior inferiortemporal gyrus (aIT), 2 in bi-hemispheric the inferior frontal gyrus, and the orbitofrontal cortex in the right hemisphere only.

Face ROI abbreviation	Cortical location
IOG	Inferior occipital gyrus (bi-hemispheric)
mFus	Middle fusiform (bi-hemispheric)
STS	Superiortemporal sulcus (bi-hemispheric)
aIT	Anterior inferortemporal gyrus (bi-hemispheric)
IFG	Inferior frontal gyrus (bi-hemispheric)
OFC	Orbitofrontal gyrus (right-hemisphere)

Table 4.2: **Identified regions of interest in the face network.**

4.6 Appendix II: Additional results

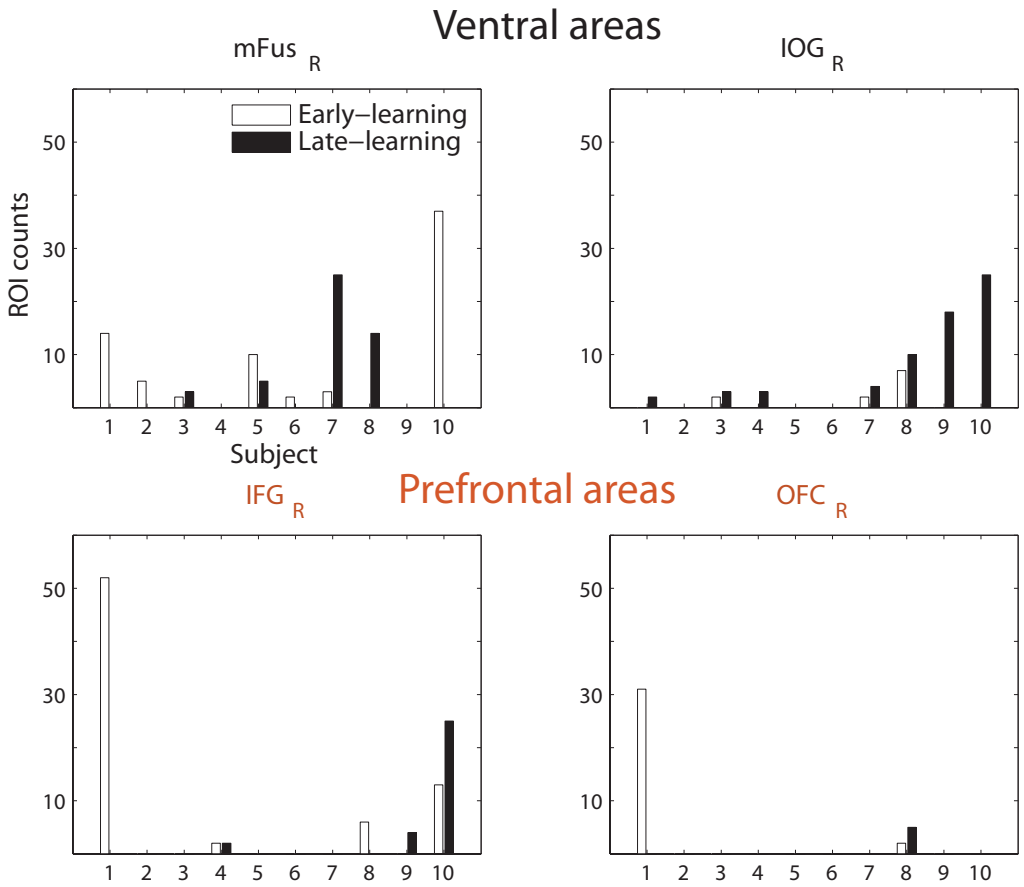

Figure 4.8: **Category-discriminative information in selected regions of the face network with excursion for individual subjects.** Each panel shows the raw tallies of hot spots in time course of a right-hemispheric region during early and late learning. 7 of 10 subjects exhibit an increase of hot spots in time course of the right inferior-occipital gyrus (IOG). Patterns are less consistent in the middle fusiform gyrus (mFus) and prefrontal cortex (e.g. inferior-frontal gyrus or IOG and orbitofrontal cortex or OFC).

Chapter 5

Method I: Characterizing spatiotemporal hot spots

Identifying regions with high differential response under multiple experimental conditions is a fundamental goal of brain imaging. In this chapter, I present a novel statistical method for characterizing spatiotemporal cortical regions of interest in MEG (and EEG). In many studies, regions of interest are not determined *a priori* but are instead discovered from the data, a process that requires care due to the great potential for false discovery. An additional challenge is that MEG sensor signals are very noisy, and brain source images are usually produced by averaging sensor signals across trials. As a consequence, for a given subject there is only one source data vector for each condition, making it impossible to apply testing methods such as ANOVA. We solve these problems in several steps. (1) To obtain within-condition uncertainty we apply the bootstrap across trials, producing many bootstrap source images. To discover hot spots in space and time that could become ROIs, (2) we find source locations where likelihood ratio statistics take unusually large values. We are not interested in isolated brain locations where a test statistic might happen to be large. Instead, (3) we apply a clustering algorithm to identify sources that are contiguous in space and time where the test statistic takes an "excursion" above some threshold. Having identified possible spatiotemporal ROIs, (4) we evaluate global statistical significance of ROIs using a permutation test. After these steps, we check performance via simulation, and then illustrate their application in a MEG study of 4-direction center-out wrist movement,

showing that this approach identifies statistically significant spatiotemporal ROIs in the motor and visual cortices of individual subjects. In general, our method is applicable to any MEG or EEG studies that require contiguous spatiotemporal ROIs to be defined under multiple conditions.

5.1 Spatiotemporal regions of interest in MEG

Magnetoencephalography and electroencephalography are non-invasive techniques that record effects of brain activity with millisecond precision. They provide brain activity images when signals recorded from sensors are mapped to activation levels on a grid of possible source locations in selected parts of the brain. There are typically many more source locations than sensor signals, which makes source localization an ill-posed inverse problem. Common methods of solving this inverse problem apply L1 and L2 constrained linear regression [2, 108]. MEG and EEG experiments typically involve two or more alternative experimental conditions, with the goal of identifying brain regions having strong differential activation levels. Such "hot spots" of differential activity are often called regions of interest (ROIs). Sometimes ROIs are determined *a priori* from substantive scientific hypotheses, but it is often desirable to identify them empirically. Because there are usually hundreds or thousands of possible source locations over hundreds or thousands of time points in MEG and EEG, there are substantial opportunities to find spurious ROIs due to chance alone. This chapter presents a new method of discovering ROIs in space and time simultaneously using an excursion algorithm via spatiotemporal clustering, and global significance tests (p-values) for discovered ROIs. The method may be applied to source images of individual subjects.

There are three main components of our work. First, because MEG and EEG signals are extremely noisy, source localization usually need averaging across many experimental trials. With only a single trial in each experimental condition, the usual methods of comparing responses across treatments (ANOVA or MANOVA) can not be applied in source space. To overcome this limitation we apply the bootstrap [126], repeatedly resampling from the trials in the sensor space, thereby generating multiple images that reflect variability of the source localization procedure. Second, complex brain activities during continuous recording yield high variability in brain modulations across space and time. To become candidate ROIs, hot spots of differential rates of activity should consist of many source locations that are

contiguous in space and time. We apply a likelihood ratio test to the bootstrapped images, computing the test statistic at every location in both space and time, and then threshold the test statistics and cluster the results in terms of both test statistic values and space-time coordinates. As we show, this identifies candidate regions of modulated activity. Finally, to overcome problems with multiple comparisons, we apply a permutation test, repeatedly permuting the resampled images across conditions while computing the likelihood ratio test and performing clustering for each set of permuted images. This provides a valid global p-value. Figure 5.1 summarizes the methodology in six steps. For the clustering step we apply Bayesian hierarchical clustering [127] because, from our previous experience, we have found it an effective method that automatically determines the number of clusters to examine.

In Section 5.2 we describe our procedures. In Section 5.3 we evaluate our algorithms in simulation studies and then illustrate the methodology in the context of a MEG study of hand movement, in which brain activities of human subjects were recorded while they performed a visually cued 4-direction center-out wrist movement task [128, 129]. In this case, we had 4 alternative experimental conditions corresponding to 4 movement directions and the goal was to identify regions on the cortical surface containing neural activity exhibiting differential modulations across the movement directions. Using the discovered ROI we show highly distinguishable patterns among the directions. In Section 5.4, we discuss further issues and draw conclusions.

5.2 Methods

Suppose we have MEG or EEG recordings from M sensors at T time points for R trials under C conditions. If we index time by $t = 1, ..., T$, trial by $r = 1, ..., R$, condition by $c = 1, ..., C$ and let $\mathbf{y}_{cr}$ be a $M \times T$ matrix representing recordings at trial r under condition c, then the ensemble recordings $\mathbf{y}$ can be represented as follows

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_{11} & \cdots & \mathbf{y}_{1R} \\ \mathbf{y}_{21} & \cdots & \mathbf{y}_{2R} \\ \vdots & \ddots & \vdots \\ \mathbf{y}_{C1} & \cdots & \mathbf{y}_{CR} \end{pmatrix}. \tag{5.1}$$

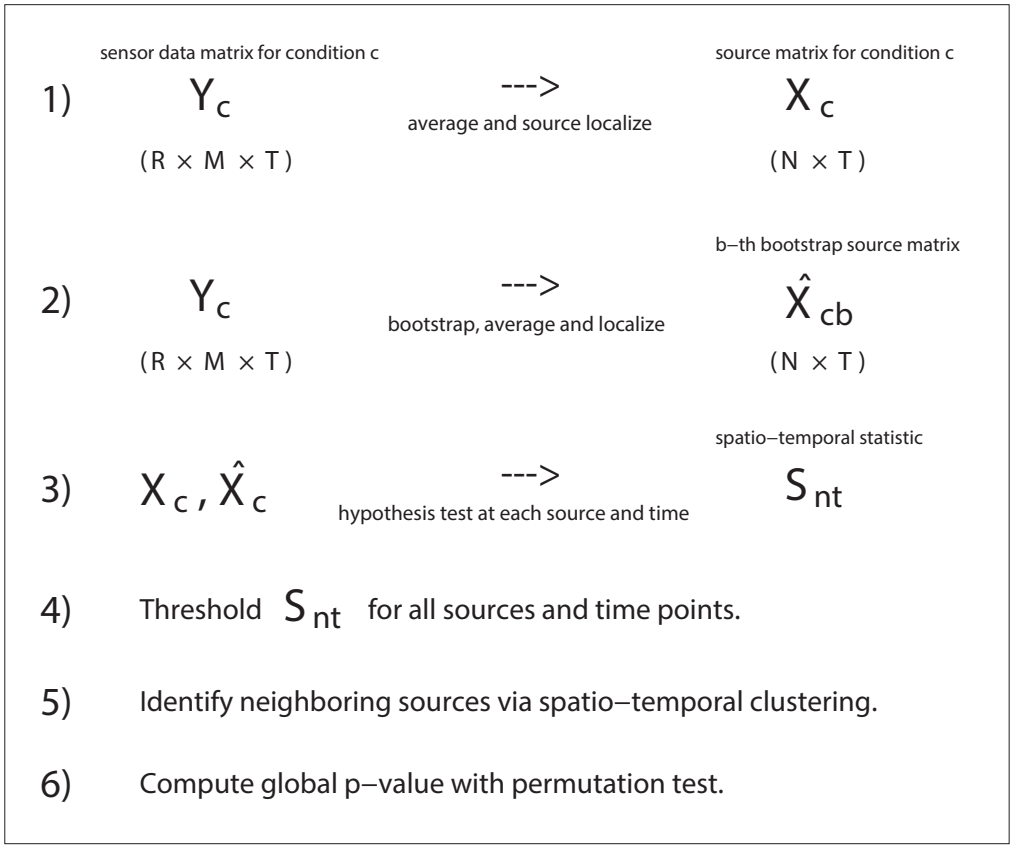

Figure 5.1: **Flow diagram of methodology.** The data for condition c consist of M sensor signals at each of T time points across R replications (trials). In step 1, we average the sensor signals across trials and then localize (see text) to N sources in the brain. Step 2 repeats this process for bootstrapped set of trials. In step 3, we perform hypothesis tests at every source and time point to test the null hypothesis that the mean source activities are equal across conditions. We threshold these test statistics in step 4 and then identify spatiotemporal neighbors as clusters, representing potential "hot spots" in step 5. We carry out a global significance test in step 6.

To map these sensor recordings to the cortical surface, MEG source localization algorithms such as Minimum Current Estimate (MCE) [108] typically average sensor signals across many trials, and then solve the inverse problem via constrained optimization

$$\mathbf{y}_c = \mathbf{A}\mathbf{x}_c + \mathbf{e} \tag{5.2}$$

where $\mathbf{y}_c = \frac{1}{R}\sum_{r=1}^{R} \mathbf{y}_{cr}$ (i.e. row averages of $\mathbf{y}$ in Equation 5.1) is the sensor trial mean under condition c, $\mathbf{A}$ is specified by the forward model from the quasi-static solution to Maxwell's equations [2], $\mathbf{x}_c$ ($N \times T$, assuming there are N sources indexed by n) is the transformed source currents on the cortical surface where each row in $\mathbf{x}_c$ is the time course of a source, and $\mathbf{e}$ is the additive noise. Since the number of sources far exceeds that of sensors, the inverse problem is ill-posed [2] (it is a $p \gg n$ problem). MCE offers a sparse solution to the inverse problem via weighted L1 regularization, effectively assuming that there are only a few sources that are activated at one time point (e.g. an image of the cortex with very few hot spots).

Solving the inverse problem based on averaged trials yields only a single trial under each experimental condition, hence it is difficult to carry out statistical analysis on these sparse data. To resolve this issue, we use bootstrap to generate multiple samples of source images under each condition. We utilize these bootstrap samples as our data and apply hypothesis tests in the brain source space. We then apply an excursion algorithm to find statistically significant and nearby spatiotemporal points ("hot spots")—it turns out that the excursion can be implemented in terms of a clustering algorithm. Finally, we use a permutation test to evaluate the significance of these regions of interest. In the following sections, we describe each of these procedures in detail.

5.2.1 Bootstrapped source images

To produce multiple "copies" of trials in the source space, we resample the trials under condition c by B times. For bootstrap sample b with $b = 1, ..., B$, we then let $\mathbf{y}_{crb}$ be the resulting sensor signal vector, take $\mathbf{y}_{cb} = \frac{1}{R}\sum_{r=1}^{R} \mathbf{y}_{crb}$, and write

$$\mathbf{y}_{cb} = \mathbf{A}\mathbf{x}_{cb} + \mathbf{e} \tag{5.3}$$

By bootstrapping we generate uncertainty under each experimental condition—each of these samples is a slight variant of the original source image based on trial averages. By solving the inverse problem (Equation 5.3), we obtain $\hat{\mathbf{x}}_{cb}$ as an estimate of $\mathbf{x}_{cs}$ and thereby obtain the ensemble of source signals (or time-varying source images)

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{\mathbf{x}}_{11} & \cdots & \hat{\mathbf{x}}_{1B} \\ \hat{\mathbf{x}}_{21} & \cdots & \hat{\mathbf{x}}_{2B} \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{x}}_{C1} & \cdots & \hat{\mathbf{x}}_{CB} \end{pmatrix}. \tag{5.4}$$

5.2.2 Likelihood ratio test

To examine whether any individual brain regions are modulated under differing experimental conditions, we perform hypothesis tests making use of the bootstrapped source data (Section 5.2.1). It should be emphasized that in reality, any standard testing methods such as ANOVA would be applicable, although here we choose to use a likelihood ratio test similar to that described in [130], which does not assume that the conditions have equal variance and explicitly takes that variability into account. We write the likelihood ratio test in extension to bootstrapped data following [130]. A likelihood ratio test has the form

$$LR = \frac{sup_{\theta \in \Theta_0} f(x|\theta)}{sup_{\theta \in \Theta} f(x|\theta)} \tag{5.5}$$

where $f(x|\theta)$ is the likelihood parameterized by θ, and Θ_0 and Θ are the parameter space under the null and in the unrestricted space respectively. Following Equation 5.4, we denote the signal from a source at a single time t from bootstrapped source trials in all conditions as $\mathbf{x} = [x_{11}, ..., x_{1B}, ..., x_{C1}, ..., x_{CB}]^T$, where x_{cb} is the signal strength in trial b under condition c. Assuming that the mean signal under condition c, $x_c = \frac{1}{B}\sum_b x_{cb}$, is normally distributed $x_c \sim \mathcal{N}(\mu_c, \sigma_c^2)$, and under the null hypothesis H_0 that the source current has equal mean strength under all conditions, we construct the following test

$$LR = \frac{\prod_c (2\pi\sigma_c^2)^{-\frac{1}{2}} \exp(\frac{-(x_c - {\mu_0}^*)^2}{2\sigma_c^2})}{\prod_c (2\pi\sigma_c^2)^{-\frac{1}{2}} \exp(\frac{-(x_c - {\mu_c}^*)^2}{2\sigma_c^2})} \tag{5.6}$$

where it can be easily shown that the following is the maximum likelihood estimate under H_0

$$\mu_0{}^* = \frac{\sum_c \frac{x_c}{\sigma_c^2}}{\sum_c \frac{1}{\sigma_c^2}} \tag{5.7}$$

where σ_c^2 is the variance of $\mathbf{x}$ estimated from the bootstrapped data. It also follows that $\mu_c{}^* = x_c$ is the maximum likelihood estimate in the unrestricted parameter space. Equation 5.6 can be thus simplified to

$$-2 \log LR = \sum_c (\frac{x_c - \mu_0{}^*}{\sigma_c})^2 \tag{5.8}$$

which if all σ_c were known, would follow a χ^2 with $C-1$ degrees of freedom. We perform this likelihood ratio test for each source at each time point and repeat across time. Assuming there are $T^{'}$ points after smoothing, we write the $\log LR$ test statistic for source n at time t as s_{nt} with $n = 1, ..., N$ and $t = 1, ..., T'$. In practice, we can smooth the signals by averaging them over a small window every few time points to reduce the number of tests. We then obtain a matrix of smoothed $\log LR$ statistics for every source across time. We then collect these into a matrix

$$\mathbf{S} = \begin{pmatrix} s_{11} & \cdots & s_{1T'} \\ s_{21} & \cdots & s_{2T'} \\ \vdots & \ddots & \vdots \\ s_{N1} & \cdots & s_{NT'} \end{pmatrix}. \tag{5.9}$$

We also obtain a corresponding matrix of p-values ($\mathbf{P}$), based on the χ^2 distribution.

5.2.3 Spatiotemporal excursion algorithm

After obtaining the $\log LR$ statistical map $\mathbf{S}$ (Equation 6.13) from the likelihood ratio tests, we wish to locate ROIs by identifying spatiotemporal clusters that have significant modulations. In this section, we describe a spatiotemporal excursion (STE) algorithm that achieves this.

The STE algorithm finds the peaks in the spatiotemporal map of the $\log LR$ statistics defined in Equation 6.13. Because peaks of interest occur across contiguous points in time and space, we cluster and threshold these spatiotemporal events. The STE follows three main steps (1) we prune "insignificant" cortical sources at a pre-defined α_{thresh} level (e.g. $\alpha_{thresh} = 0.01$) based on the likelihood ratio test statistics, (2) we partition the rest of the sources via a clustering algorithm where the inputs are four-dimensional vectors (3-D for spatial source coordinates and 1-D for a specific time instance), (3) we locate the hot spots by finding large aggregated source statistics within the obtained clusters. We illustrate our algorithm in a simplified 3-D example shown in Figure 5.2. The height on L-axis indicates the magnitude of the test statistic and the X and Y axes represent space and time. In step 1, we threshold using a sectioning plane at $L = 0.5$ and prune those points that are under the threshold. In step 2, we partition the remaining areas into contiguous clusters (i.e. the two peaks in Figure 2). In step 3, we localize ROIs by finding spatiotemporal clusters that have large statistics summed over the individual sources (i.e. the cluster with the larger area in this particular example). As a result, we obtain a spatiotemporal representation of ROIs where significant activities may take place. We summarize the STE algorithm as Algorithm 1. Specifically, the STE algorithm calls three subroutines which are fully described in Algorithms 2 – 4.

It turns out that the excursion procedure can be implemented in the vehicle of a clustering algorithm. Specifically, if we treat the spatial and temporal coordinates of individual sources as the input, then a clustering algorithm defined with some measure of similarity metric would serve the purpose of grouping these sources based on their spatiotemporal "closeness". To cluster the spatiotemporal sources, we use an algorithm (Algorithm 3) which partitions the sources based on their spatial and temporal coordinates. In our analysis, we use Bayesian hierarchical clustering [127]. It is worth mentioning that in principle, any clustering method (e.g. K-means, spectral-clustering) would work for excursion, and BHC is only one of these methods. The strength of BHC, however, is that it automatically determines the number of clusters in the data and has been proven to perform well in general. In the following section, we describe the BHC algorithm following [127].

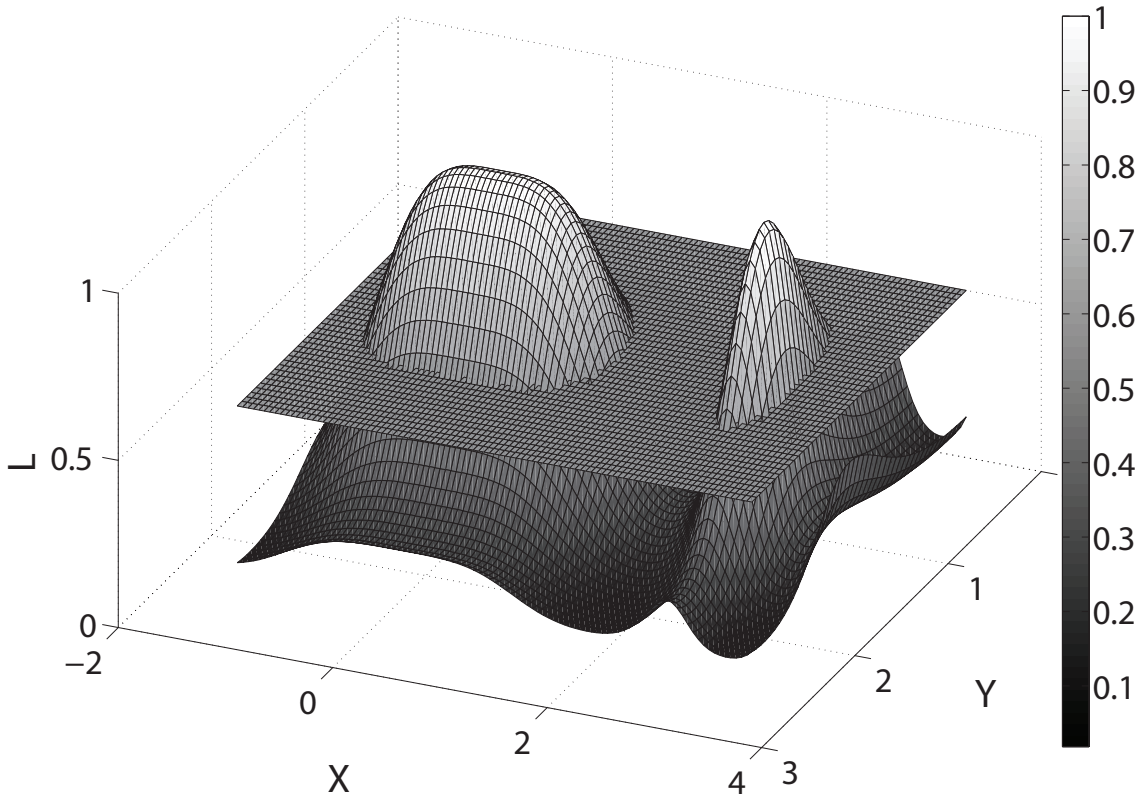

Figure 5.2: **Illustration of the excursion algorithm in a 2-D space.** The curved surface represents the magnitude of statistic from the hypothesis tests. The sectioning plane prunes insignificant sources at a pre-defined threshold level. We subsequently grouped the remaining two peaks into two distinct clusters based on their neighboring profiles.

Bayesian hierarchical clustering

Consider a data set $\mathcal{D} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}$ and tree T where $\mathcal{D}_i \subset \mathcal{D}$ is the set of data points at the leaves of the sub-tree T_i of T. In our context, each data point corresponds to a 4-element vector containing the spatial coordinates of a source and its time instance. BHC is similar to traditional agglomerative clustering in that it is a bottom-up agglomerative method which initializes n clusters (leaves of the hierarchy) each containing a single data point $\mathcal{D}_i = \{\mathbf{x}^{(i)}\}$. It then iteratively merges pairs of clusters to construct a hierarchical binary tree. The main difference between BHC and traditional hierarchical clustering methods is that BHC uses a statistical hypothesis test to choose which clusters to merge based on the odd ratio of posterior probabilities [131], instead of a distance metric.

In considering each merge, two hypotheses are compared. The first hypothesis ($\mathcal{H}_0^k$) is that all the data in $\mathcal{D}_k$ were generated independently and identically from the same probabilistic model (i.e. the merged hypothesis), $p(\mathbf{x}|\theta)$ with unknown parameters θ (e.g. a Gaussian with $\theta = (\mu, \Sigma)$). We compute the probability of data $\mathcal{D}_k$ under $\mathcal{H}_0^k$ by specifying a prior over the parameters of the model (if we use conjugate priors the following integral is tractable):

$$\begin{aligned} p(\mathcal{D}_k|\mathcal{H}_0^k) &= \int p(\mathcal{D}_k|\theta)p(\theta|\beta)d\theta \\ &= \int \Big[\prod_{\mathbf{x}^{(i)} \in \mathcal{D}_k} p(\mathbf{x}^{(i)}|\theta) \Big] p(\theta|\beta)d\theta \end{aligned} \tag{5.10}$$

The alternative hypothesis ($\mathcal{H}_1^k$) would be that $\mathcal{D}_k$ has two or more clusters in it (i.e. the split hypothesis). Summing over the exponentially many possible ways of dividing $\mathcal{D}_k$ into two or more clusters is intractable. However, if we consider only clusterings that partition the data consistent with the sub-trees T_i and T_j, which are built from the agglomerative bottom-up process, we can efficiently sum over exponentially many alternative clusterings using recursion. The probability of the data under the alternative hypothesis is then simply $p(\mathcal{D}_k|\mathcal{H}_1^k) = p(\mathcal{D}_i|T_i)p(\mathcal{D}_j|T_j)$. The marginal probability of the data in any sub-tree T_k is computed as follows:

$$p(\mathcal{D}_k|T_k) = \pi_k p(\mathcal{D}_k|\mathcal{H}_1^0) + (1 - \pi_k)p(\mathcal{D}_i|T_i)p(\mathcal{D}_j|T_j) \tag{5.11}$$

where $\pi_k \stackrel{\text{def}}{=} p(\mathcal{H}_1^k)$. Note that this equation is defined recursively, where the first term considers the hypothesis that there is a single cluster in $\mathcal{D}_k$ and the second term efficiently sums over all other clusterings in $\mathcal{D}_k$ which are consistent with the tree structure. At each iteration, BHC merges the two clusters that have the highest posterior probability of the merged hypothesis $r_k \stackrel{\text{def}}{=} p(\mathcal{H}_1^k|\mathcal{D}_k)$ which is defined by the Bayes rule:

$$r_k = \frac{\pi_k p(\mathcal{D}_k|\mathcal{H}_0^k)}{p(\mathcal{D}_k|T_k)} \tag{5.12}$$

The quantity π_k, which can also be computed bottom up as the tree is built, is defined to be the relative prior mass in a Dirichlet process mixture model (DPM) with hyperparameter α, of the partition where all data points are in one cluster, versus all the other partitions consistent with the subtrees. As shown in [127], $\pi_k = \frac{\alpha\Gamma(n_k)}{d_k}$ where $d_k = \alpha\Gamma(n_k) + d_{\text{left}_k}\, d_{\text{right}_k}$, right (left) refer to the children of internal node k, and at the leaves, $d_i = \alpha$, $\pi_i = 1$. BHC automatically infers the number of clusters by cutting the tree at $r_k < 0.5$. We summarize the BHC in Algorithm 3.

Although BHC is a greedy algorithm (i.e. it iteratively merges the two most probabilistically similar clusters), at each iteration it performs model selection by evaluating the odd ratio of the marginal likelihoods under the merged and the split hypotheses (i.e. Bayes factor of the two hypotheses). Since the algorithm is recursive, the marginal likelihood sums over configurations that are consistent with the tree structure, which takes into account a large space of partitioning of the data (the resulting hierarchical tree is a rich mixture model). The algorithm can be understood as performing iterative hypothesis tests that evaluate the odds of merging or splitting the pair of clusters as it builds up the tree, and the number of clusters is determined where the merged odds is lower than the split odd that cuts the tree. Work in [127, 132] relates BHC with Dirichlet process mixtures, a clustering model which considers the space of all possible partitions of the data yet is computationally intractable, by proving that BHC provides a theoretical lower bound on the marginal likelihood of DPM. That work also empirically demonstrates that BHC offers superior clustering performance and a tighter bound compared to other alternative approximate methods to DPM. It is worth mentioning, however, that since BHC outputs a single binary tree, it is possible that it may not capture fully the uncertainty associated with alternative clusterings. Work in [132] shows how BHC

can be modified to consider alternative clusterings, although the gain of incorporating uncertainty seems only marginal, suggesting that BHC generally yields good clustering results. We demonstrate the algorithm in Section 5.3.1. It is worth mentioning that the computational complexity of BHC is quadratic in the number of data points, hence the algorithm can be computationally demanding with very large data sets. In such cases, simpler algorithms such as K-means might serve as an alternative for practical purposes.

Computing the ROI statistic

Within each of the partitioned cortical area clustered by BHC, we compute the cluster statistics by summing over time and space and assign the clusters that have large statistics as the ROI. Suppose BHC partitions the sources into K clusters (including those pruned by thresholding, which have zero integral) and let $A = [A_1, ..., A_K]$, then ROI can be defined as

$$ROI \leftarrow \underset{k}{\operatorname{argmax}}\, int(A_k) = \int_{n \in A_k} \int_{t \in A_k} s_{nt} dn dt \tag{5.13}$$

where s_{nt} is the $\log LR$ statistic for source n at time t. Intuitively, this means that ROI is the region that contains contiguous sources that have the largest statistics summed over time and space. The algorithm is summarized in Algorithm 4. It should be noted that there are a variety of choices for ROI statistic, e.g. mean or median of the clustered sources, and the aggregated statistic defined in Equation 5.13 is only one possibility. In the case of single point sources, for example, using maximal or mean statistic within the clusters may be more appropriate than aggregated statistics. Our method, however, uses a modular approach that potentially allows different choices of statistic to be plugged into the algorithm for studying the ROIs.

Algorithm 1 Spatial-temporal excursion algorithm (*STE*)

input: threshold, averaging window, source signals and coordinates
run likelihood ratio test and threshold (Algorithm 2 – *Hyptest*)
cluster spatiotemporal sources (Algorithm 3 – *BHC*)
compute ROI statistics (Algorithm 4 – *ROIstats*)
output: ROI coordinates, time points and statistics

Algorithm 2 Routine *Hyptest*

input: signals $\mathbf{x}$ (N sources) and coordinates $\mathbf{POS}$ ($N \times 3$), threshold level α_{thresh}, averaging window δ

for $n = 1$ to N **do**

run likelihood ratio test on source n in time step δ (T steps)

compute statistics $\mathbf{s}_n = [s_{n1}, ..., s_{nT}]$ and corresponding p-values $\mathbf{p}_n = [p_{n1}, ..., p_{nT}]$

for $t = 1$ to T **do**

if $p_{nt} < \alpha_{thresh}$ **then**

obtain 3-D coordinates $\mathbf{POS}_n = [x, y, z]_n$

store $\mathbf{c}_{nt} = [t, \mathbf{POS}_n]$ in $\mathbf{C}$

store n and t in N^* and T^* respectively

end if

$t \leftarrow t + 1$

end for

$n \leftarrow n + 1$

end for

output: 4-D spatiotemporal coordinates $\mathbf{C} = \{\mathbf{c}_{nt} | n \in N^*, t \in T^*\}$

Algorithm 3 Routine *BHC*

input: spatiotemporal source coordinates $\mathbf{C} = \{\mathbf{c}_{nt} | n \in N^*, t \in T^*\}$

initialize: number of clusters $m = |N^*| \times |T^*|$ where each cluster contains a single 4-D coordinate $\mathbf{c}_{nt}$ in $\mathbf{C}$

while $m > 1$ **do**

Merge pair of coordinates with the highest probability of the merged hypothesis (see Section 5.2.3)

end while

output: clustered coordinates in space and time $A = [A_1, ..., A_K]$

Algorithm 4 Routine *ROIstats*

input: partitioned spatiotemporal source coordinates $A = [A_1, ..., A_K]$

for $k = 1$ to K **do**

compute cluster statistic $int(A_k) = \int_{n \in A_k} \int_{t \in A_k} s_{nt} dndt$

$k \leftarrow k + 1$

end for

ROI $\leftarrow \arg\max_{k*} int(A_k)$ and $S_{obs} = \arg\max_{k*} int(A_k)$

output: ROI topography $\{\mathbf{POS}_n | n \in A_{k^*}\}$, time points $\{t | t \in A_{k^*}\}$ and statistic S_{obs}

5.2.4 Computing global statistical significance

The likelihood ratio test described in Section 5.2.2 involves testing at a large number of source locations and time instances. To account for multiple comparisons, we use a permutation test which characterizes the statistical significance globally. Nichols and Holmes (2001) first introduced permutation test in ROI analysis in fMRI [133]. Maris and Oostenveld (2007) also applied similar methods to the time-frequency analysis of MEG and EEG data [134]. Here we apply the idea to spatiotemporal events in MEG and EEG source space.

In standard ANOVA permutation tests, we test the null hypothesis that the data distributions are the same across conditions. Here we consider the analogous null hypothesis that the multivariate source distributions across conditions are identical. In our context, once we find an ROI using the STE algorithm (Algorithm 1), we compute a p-value which quantifies the global significance by considering the sources in the ROI clusters as a whole. To do this, we use a permutation test where the null hypothesis is that the trials are identically distributed across conditions. In other words, we compute the ROI statistic (Equation 5.13) for each set of permuted trials, and compute the number of permutations where that statistic exceeds the observed ROI statistic calculated from the original data. Specifically, we reject the null at level $\alpha_{thresh} = \frac{J^*}{N_p}$ if in J^* out of N_p permutations, the ROI statistic found by permuting the trials across conditions is larger than that of the observed ROI. If we denote the statistic of observed ROI as S_{obs} and that of permuted trials j where $j = \{1, ..., J\}$ as S_j, we can approximate the global p-value

$$p = \frac{1}{J} \sum_{j=1}^{J} I(S_j \geq S_{obs}) \tag{5.14}$$

where each S_j can be calculated from the STE algorithm (Section 5.2.3). We summarize the complete procedures for computing the global p-value in Algorithm 5.

Algorithm 5 Global p-value of ROI

input: MEG/EEG sensor signals $\mathbf{y}$

bootstrap $\mathbf{y}$ under each condition
store source-localized bootstrapped signals in $\mathbf{x}$
run *STE* (Algorithm 1) on $\mathbf{x}$ and compute ROI statistic S_{obs}
for $j = 1$ to J **do**
  permute source signals across conditions and obtain $\mathbf{x}_j^{perm}$
  run *STE* on $\mathbf{x}_j^{perm}$ and tally $I(S_j \geq S_{obs})$
  $j \leftarrow j + 1$
end for
compute p-value $p = \frac{1}{J} \sum_j I(S_j \geq S_{obs})$

output: approximate global p-value

5.3 Results

In this section, we first evaluate our algorithms in three separate simulation studies. We then apply our method to a MEG study of center-out wrist movement where we discover statistically significant spatiotemporal ROIs in the motor and visual areas of cortex. Finally, we use the discovered ROIs to visualize the within and between condition variability based on the bootstrap.

5.3.1 Simulation

Clustering spatiotemporal events

We simulated a 3-dimensional space (2 spatial dimenions (x, y) and 1 temporal dimension t) with contiguous spatiotemporal "hot-spots" and used Bayesian hierarchical clustering to automatically group these events based on their coordinates. Figure 5.3A shows the true events in symbols of different shapes. The cluster of squares has 8 points that lie at $4:5$, $1:2$ and $10:11$ in x, y and t axes. The cluster of stars consists of 9 points and occurs at a single time point, lying at $6:8$ and $1:3$ in x and y axes. The cluster of triangles has 18 points that lie at $4:6$, $1:3$ and $1:2$ in x, y and t axes. Finally the cluster of circles contains 12 points and lie at $7:8$, $6:7$ and $1:3$ in x, y and t axes. Note that the triangle and square clusters of points are spatially overlapped (the square cluster constitutes a subset of the triangle cluster spatially) but occur at different time points. The triangle and the circle cluster overlap temporally but are discrete in space. The star cluster is relatively isolated on its own. Figure 5.3B shows the clustering results from BHC over these points. The algorithm identifies exactly the clusters despite the overlap in space and time. The negative weights on the dendrogram suggest that the ratio of the merged hypothesis against the split hypothesis is less than 1 and hence the tree could be cut off at these places, automatically yielding 4 distinct clusters.

Uniform p-values under null hypothesis

We used a simulation study to show that the p-values defined in Section 5.2.4 are roughly uniformly distributed under the null hypothesis for randomly generated data. We generated 1000 synthetic datasets in the MEG source space. For each source (853 in total in MCE software), we used 3 conditions each of which has 25 trials. Each trial is a flat 5-point time series of random amplitude between 1 and 100, where each point is subject to a Gaussian noise with $\mu = 0$ and $\sigma = 5$. In each of 1000 datasets, we applied the algorithm described in Section 5.2 to compute a global p-value by permuting 10000 times in the permutation test. The histogram of the resulting 1000 p-values roughly follow a uniform distribution.

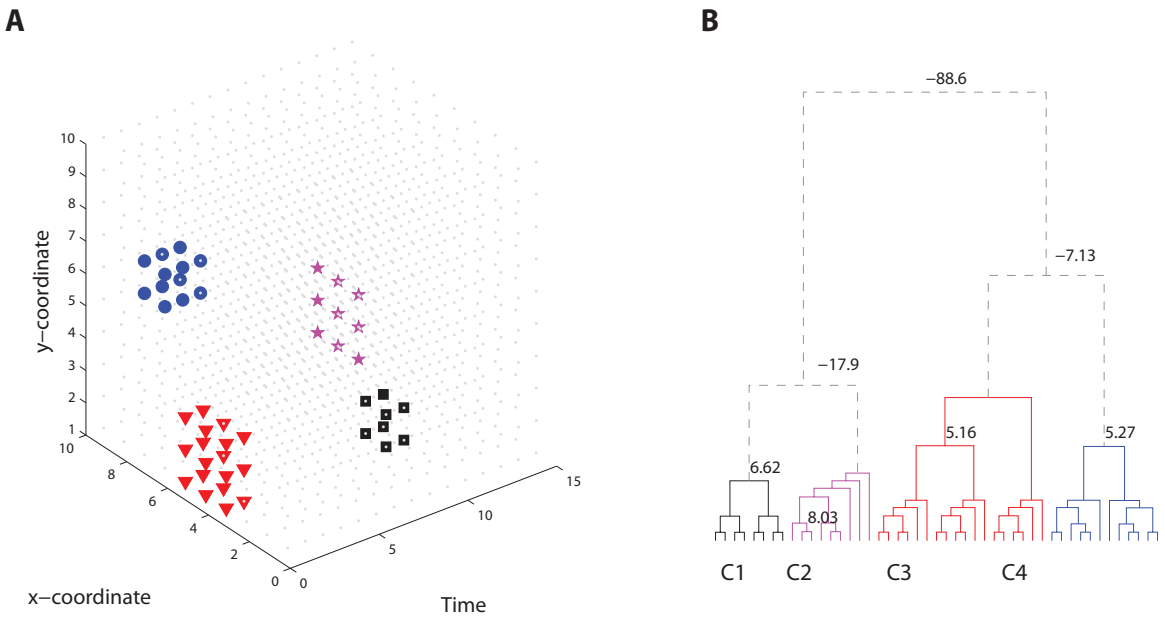

Figure 5.3: **Demonstration of spatiotemporal clustering in a 3-D space. (A)** We set up four "hot spots" in different shapes that extend through space and time (see text for details). **(B)** Clustering in the hierarchical tree. The algorithm prunes the tree where the odds ratio of the merged hypothesis falls below the split hypothesis, identifying four distinct clusters.

5.3.2 A MEG study

Experimental setup and data pre-processing

The experiment involved a center-out wrist movement task. A screen was set up in front of the subjects to provide visual feedback throughout the behavioral task. A 306-channel ElektaNeuromag MEG system was used to record their brain activities. Two right-handed subjects were asked to perform wrist movement by manipulating a joystick with their right hand following one out of four directions (radial and ulnar deviation, flexion and extension) indicated by a corresponding visual cursor cue (up, down, left, and right). Each subject was asked to perform the task in repeated trials (120 in each direction). Meanwhile, electrooculography was used to detect eye movement to remove any artifacts. All experimental procedures were approved by the Institutional Review Board at the University of Pittsburgh, and all experiments were performed in accordance with the approved protocol. The subjects gave informed consent before the experiments. Spatial filtering was performed on the raw MEG data using Signal Space Separation [135]. Trials with apparent eye movement detected by electrooculography were discarded. Each trial was aligned to recorded movement onset which was defined as the first time when 15% of maximal cursor speed was reached. Minimum current estimate (MCE) was used for source localization.

Experimental results

We applied our method to the center-out wrist movement study. For each subject, we bootstrapped the trial average of sensor signals within each of the 4 directions 50 times (i.e. 200 trials in total; to reduce computation, we used a small sample size, although by examining the samples we found the non-zero source currents follow the normality requirement of hypothesis tests) using the procedure described in Section 5.2.1. We then source localized each bootstrapped sample via MCE obtaining 200 images of source currents. For each source (853 in total) given by MCE, we ran the likelihood ratio test (the null hypothesis is that the mean source current is equivalent under 4 movement directions) through time by averaging the signal every $10msec$. The entire time course is $1000msec$, so there are 100 time steps after averaging. In each source and time step, we computed the $\log LR$ statistic and p-value forming a statistical map as in Equation 6.13. We note that the bootstrap variance across

conditions varied as much as 20-fold, hence the standard ANOVA assumption would not be valid in this case.

We applied the STE algorithm to the $\log LR$ statistical map thresholding at $p = 0.01$. Figure 5.4 shows the results for subject $S1$. Figure 4a illustrates the $\log LR$ map where the black dots indicate above-threshold test statistics. The onset of subject movement is at $0msec$ and the onset of visual cue is at approximately $-300msec$. We observe that significant modulations occur at about $100msec$ prior to movement and persist for $200msec$ through movement. These activities mostly occur in the motor cortex, and they correspond to the motor planning and execution stages. In addition to modulation in the motor cortex, we also observe significant modulations at about $-260msec$ in the occipital visual area after the cue onset, which are presumably due to processing of visual stimuli. Figure 5.4B shows the normalized sum of ROI statistics of 9 clusters discovered by STE. We see that cluster 9 (C9) has the largest ROI statistic whereas cluster 0 has zero ROI statistic corresponding to the under-threshold sources. Figure 5.4C maps these clusters topographically on the cortex, although it is worth mentioning that the sum of ROI statistics also integrates over time (Equation 5.13). We see that C9 with maximal ROI statistic is in the contralateral motor region, and large ROIS also occurs in the occipital and frontal areas. Figure 5.5 and 5.6 further demonstrate the results for the two subjects ($S1$ and $S2$) respectively. We note that both subjects have similar patterns of modulation with contralateral motor region owning the maximal ROIs and less significant modulations in the visual and frontal areas. These observations suggest that the motor cortex encodes differing directional information, which agrees with the phenomenon of directional tuning observed in single-neuron studies [136–138]. Our observation is also consistent with the results in a recent MEG study of decoding center-out movement via motor-related sensors [128].

To evaluate the global significance of the observed ROI (i.e. the contralateral motor clusters), we permuted the bootstrapped source trials 100000 times and used STE to compute the sum of ROI statistics for each permutation. For both subjects there was zero case where the permuted ROI statistic exceeds that of the original ROI. Hence we conclude that the ROI in both cases statistically significant with a global p-value $p < 10^{-5}$. To account for multiple ROIs, we repeated the entire permutation procedure and compared instead to the smallest ROI in the clusters partitioned from BHC, hence obtained a more conservative estimate for p-value. The intuition is that if the ROI with the smallest statistic is found

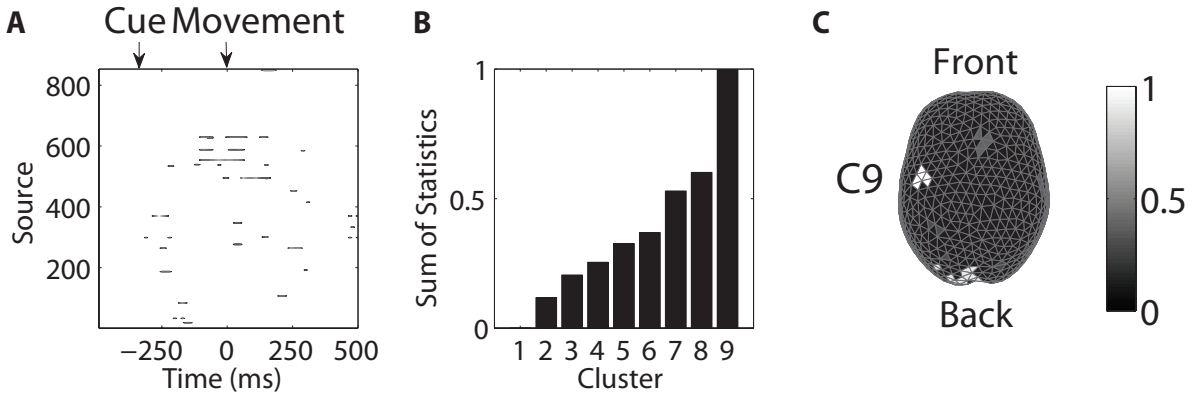

Figure 5.4: **Spatiotemporal excursion analysis in a visuomotor MEG study. A** Map of time-evolving chi-square statistics of 853 sources thresholded at $\alpha_{thresh} = 0.01$ for subject S1. The black dots indicate test statistics that exceed threshold. **B** Normalized sums of statistics of nine spatiotemporal clusters found using the STE algorithm for S1. Cluster 1 consists of under-threshold spacetime events. Cluster 9 has the maximal sum of statistics and corresponds to the contralateral motor area (on the right). **C** Spatial visualization of ROI on the cortical surface. The white area indicates the cluster with the maximal sum of statistics.

significant, then those that have larger statistics would also be significant. Thus we are able to establish the significance for multiple clusters. In these experiments, we also found that $p < 10^{-5}$. Finally, to visualize spatiotemporal modulations, we took 4 snapshots across the time course to see how the modulations varied spatially on the cortex (Figure 5.7). We observe that the modulations start from the occipital visual area during the cue onset and transit to the motor area before the movement onset and persist through the movement, which matches the observed $\log LR$ map in Figure 5.4A and intuitively explains the process in this visuomotor task.

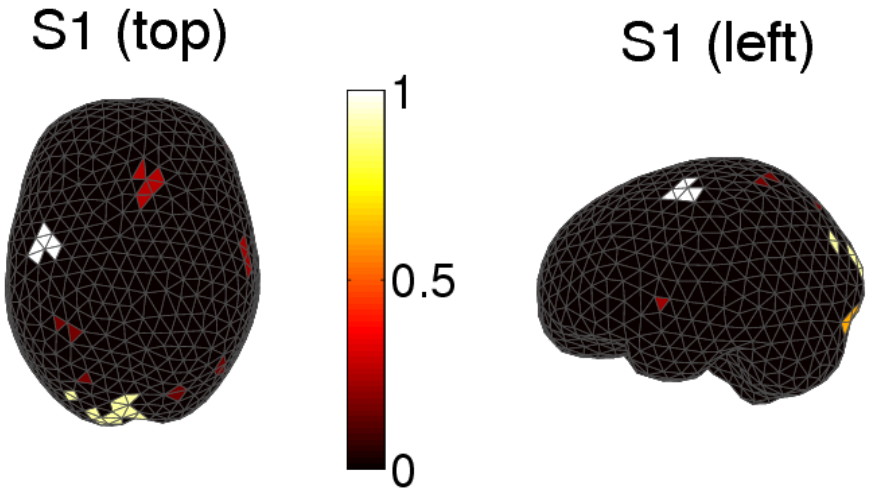

Figure 5.5: **Spatiotemporal "hot spots" for subject S1.** The intensity of the bar indicates the magnitude of the normalized sum of statistic.

5.3.3 Visualizing bootstrapped trial variability

We extracted the signals in the ROI in the contralateral motor area (Section 5.3.2) and examined the variation within and across the four movement directions. To do this, we projected the bootstrapped source signals at locations inside the ROI cluster discovered by our algorithm (these occurred 100 to $0msec$ prior to movement onset) onto a lower-dimensional space via PCA. Figure 5.8 visualizes these trials in a 2-D space spanned by the eigenvectors that correspond to the two largest eigenvalues in PCA. We observe that whereas

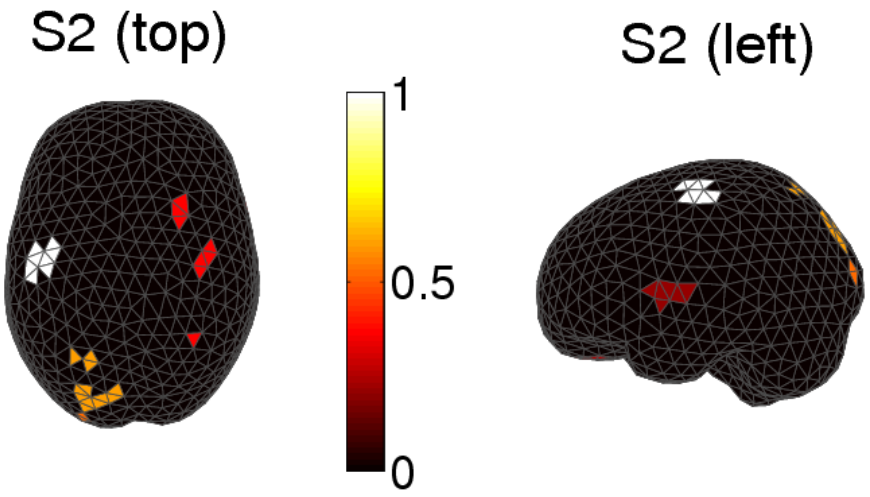

Figure 5.6: **Spatiotemporal "hot spots" for subject S2.** The intensity of the bar indicates the magnitude of the normalized sum of statistic.

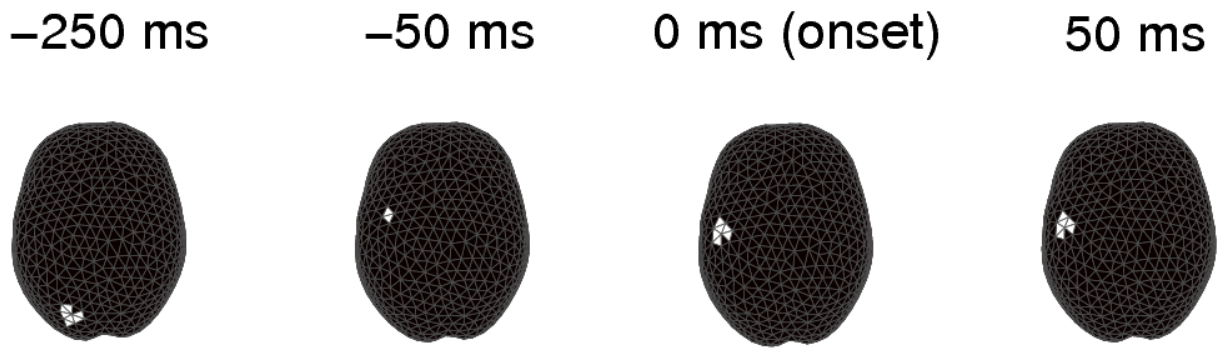

Figure 5.7: **Temporal snapshots of hot spots.** Cortical hot spots migrate from occipital visual region to motor region during the time course in a center-out visually cued motor task with movement onset at $0msec$.

the trials in each of four direction form their own clusters in the projected space, there is also noticeable trial-trial variation within each movement direction.

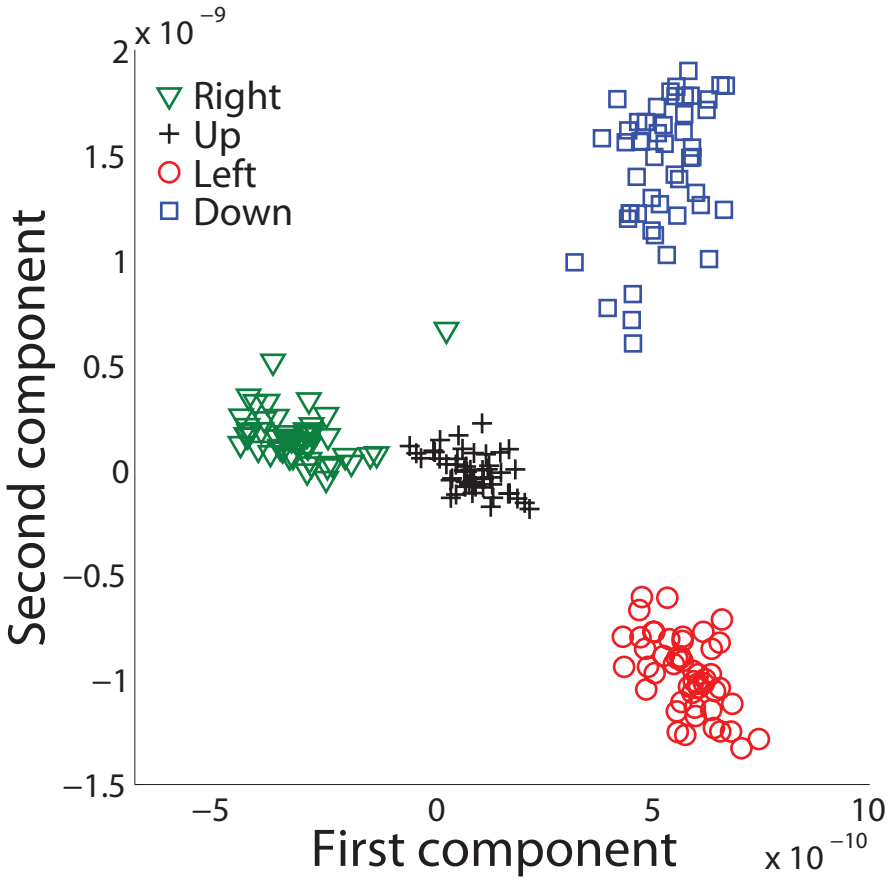

Figure 5.8: **Low-dimensional projection of bootstrapped trials.** Visualization of 200 bootstrapped trials in four movement directions (50 in each direction). The signals from the ROI in each trial are projected onto the first two principal components via PCA.

5.4 Discussion

To solve the problem of discovering brain regions having differential activity under varying experimental conditions, within subjects, from MEG or EEG data, we have pursued two ideas. The first uses bootstrap resampling of trials to produce uncertainty in source-localized images, within conditions. This is analogous to performing a somewhat unusual,

but valid bootstrap when solving an ordinary ANOVA problem. The usual bootstrap solution to ANOVA is to combine the data across conditions, and resample the whole, assigning at random each resampled observation to an experimental condition—each resampled set of data leading to a single bootstrapped F statistic. An alternative is to begin with the data means under each condition, compute a standard error of each mean by resampling the observations under each condition separately, and then apply a likelihood ratio test by assuming each mean to be normally distributed. This latter bootstrap would be unnecessary in the ordinary ANOVA problem because the same standard errors may be obtained analytically. In the case of MEG, however, we are not working with a sample mean but rather with the source-localized image strength at each source, and for this, particularly in the case of L1-penalized source localization, the bootstrap is very helpful.

The second idea was to apply a clustering method to thresholded likelihood ratio values in order to find contiguous spatiotemporal sources of high differential activity, and then to evaluate global significance using a permutation test. We used BHC as our clustering method, but this is not essential to the logic of our approach. We illustrated the methodology using data taken from subjects during wrist movement. Elsewhere we have shown that directionality can be decoded from individual trials based on MEG signals [129, 139]. The point, here, has been to discover regions responsible for this ability to decode, while producing a p-value that assesses the strength of the evidence against the rate of spurious null results. We have emphasized that the methodology applies to source images produced from individual subjects. Using this approach to examine inter-group differences by combining results across subjects is a topic for future research.

Chapter 6

Method II: Face network constrained source localization

The most fundamental challenge in MEG (or EEG) imaging is accurately reconstructing cortical source activities from a limited number of sensor recordings—a procedure known as *source localization*. Conventional approaches apply regularized regression to this problem, but most of these methods resort to a general solution without incorporating experimental or domain-specific knowledge. In this chapter, I present an alternative method that is designed for a class of problems where learning effects are pursued in cortical regions of interest defined *a priori*. In particular, we tailored the method to investigating spatiotemporal properties of the "face" network in a trial-and-error face category learning experiment (see Chapter 4). Our method features three steps that leverage both domain knowledge and the experimental structure. First, we registered regions of interest in the cortical face network using a MEG localizer experiment, and constrained the search for these ROIs by using a combination of excursion algorithms (see Chapter 5) and knowledge about the face network from the literature. Next, we partitioned the experimental trials into two parts for *model estimation* (training) and *hypothesis evaluation* (testing). Using trials in the midst of learning, we developed a spatiotemporal source model that weighs contributions of regions within and outside the face network differentially in light of a large number of trials—this ensures a good estimation of model parameters that accentuate the role of face network relative to irrelevant sources. As a final step, we applied the estimated model to trials in the initial and

final stages of learning—this provides a strong validation to the estimated model and allows hypotheses about learning to be tested. We showed that, overall, our proposed method yielded significant improvement in localizing source activities for held-out trials over an off-the-shelf source localization method.

6.1 An alternative approach to the inverse problem in MEG

Both magnetoencephalography and electroencephalography offer millisecond precision in recording cortical responses, but these imaging modalities also suffer from coarse reconstruction of source activities. Known as the inverse problem [2], recovering activities of thousands of cortical sources from a sparse array of sensors is highly under-constrained. Despite the ill-posed nature, cortical source activities are related to sensor signals by a simple forward model—a linear matrix operator derived from Maxwell's equations. In addition, the sensor noise in MEG is typically observed to be white noise [140], which is modeled sufficiently by a Gaussian distribution—parameters of the distribution can be estimated from empty room sensor recordings. Therefore, the crux of solving the inverse problem is making structured and reliable assumptions about cortical sources.

A popular approach to solving the inverse problem is using distributed source modeling. These models approximate joint activities of neuronal populations as dipoles, or polarized magnetic fields, which are typically distributed with small separations, e.g. $5 - 7mm$, and situated perpendicularly to the cortical surface [141]. Almost all distributed models use regularized regression. These models use either $L2$-regularized [1, 107] or $L1$-regularized [108, 109] regression to recover activities of source dipoles. Recent works have developed variants of these models by introducing penalties on dipoles and their time courses [142] or using a mixture of $L2$ and $L1$ norms to achieve smoothness in the frequency or time domain and sparsity in the source space [110, 143].

Despite the success and variety of distributed source modeling, relatively fewer works have explored the potential to design models that leverage experimental or domain-specific knowledge for a scientific investigation. To be precise, it is possible to manually select sensors of interest based on expert knowledge, but such methods are neither automatic nor neces-

sarily principled. Recent studies have proposed methods that incorporate domain-specific constraints in the modeling process. For example, Liu and colleagues [144, 145] have suggested that incorporating spatial constraints from regions of interest defined from functional imaging can improve localization in MEG. Henson et al. [125] have introduced an empirical Bayesian model that is shown to be effective in estimating restricted cortical sources for face perception. Our method extends beyond these approaches in several respects. Firstly, to provide more accurate spatial constraints in source space, we designed a localizer experiment in MEG as opposed to in functional imaging—this provides a more direct means to defining ROIs in MEG. Secondly, we developed a source model that allows parameters in a network of cortical regions to be automatically determined as opposed to manual thresholding; extending the model to a network of regions allows cortical dynamics to be explored at a broader scale. Finally, we partitioned the experimental data for model estimation (or training) and testing separately. Preparing a training set and an additional held-out set provides a much stronger evaluation for the source model and avoids the overuse or overfitting of data. To be more specific, we used the training data, a sufficiently large sample of trials, to estimate model parameters. We then applied the estimated model to the held-out data—trials excluded during training—for evaluating the accuracy of the model. Splitting trials this way allows our source model to be tested against a novel set of data. In addition, it makes it possible to source-localize data in the held-out set trial-by-trial and provide a procedure for quantifying standard errors in the estimation.

We applied our source localization method to MEG data obtained from a trial-and-error face category learning task in which subjects learned to discriminate two face categories. A learning paradigm serves as an ideal application because cortical learning effects are often pursued by contrasting data at the start and end of learning. Consequently, trials in the middle of learning are often under-used. The localization method we proposed utilizes these data, whereby trials in the middle of learning are leveraged for model training, and trials at both ends are used for model testing. Moreover, since recent works have studied extensively a distributed cortical network related to face perception [57, 92, 94, 98], this line of research provides valuable resources for constraining the ROIs in the source localization process.

6.2 Methods

Figure 6.1 describes the overall work flow of our method in the context of the face category learning experiment, but the proposed framework should be applicable to other (learning) paradigms if some portions of the experimental data can be used for model estimation. In the following section, we derive the cortical face network constrained source model. Chapter 4 (Appendix I) discusses the registration of regions of interest for face perception using an independent MEG localizer experiment (Step 1 in Figure 6.1), so we omit these details and assume the ROI source dipoles are defined *a priori*.

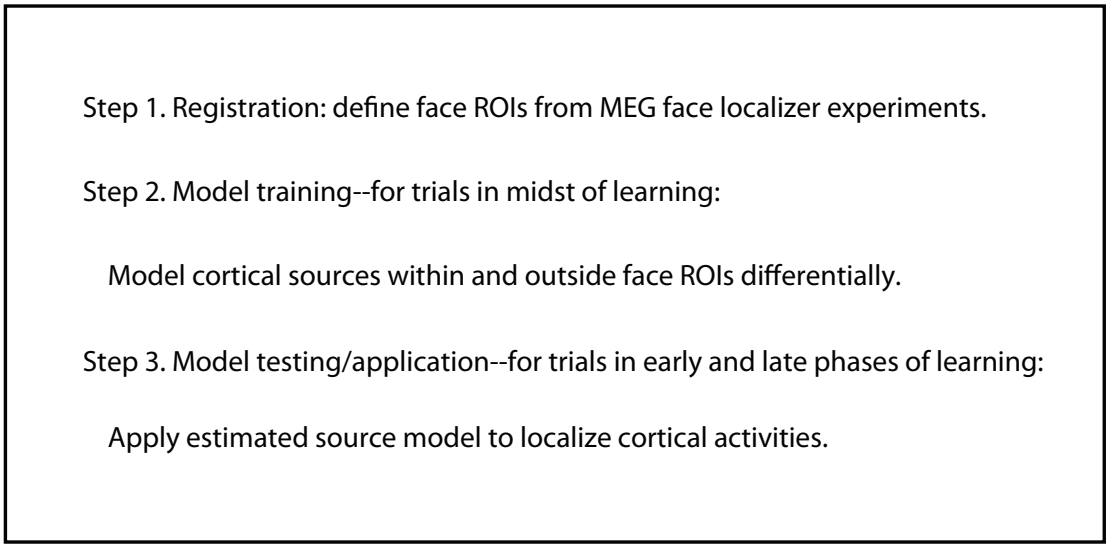

Figure 6.1: **Procedures for the proposed source localization method.**

6.2.1 Cortically constrained source model

This section describes the source model in full following the three procedures in Step 2 in Figure 6.1. We start with a formulation similar to a classical method called minimum-norm estimates, or MNE [1, 2]. Let $\mathbf{Y}$ $(M \times T)$ represent the ensemble of M observed sensor signals over time $t = 1, ..., T$, $\mathbf{X}$ $(N \times T)$ be the latent source currents from N dipoles, $\mathbf{A}$ $(M \times N)$ be a known constant forward matrix defined by Maxwell's equations, and $\mathbf{E}$ $(M \times T)$ be the sensor noise. Then the current estimates are linearly related to the sensor signals:

$$\mathbf{Y} = \mathbf{AX} + \mathbf{E}. \tag{6.1}$$

To simplify notations, we consider the same formulation at a snapshot, denoting $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{e}$ as a column vector in $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{E}$ at a single time instance, so it follows that

$$\mathbf{y} = \mathbf{Ax} + \mathbf{e}. \tag{6.2}$$

The MNE sets up a generative model where both the source and noise variables are assumed to be normally distributed:

$$\mathbf{x} \sim \mathcal{N}(0, \mathbf{R}) \tag{6.3}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{Ax}, \mathbf{C}). \tag{6.4}$$

Here $\mathbf{R}$ and $\mathbf{C}$ are covariances of cortical sources and sensors respectively. Note that $\mathbf{R}$ is assumed to be an identity matrix multiplied by a scaling factor (typically set proportional to the signal-to-noise ratio), and $\mathbf{C}$ can be directly estimated from empty-room sensor data. The MNE then computes the maximum *a posteriori* (MAP) solution to obtain current estimates for the sources:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} \left(P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \right). \tag{6.5}$$

It turns out that the generative model is equivalent to a $L2$ penalized regression:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left((\mathbf{y} - \mathbf{Ax})^T \mathbf{C}^{-1} (\mathbf{y} - \mathbf{Ax}) + ||\mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}||^2 \right), \tag{6.6}$$

where the first term to the right of the equation is effectively the negated likelihood function specified in Equation 6.5, and the second term is the negated prior distribution on sources.

Thus the MAP is equivalent to minimizing the $L2$ norm, both yielding an analytical inverse operator as the final solution:

$$\hat{\mathbf{x}} = \mathbf{R}\mathbf{A}^T(\mathbf{A}\mathbf{R}\mathbf{A}^T + \mathbf{C})^{-1}\mathbf{y}. \tag{6.7}$$

A major caveat of this approach is that all sources are treated equally and shrunken towards a zero mean. This assumption is unsatisfactory for cases where ROIs are known. In our model, we instead propose to group dipoles in each ROI within a pre-defined network and model them with different normal distributions. Suppose there are K available ROIs, we extend Equation 6.3 to

$$\begin{aligned} \mathbf{x}_0 &\sim \mathcal{N}(0, \sigma_0^2\mathbf{I}) \\ \mathbf{x}_1 &\sim \mathcal{N}(\mu_1, \mathbf{R}_1) \\ &... \\ \mathbf{x}_K &\sim \mathcal{N}(\mu_K, \mathbf{R}_K) \end{aligned} \tag{6.8}$$

where $\mathbf{x}_0$ corresponds to sources outside regions of interest, so they are modeled similarly as in MNE except their variances could differ (e.g. σ_0 can be potentially estimated for each individual source outside the ROIs). But differing from MNE, pre-defined ROI clusters $\mathbf{x}_1, ..., \mathbf{x}_K$ are modeled with their own means and covariances using Gaussian likelihood functions. This formulation allows differential activities outside and within ROIs to be captured flexibly in our model by shrinking irrelevant sources to zero and shrinking ROIs to values other than zero. In addition, covariances within the ROIs can also be modeled other than a diagonal matrix (e.g. a block-diagonal matrix).

Following the Bayes theorem and indexing trials by $i = 1, ..., R$, the posterior, assuming a unit prior on the model parameters, is:

$$
\begin{aligned}
P(\mathbf{x}|\mathbf{y}) &\propto P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \\
&= \prod_i p(\mathbf{y}^i|\mathbf{x}^i)p(\mathbf{x}_0^i)\prod_k p(\mathbf{x}_k^i) \\
&= \prod_i \mathcal{N}(\mathbf{y}^i|\mathbf{A}\mathbf{x}^i,\mathbf{C})\left(\mathcal{N}(\mathbf{x}_0^i|0,\sigma_0^2\mathbf{I})\prod_k \mathcal{N}(\mathbf{x}_k^i|\mu_k,\mathbf{R}_k)\right)
\end{aligned} \tag{6.9}
$$

We can take the logarithmic form of the posterior since it is a monotonous transform and remove all constant terms in the expression that are irrelevant to the parameter estimation:

$$
\begin{aligned}
L = \log P(\mathbf{x}|\mathbf{y}) \propto -[\sum_i (\mathbf{y}^i - \mathbf{A}\mathbf{x}^i)^T\mathbf{C}^{-1}(\mathbf{y}^i - \mathbf{A}\mathbf{x}^i) + \frac{\mathbf{x}_0^{iT}\mathbf{x}_0^i}{\sigma_0^2} + ... \\
... \sum_k \left((\mathbf{x}_k^i - \mu_k)^T\mathbf{R}_k^{-1}(\mathbf{x}_k^i - \mu_k) + log|\mathbf{R}_k|\right)]
\end{aligned} \tag{6.10}
$$

Since the ROIs are assumed to be independent of each other and with the rest of the sources, parameters of the model can be estimated using data from the middle part of learning in the following analytical forms:

$$
\begin{aligned}
\mu_k &= \frac{1}{R}\sum_i \mathbf{x}^i,\ \forall \mathbf{x} \in ROI_k \\
\mathbf{R}_k &= \frac{1}{R}\sum_i (\mathbf{x}^i - \bar{\mathbf{x}})(\mathbf{x}^i - \bar{\mathbf{x}})^T,\ \forall \mathbf{x} \in ROI_k \\
\sigma_{0(j)}^2 &= \frac{1}{R}\sum_i (\mathbf{x}_j^i - \bar{\mathbf{x}_j})^2,\ \forall \mathbf{x}_j \notin ROI
\end{aligned} \tag{6.11}
$$

These estimated parameters can then be used to source-localize trials in the initial and final part of learning. Note that if we set the means of ROIs to zero and use diagonal covariance matrices, this formulation is similar to the model proposed in [145] except the weights (or covariances) on ROIs are automatically inferred. However, if we allow parameters to be unconstrained for the ROIs, then the model offers much more flexibility by incorporating

both differential means for the ROIs and a block-diagonal matrix for the covariance of ROIs. Specifically, the mean (vector) of all source dipoles would be in the form:

$$\mathbf{m} = \begin{pmatrix} 0\,(\forall \mathbf{x} \notin ROI) \\ \mu_1\,(\forall \mathbf{x} \in ROI_1) \\ \vdots \\ \mu_K\,(\forall \mathbf{x} \in ROI_K) \end{pmatrix}, \tag{6.12}$$

and the covariance matrix of the sources would be a mixture of diagonal (for non-ROIs) and block-diagonal (for ROIs) forms:

$$\mathbf{R} = \begin{pmatrix} \sigma^2_{0(1)} & 0 & \cdots & & 0 \\ & \sigma^2_{0(2)} & & & \vdots \\ & & \ddots & & \\ \vdots & & & \sigma^2_{pp(k)} & \sigma^2_{pq(k)} \\ & & & \sigma^2_{qp(k)} & \sigma^2_{qq(k)} \\ 0 & & \cdots & & \ddots \end{pmatrix}, \tag{6.13}$$

where $\sigma^2_{0(1)}, \sigma^2_{0(2)}, ...$ represent variances of individual dipoles outside the ROIs (diagonal), and $\sigma^2_{pp(k)}$ and $\sigma^2_{pq(k)}$ represent variance-covariance for and between the pth and qth dipoles in cluster k (block-diagonal). Depending on the maximal number of sources within an ROI and the amount of available data, the quality of estimating the covariance of an ROI can vary. For example, for cases where the number of sources in an ROI clearly exceeds the number of trials, it would be better to avoid estimating a full-ranked covariance matrix. For cases where the number of sources in an ROI is substantially less than the number of trials, it is possible to more reliably estimate the covariance matrix.

6.3 Results

6.3.1 Simulation

To illustrate the effectiveness of spatial priors or ROI constraints in localization, we first conducted a simple simulation study. Here we created two artificial seeds in the source space located in the posterior fusiform and orbitofrontal gyrus. We then simulated source currents and projected these into the sensor space via a forward model taken from an arbitrary subject's data using the single-layer boundary element model. Figure 6.2A and B illustrate the sensor patterns and true source locations respectively. We estimated the noise component from empty-room sensor recordings and generated 100 normally distributed trials of source data in the designated seed regions, maintaining a signal-to-noise ratio of approximately 5db and a small variance of 20% off the mean source signal.

Using both MNE (without spatial constraints) and our proposed source model (ROI-constrained and weighted), we compared the recovered source pattern from each of these localization algorithms. Figure 6.2C shows that without spatial constraints, the orbitofrontal seed cannot be recovered at all. Figure 6.2D shows, however, that our ROI-weighting scheme helped recover the missing seed, suggesting it is advantageous to constrain on ROIs if these can be defined *a priori*.

6.3.2 Application to face category learning

A more realistic evaluation of the source localization method can be conducted on real-world data, and a more robust test requires some data to be held out for testing purposes exclusively. To achieve these goals, we applied our source localization method to the data acquired from a MEG face category learning experiment (see Chapter 4 for details). To provide a reliable estimate and properly evaluate the source model, we partitioned the experimental trials into training and testing portions. More specifically, we held out 160 trials (close to the length of the earliest and latest blocks in the experiment) at initial and final phases of learning for testing purposes, and used the rest of the trials in midst of learning (approximately 300+ trials for each subject) for model estimation, or training.

We applied the source model described in Section 6.2 to data in the middle portion of the

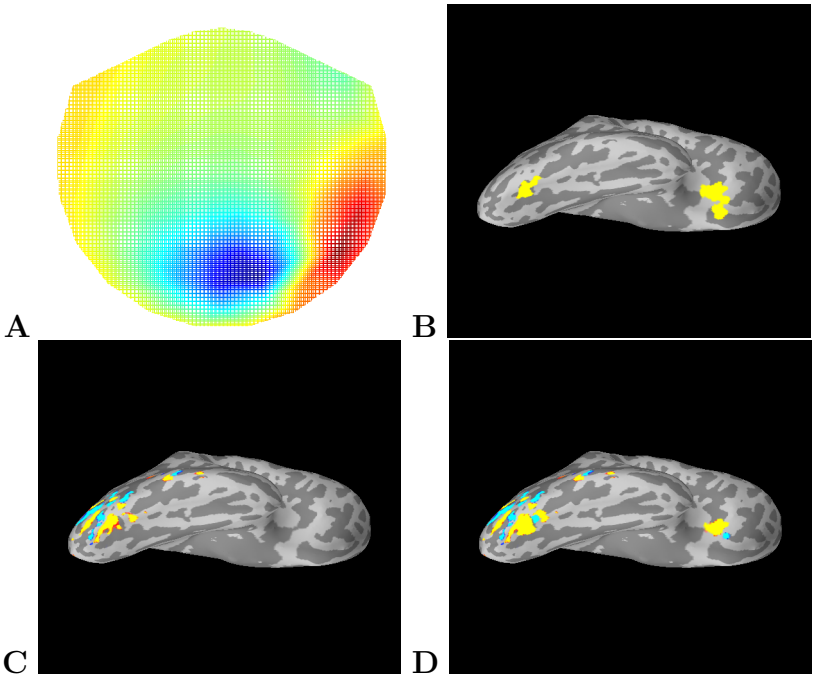

Figure 6.2: **A simulation that illustrates effectiveness of ROI weighting in localization.** **(A)** Simulated sensor signals. **(B)** True sources. **(C)** Source estimates without ROI constraints. **(D)** Source estimates incorporating ROI constraints.

face learning data set. This step allowed us to estimate parameters of the network of cortical ROIs and on rest of the sources. To evaluate the performance of our source model, we source-localized trials in early and late phases of learning using the estimated model parameters, and compared our method with the MNE method. Specifically, we used mean-squared error (MSE) as an objective criterion to evaluate how well our model performances did relative to MNE in terms of recovering sensor signals based on the estimated source currents. We computed MSE for both methods based on the reconstruction error on the sensor signals averaged over each of 320 (160×2) trials (trials indexed by i) and over a $600msec$ duration after the visual onset of the face stimulus (time indexed by t):

$$MSE(\hat{\mathbf{X}}) = \int_t E[\,||\mathbf{Y} - \mathbf{A}\hat{\mathbf{X}})||^2\,]dt \tag{6.14}$$

Figure 6.3 summarizes the performance at the group-level across 10 subjects. Regardless

of whether it is during early or late learning, our method reduces the MSE significantly over MNE in the held-out trials ($p < 0.022$ from binomial tests), which suggests that our method is highly effective and proves that there is an advantage to using a cortically constrained and spatiotemporal approach. Figure 6.4 further illustrates how MSE is reduced substantially by our method for almost all subjects in held-out data from different stages of learning. This provides strong evidence that using trials from middle-learning for model training can help to make source localization more precise in both early and late learning, which is desirable when examining learning effects at these end points.

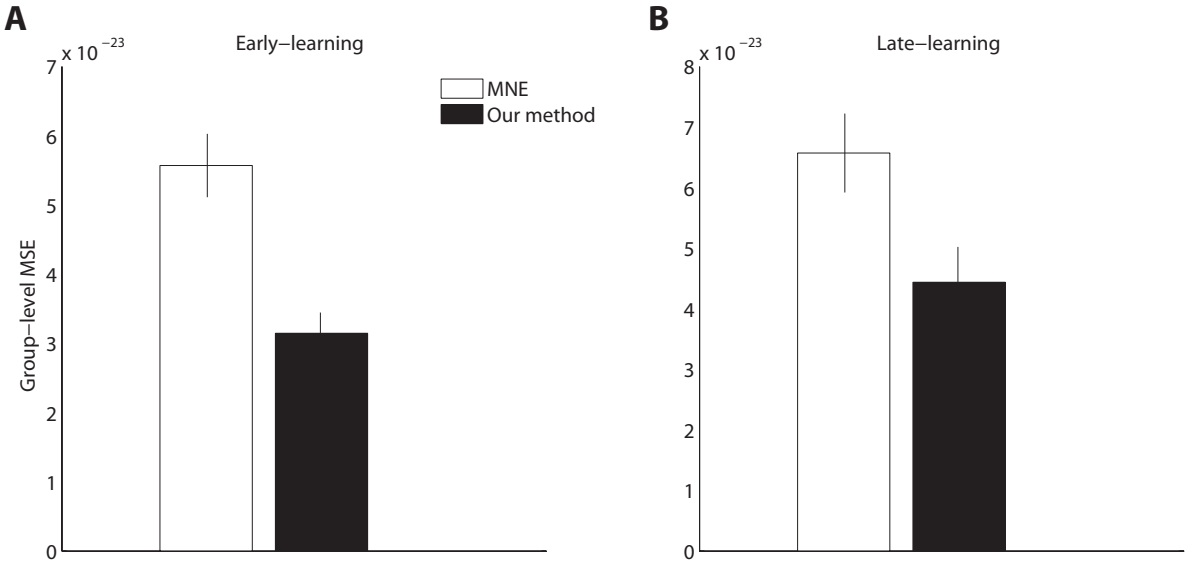

Figure 6.3: **Group-level mean-squared error comparison between minimum-norm estimates (MNE) and our source localization method. (A)** Comparison using held-out trials from earliest stage of learning. **(B)** Comparison using held-out trials from final stage of learning. In both early and late learning, our method outperforms MNE.

6.4 Discussion

We presented a novel source localization method for MEG that takes into account both domain-specific and experimental knowledge. Such a method offers significant advantage over an off-the-shelf method in localizing cortical source activities on held-out data. Using

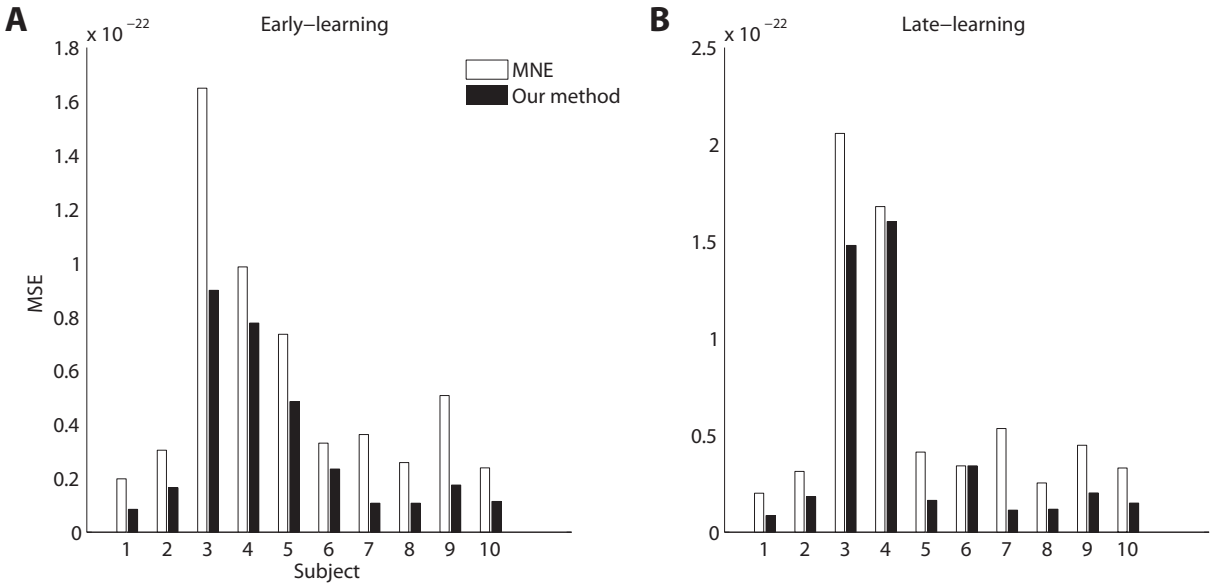

Figure 6.4: **Individual-level mean-squared error comparison between minimum-norm estimates (MNE) and our source localization method. (A)** Comparison using held-out trials from earliest stage of learning. **(B)** Comparison using held-out trials from final stage of learning.

domain-specific knowledge, our method allows a network of cortical regions, e.g. the face network in our application, to be differentially modeled and distinguished from task-irrelevant sources. Based on the unique structure of a learning experiment, our method uses trials in different phases of learning for model training and testing separately, effectively utilizing trials in midst of learning for robust estimation of model parameters, and increasing precision in localizing source activities from trials at both ends of learning. Importantly, we proposed the idea of trial partitioning so that some portion of data can be used exclusively for testing purposes. This promotes an alternative and much more convincing scheme for evaluating the quality of localization models.

A large body of research on the inverse problem has emphasized the development of generic source localization algorithms, e.g. [1, 107–110]. Models like these have been popularized because they are generally accessible, scalable and widely applicable to a range of problems. But as a consequence, they often ignore specificities in individualized experiments that could be vitally important for scientific inference. The method we have developed takes

an opposite approach that utilizes experimental data structure in the procedure of source localization. It also encourages a experimental design that helps constrain localization (e.g. the cortical face network) for the scientific domain in question. In doing so, our method offers the benefits to 1) reduce the number of sources that need to be estimated and thus alleviate the under-constrained inverse estimation problem and 2) offer an automatic procedure to differentially model the ROIs, accentuating those that are important to the data and down-weighting those that are less involved in the task. It is worth acknowledging, however, that it is not always possible to define regions of interest *a priori*, as in many cases ROIs can themselves be a subject of investigation. Under these circumstances, our method will find more limited applications, although we would still recommend performing some ROI-based analysis wherever possible, e.g. by using other imaging modalities. The hope is that ROIs defined elsewhere can eventually be fused to make localization more precise, e.g. [125, 145].

The learning paradigm we have introduced is unique in that it offers a natural way to partition the data, so we can afford to use different portions of the data for the purposes of model estimation and evaluation. But it is also general from two perspectives. Methodologically, it suggests that any localization methods should in principle be tested against held-out data. The notion of data partitioning or splitting is similar to cross validation commonly used in machine learning and statistics [146], but it is critically different because no data is ever used for more than one purpose (or doubly used). Thus from the perspective of model evaluation, it is far more strict and desirable than cross validation. Besides, it also offers the possibility of quantifying standard errors of the estimates. Scientifically, the procedure we have proposed also suggests that for any scientific conclusions to be drawn from the MEG source space, it may be better to prepare separate sets of data, so that localizing in one set can potentially improve the precision obtained from the other set, hence improving the accuracy of findings in MEG source space.

Chapter 7

Conclusions

Visual category learning is key to fast and accurate discrimination of novel visual stimuli. In this thesis, I used MEG to develop methodologies that captured cortical time course during online learning of visual categories, and showed that visual category learning is facilitated by spatiotemporal plasticity in the human cortex. Specifically, I focused on characterizing changes in information flow in the ventral stream and prefrontal cortex, which are both essential to visual category learning. With this proposed framework, I helped to bridge the gap between the human cortex and the non-human primate brain in regard to theories and methodologies in cortical spatiotemporal properties during visual category learning.

From a theoretical perspective, the current work extends previous theories by elucidating differential roles of the ventral stream and prefrontal cortex. Using objects and faces, I determined that information about visual categories is increasingly supplied in the ventral stream, but it becomes less prominent over time in prefrontal cortex. These findings suggest that the ventral stream supports encoding of visual categories in the long term, but prefrontal cortex jointly encodes categories mostly during initial learning.

From a methodological perspective, I developed novel statistical methods to tackle two key issues in MEG: high dimensionality and coarse source localization. In both methods, I showed how to more precisely characterize spatiotemporal properties of the cortex by incorporating constraints based on domain knowledge or experimental structures.

In this final chapter, I explain the implications of my theoretical findings, explain what we can learn from statistical modeling and discuss the limitations and extensions of my thesis.

7.1 The ventral stream is crucial to visual categorization in the long term

Neural substrates in the ventral stream have been proposed by researchers to be the dominant sites for representation of visual categories. This is supported by studies that show lesions in the ventral stream cause deficits in visual recognition (e.g. [30–32]), functional imaging studies that detect category-selective responses in this region (e.g. [33–38]), as well as EEG and MEG studies that source category-selective waveforms in occipito-ventral regions (e.g. [40, 41]). In this thesis, I addressed a question that has not been answered as clearly: does the ventral stream support the acquisition and online discrimination of visual categories? Critically, I investigated the cortical time course during visual category learning at a fine temporal resolution—such studies are scarcer than work on non-human primate brains [21, 22].

Scientists have debated whether or not the ventral stream is able to acquire discriminability of visual categories that differ little in feature space [19, 20]. Part of the contention on this issue stems from discrepancies in designing categorical visual stimuli [44]. In Chapter 3 and 4, I suggested that the design of visual categories should yield clear category boundaries, such that there would be no confusion in the perceptual space. This is not to imply, however, that all real-world visual categories are separated by clean boundaries, but rather those cases are exceptionally rare. From a behavioral point of view, category separability is a basic criterion for successful learning. Using visually similar yet separable categories of objects and faces, I demonstrated that categorical information is available in the ventral stream before category decisions are made. This information flows in certain regions, including inferior-lateral-occipital complexes (for objects), fusiforms (for objects and faces) and anterior inferior-temporal cortices (for faces).

The current findings do not rule out the possibility that category encoding involves other parts of the cortex. But, I did show category discriminability in the ventral stream increases over the course of learning, which suggests that spatiotemporal plasticity in the ventral stream is crucially related to proficiency in visual categorization. Thus, the results of my research agree with those of developmental studies, which find in-depth neural modulations in the ventral stream over time (e.g. [147, 148]). For example, Scherf et al. [148] reported

that whereas children do not exhibit category-selectivity for faces in the fusiform gyrus, adolescents and adults do, indicating a correlation between plasticity in the ventral stream and proficiency in visual recognition. I further suggested with my thesis that plasticity can occur during a short period of time—i.e. during online learning. Overall, these findings support the theory that the ventral stream plays a crucial role in determining efficient visual categorization over time.

7.2 Prefrontal cortex encodes categories during early learning

The results of recent works on primate brains have suggested that prefrontal cortex dominates the encoding of visually similar categories [21, 22]. However, little work has explored fine-temporal properties of the human prefrontal cortex. In this thesis, I showed that categorical information flows in prefrontal cortex during visual discrimination. Contrary to the previously proposed PFC-dominance theory [20–22, 45], I suggested that PFC involvement occurs early on during learning, and thus plays a more complimentary role over time.

Across object and face categories, I found that discriminability of visual categories decreases in PFC, but increases in the ventral stream. This finding is incompatible with the PFC-dominant view, which proposes that the PFC dictates category representation (especially at decision boundaries).

However, I did find evidence that PFC is better at encoding categories early on during learning, possibly because of its initial role in helping the ventral stream to map the stimulus to categories (i.e. working memory and rule formation). This finding coincides better with a complementary-PFC view [24, 44, 50–52], which posits that the PFC coordinates with the ventral stream to form categories. I proposed, though, that less help is required from PFC as categorization performance improves. This view is in agreement with a recent work [149] that showed repeated exposures to stimulus reduce neural responses in prefrontal cortex but enhance communications between frontal and temporal regions. However, it is possible that category coding in prefrontal cortex becomes more sparse as learning progressed, and the current work is limited by the spatial resolution of MEG. Future work is needed to confirm

whether visual categorization can be supported by the ventral stream along in the long term with minimal support from PFC.

In regards to the type of category, I found that PFC was more involved in the encoding of faces than objects. This finding could be partially explained in terms of differences in the stimuli used in these experiments. In Chapter 3, I used blob-like stimuli. The category boundary of these blobs depends on multiple features (around the edges). Therefore there are no simple rules to leverage in categorizing these blobs, and categorization likely involves an abstraction of prototypes. This is in contrast to the design of the face categories in Chapter 4, which differed only in two components such as the eyes and nose. Consequently, the categorization strategy for faces is more likely to involve attention to diagnostic facial parts than whole faces. It is plausible that such differences have involved PFC to differential degrees in encoding blob or face categories. Specifically, blob categorization relies on a prototype representation, but face categorization requires a more rule-based approach that potentially favors PFC [46–48].

Differences like these have been discussed in the literature. For example, Sloutsky and others [8] pointed out that depending on the nature of the stimulus, different cortical (and subcortical) systems can be recruited for category learning. More specifically, the blob stimuli are close to what they called dense categories, where categorization relies on multiple features. Face categories, though, correspond more with sparse categories, which rely on selection-based strategies. Consequently, the recruitment of cortical processes differs in the type of stimulus and categorization strategy. For example, a selection-based strategy might recruit PFC more for inhibition control and attention. Nevertheless, the ventral stream must support selection-based categorization, e.g. IOG involvement in coding facial parts. This complementary relationship between the ventral stream and prefrontal cortex is supported by evidence from [102], in which researchers found co-activation of the fusiform and prefrontal cortex in discriminating face categories that differ only in parts.

In summary, PFC may play less of a role in categorization of dense visual categories that require representation that is prototype-based. It may, however, contribute more to rule formation for categorizing sparse categories such as the face categories. But overall, I showed that the involvement of the PFC in category encoding decreases during learning, and this is comparably more in line with a complementary view of PFC in long-term visual categorization.

7.3 Faces help elucidate cortical spatiotemporal properties

Beyond the comparisons between the ventral stream and prefrontal cortex, of further concern is how visual categories emerge over time, and particularly in the ventral stream. To answer this question, I used faces as a model category to probe—with better spatial resolution—a cortical network during the learning of face categories, and suggested that, depending on strategies in visual categorization, such a hierarchical scheme is amenable to change.

Much evidence has suggested that recognition of faces involves a distributed cortical network (see [55, 92]). Although recent works have started to unveil the functions of this network (e.g. [57–59, 61]), we have not yet specified how information about face identities flows during the discrimination time course. In Chapter 4, by zooming into core regions in the ventral stream, I showed that face identities emerge in time course following the ventral hierarchy. This is instantiated in a hierarchical temporal coding pattern among aIT, mFus and IOG in the right ventral pathway.

This finding supports a recent work by Avidan and colleagues [58], which suggests that the disrupted connection between aIT and the posterior face network is key to deficits in face recognition. Here I showed in a more rapid time course that identity information is delivered hierarchically along the ventral stream (initially in learning), and this information is most prominent in aIT and least prominent in IOG. This implies a feed-forward mechanism in the ventral stream as proposed in computational models [28]. Thus, disruption of the connection between aIT and the rest of the core network would prevent the transmission of low-level information to the upper ventral stream. The current finding, however, cannot determine the direction of this information flow, which would help elucidate whether aIT provides top-down supervision on the core posterior network, or whether the feed-forward mechanism dominates (or connections are more recurrent).

In contrast to the hierarchical coding in initial learning, I showed that the posterior face network, e.g. IOG, encodes more information about face categories as learning progresses. Additionally, activities in IOG and mFus become more synchronous possibly due to increased local communication. These findings support previous proposals about the role of the IOG in processing facial parts [59, 60], but also suggest that IOG and mFus jointly encode face

categories. It is plausible that mFus selects informative facial features [114] based on information processed in IOG at an entry level, and the increased face discriminability in IOG is a result of top-down influence from mFus.

In summary, the current set of results confirms a hierarchical processing scheme in the ventral stream with a division of labor among the core "face" areas. It suggests that a shift in categorization strategy, e.g. increasingly part-based categorization, can perturb this hierarchy. Overall, the extensively studied model category of faces provides resourceful venues for investigating dynamic properties of the cortex at a better spatiotemporal resolution.

7.4 Lessons from statistical modeling

We face two major challenges in using MEG as an emerging imaging technology. First, a high temporal resolution and a large spatial coverage in MEG result in high-dimensional data. Validating discovered regions of interest across time and space thus becomes essential but crucially, few generic algorithms enforce spatiotemporal contiguity constraints in the ROI discovery process. Second, a coarse reconstruction of cortical activities in source localization can hinder analysis at a fine spatial scale, yet most state-of-the-art source models do not build in knowledge about the scientific domain of interest (or about the experiment). In this thesis, I developed two methods that tackled these issues by incorporating spatiotemporal constraints into the modeling process. This yielded better precision and helped scientific hypothesis testing.

Apart from building knowledge into statistical models, I also proposed 1) to design experiments to help constrain the models and 2) to potentially incorporate experimental knowledge into the modeling process. In particular, to provide spatial constraints for the face learning experiment in Chapter 4, I designed a separate localizer experiment that uses a contrast between face and object conditions to locate a cortical "face" perception network. These constraints allowed the source model to apply differential levels of shrinkages to regions of interest and those that are less relevant to face processing. This in turn reduced the diffuseness in the source solution and gave a better reconstruction of activities in the source space. By partitioning data into different stages of learning, I used data in the middle of learning—a substantial proportion of trials—for training the source model. To verify its effectiveness, I

used the rest of the data (at the beginning and the end of learning) to test the model. This procedure provides a strong test for the proposed model and, at the same time, improves precision in localizing trials at both ends of learning.

7.5 Limitations and extensions

In this closing section, I discuss the main limitations and some possible extensions of my thesis work.

7.5.1 Rapid learning in the cortex

The basic assumption in the experiments described in Chapter 3 and 4 is that the learning effect is minimal within the initial and ending stages as compared to the effect between these two stages. I made this assumption primarily because visual learning in the cortex is relatively slow, especially compared to subcortical learning (see [23]). However, another technical reason is that, due to a low signal-noise ratio, we cannot easily test learning effects in a handful of trials in MEG. In reality, visual category learning can be very fast [150], so the current work does not account for potential changes in the cortex during a rapid period of time. We can remedy this to some degree by using visually-similar categories because they are harder to learn and, compared to visually distinct categories, less susceptible to fast learning. However, it is still possible that rapid learning may occur.

Future work could explore the possibility of analyzing cortical changes concomitant with fast learning. For example, it would be valuable to investigate cortical plasticity associated with the steepest part of visual category learning. Specifically, we do not know how the ventral stream and prefrontal cortex would contribute to rapid learning, so one might help answer how fast category mapping is achieved by tracking the information flow in these regions and possibly other parts of the cortex. The key question, however, is whether the noise level (in MEG) allows investigation at this resolution.

7.5.2 Learning invariance in the cortex

Ultimately, theories of visual categorization should explain how the cortex learns invariant properties of objects under varying conditions, e.g. in view points, lighting, angle, size, position, orientation, occlusions and motion [5, 53]. The experiments described in this thesis used stimuli that have fixed orientations, viewpoints and no occlusions. Therefore the conclusions made were based on a restricted form of categorical abstraction.

Future work can extend this paradigm to incorporate stimuli that have varying degrees of viewpoints, orientations or partially occluded objects. These variants can be used to probe cortical spatiotemporal dynamics that become robust against varying conditions over the course of learning. In particular, if the ventral steam supports long-term visual categorization, I predict that over time categorical structure should emerge regardless of the varying conditions. I also expect, however, within a category, objects presented in different orientations, viewpoints or other conditions should not be differentiated in (the high-order areas of) the ventral stream. The remaining question is: how much does such invariance emerge in the time course of learning. Similar analysis can be conducted in prefrontal cortex and other parts of the cortex. In particular, if PFC is involved in encoding categories initially, to what extent does it support the encoding of invariance properties? It seems plausible that whereas the ventral stream may be susceptible to varying conditions such as viewpoints early in learning, prefrontal cortex may be affected to a lesser degree if it somewhat hard codes exemplars to categories (i.e. withholding them in working memory). In the long run, though, we have not determined whether PFC still plays a role in mapping exemplars to categories or the ventral stream alone is sufficient enough to account for the invariance in visual categorization.

7.5.3 Defining the null in MEG imaging

A fundamental limitation of analysis in MEG imaging is that it is difficult to define a null distribution across conditions. As such, results cannot necessarily be easily interpreted, particularly in the context of learning. As an example, it is difficult from two perspectives to confidently posit that prefrontal cortex ceases to encode visual categories in the long run. This is challenging from two perspectives. Firstly, the discriminability measure I have used to

characterize category separability is based on a ratio of the between-condition distance and the covariance pooled across conditions. It is plausible that recorded neural signals in PFC are noisier than signals recorded in the ventral stream, and this difference can potentially account for an apparent null effect in PFC. Secondly, even if the noise level were identical throughout the cortex, it is still hard to interpret how statistical discriminability reflects the actual discriminability. This is because it is almost impossible to define what parts of the cortex do not encode categories; identifying such regions helps provide a baseline measure for null category codability.

Perhaps a partial solution to resolve this issue is to manipulate experimental designs. For example, to define a null effect, one can design a pre-learning session where no learning behavior is expected or instructed, e.g. a passive viewing task with target stimuli shown randomly, possibly along with other, irrelevant stimuli. Within this pre-learning condition, it may then be possible to define a null distribution by finding a cortical region that is least sensitive to categorical differences. This should set a lower bound on the discriminability measure and can be used as a proxy for testing null effects in learning.

7.6 Concluding remarks

As summarized in Figure 1.1, this thesis presents a framework that synthesizes scientific experiments, neuroimaging technology and statistical methodologies in characterizing cortical spatiotemporal properties during visual category learning.

In the pursuit of this goal, I used MEG that proved to be critical in recording activities in the human cortex at a fine temporal resolution. Such precise timing helped reveal information flow in the cortex during the rapid visual discrimination time course. I further developed an experimental paradigm that allowed changes in information flow to be captured at a broader temporal scale that encompasses the course of learning.

To better characterize properties of the cortex, I developed novel statistical methods that improved the precision of spatiotemporal analysis in MEG. In these methods, I proposed that incorporation of domain and experimental knowledge is important for effective modeling. I also suggested that whereas statistical modeling improves experimental analysis, experimental design can also improve statistical modeling.

Using this methodological framework, I elucidated the dynamic roles of two key cortical regions in the process of visual category learning. I found that both the ventral stream and prefrontal cortex contain information about novel visual categories and such information is made available during the time course of categorization. However, emergence of categorical information becomes more prominent in the ventral stream toward the end of learning, but it diminishes in prefrontal cortex. These findings suggest that the ventral stream is crucial to the encoding of visual categories when categorization is proficient. They also suggest, however, that prefrontal cortex encodes visual categories in the initial stage of learning. These findings are in agreement with the prefrontal-complementary view in visual categorization, such that they support a coordinative relationship between prefrontal cortex and the ventral stream in the acquisition and discrimination of novel visual categories.

In summary, with the current thesis I contribute to a scientific and methodological framework for characterizing cortical spatiotemporal plasticity in visual category learning. In doing so, I have begun to unravel the complex spatiotemporal mechanisms in the human cortex and created a firm basis for future methodological innovations.

Bibliography

[1] M. S. Hämäläinen and R. J. Ilmoniemi. *Interpreting measured magnetic fields of the brain: Estimates of current distributions*. Helsinki University of Technology, Department of Technical Physics, 1984.

[2] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalographytheory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413, 1993.

[3] Y. Xu, G.P. Sudre, W. Wang, D.J. Weber, and R.E. Kass. Characterizing global statistical significance of spatio-temporal hot spots in MEG/EEG source space via excursion algorithms. *Statistics in Medicine*, 30:2854–2866, 2011.

[4] S. Pinker. Visual cognition. 1985.

[5] N. K. Logothetis and D. L. Sheinberg. Visual object recognition. *Annual Review of Neuroscience*, 19(1):577–621, 1996.

[6] I. W. R. Bushneil, F. Sai, and J. T. Mullin. Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, 7(1):3–15, 1989.

[7] L. B. Smith, S. S. Jones, B. Landau, L. Gershkoff-Stowe, and L. Samuelson. Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1):13–19, 2002.

[8] V. M. Sloutsky. From perceptual categories to concepts: What develops? *Cognitive Science*, 34(7):1244–1286, 2010.

[9] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.

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

[11] R. L. Fantz. Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, 146(3644):668–670, 1964.

[12] P. C. Quinn and P. D. Eimas. On categorization in early infancy. *Merrill-Palmer Quarterly*, pages 331–363, 1986.

[13] P. D. Eimas. Categorization in early infancy and the continuity of development. *Cognition*, 50(1):83–93, 1994.

[14] M. I. Posner and S. W. Keele. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3p1):353–363, 1968.

[15] M. I. Posner and S. W. Keele. Retention of abstract ideas. *Journal of Experimental Psychology*, 83(2):304–308, 1970.

[16] D. Homa and R. Vosburgh. Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory*, 2(3):322, 1976.

[17] E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439, 1976.

[18] K. Grill-Spector. The functional organization of the ventral visual pathway and its relationship to object recognition. *Attention and performance XX: Functional brain imaging of visual cognition*, pages 169–193, 2003.

[19] J. R. Folstein, T. J. Palmeri, and I. Gauthier. Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, 23(4):814–823, 2013.

[20] X. Jiang, E. Bradley, R. A. Rini, T. Zeffiro, J. VanMeter, and M. Riesenhuber. Categorization training results in shape-and category-selective human neural plasticity. *Neuron*, 53(6):891–903, 2007.

[21] D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, 23(12):5235–5246, 2003.

[22] E. M. Meyers, D. J. Freedman, G. Kreiman, E. K. Miller, and T. Poggio. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, 100(3):1407–1419, 2008.

[23] C. A. Seger and E. K. Miller. Category learning in the brain. *Annual Review of Neuroscience*, 33:203–219, 2010.

[24] M. Mishkin, L. G. Ungerleider, and K. A. Macko. Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, 6:414–417, 1983.

[25] A. D. Milner and M. A. Goodale. Two visual systems re-viewed. *Neuropsychologia*, 46(3):774–785, 2008.

[26] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106, 1962.

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

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

[29] T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 104(15):6424–6429, 2007.

[30] A. R. Damasio. Category-related recognition defects as a clue to the neural substrates of knowledge. *Trends in Neurosciences*, 13(3):95–98, 1990.

[31] A. R. Damasio, D. Tranel, and H. Damasio. Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience*, 13(1):89–109, 1990.

[32] M. J. Farah. Agnosia. *Current Opinion in Neurobiology*, 2(2):162–164, 1992.

[33] N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11):4302–4311, 1997.

[34] G. K. Aguirre, E. Zarahn, and M. Desposito. An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron*, 21(2):373–383, 1998.

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

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

[37] A. Puce, T. Allison, M. Asgari, J. C. Gore, and G. McCarthy. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *The Journal of Neuroscience*, 16(16):5205–5215, 1996.

[38] A. Martin, C. L. Wiggs, L. G. Ungerleider, and J. V. Haxby. Neural correlates of category-specific knowledge. *Nature*, 379(6566):649–652, 1996.

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

[40] J. Liu, A. Harris, and N. Kanwisher. Stages of processing in face perception: An Meg study. *Nature Neuroscience*, 5(9):910–916, 2002.

[41] B. Rossion, C. A. Joyce, G. W. Cottrell, and M. J. Tarr. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage*, 20(3):1609–1624, 2003.

[42] C. R. Gillebert, H. P. O. de Beeck, S. Panis, and J. Wagemans. Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape differences. *Journal of Cognitive Neuroscience*, 21(6):1054–1064, 2009.

[43] M. van der Linden, M. van Turennout, and P. Indefrey. Formation of category representations in superior temporal sulcus. *Journal of Cognitive Neuroscience*, 22(6):1270–1282, 2010.

[44] J. R. Folstein, I. Gauthier, and T. J. Palmeri. How category learning affects object representations: Not all morphspaces stretch alike. *Journal of Experimental Psychology-Learning Memory and Cognition*, 38(4):807, 2012.

[45] E. K. Miller, D. J. Freedman, and J. D. Wallis. The prefrontal cortex: Categories, concepts and cognition. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1424):1123–1136, 2002.

[46] J. D. Wallis, K. C. Anderson, and E. K. Miller. Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411(6840):953–956, 2001.

[47] J. D. Wallis and E. K. Miller. From rule to response: Neuronal processes in the premotor and prefrontal cortex. *Journal of Neurophysiology*, 90(3):1790–1806, 2003.

[48] R. Muhammad, J. D. Wallis, and E. K. Miller. A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. *Journal of Cognitive Neuroscience*, 18(6):974–989, 2006.

[49] O. E. Krigolson, L. J. Pierce, C. B. Holroyd, and J. W. Tanaka. Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *Journal of Cognitive Neuroscience*, 21(9):1833–1840, 2009.

[50] L. G. Ungerleider and J. V. Haxby. what and where in the human brain. *Cuurent Opnion in Neurobiology*, 4(2):157–165, 1994.

[51] Meenan J. P. Bulthoff H. H. Nicolle D. A. Murphy K. J. Goodale, M. A. and C. I. Racicot. Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology*, 4(7):604–610, 1994.

[52] M. Bar, K. S. Kassam, A. S. Ghuman, J. Boshyan, A. M. Schmid, A. M. Dale, M. S. Hämäläinen, K. Marinkovic, D. L. Schacter, B. R. Rosen, and E. Halgren. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2):449–454, 2006.

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

[54] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini. Human neural systems for face recognition and social communication. *Biological Psychiatry*, 51(1):59–67, 2002.

[55] J. V. Haxby and M. I. Gobbini. Distributed neural systems for face perception. *The Oxford Handbook of Face Perception*, pages 93–110, 2011.

[56] M. J. Tarr and I. Gauthier. FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3:764–770, 2000.

[57] A. Nestor, D. C. Plaut, and M. Behrmann. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences*, 108(24):9998–10003, 2011.

[58] G. Avidan, M. Tanzer, F. Hadj-Bouziane, N. Liu, L. G. Ungerleider, and M. Behrmann. Selective dissociation between core and extended regions of the face processing network in congenital prosopagnosia. *Cerebral Cortex*, doi: 10.1093/cercor/bht007, 2013.

[59] D. Pitcher, V. Walsh, G. Yovel, and B. Duchaine. Tms evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, 17(18):1568–1573, 2007.

[60] D. Pitcher, V. Walsh, and B. Duchaine. The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, 209(4):481–493, 2011.

[61] E. J. Barbeau, M. J. Taylor, J. Regis, P. Marquis, P. Chauvel, and C. Liégeois-Chauvel. Spatio temporal dynamics of face recognition. *Cerebral Cortex*, 18(5):997–1009, 2008.

[62] D. Cohen. Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(3843):784–786, 1968.

[63] JE Zimmerman. Heterodyne detection with superconducting point contacts and enhanced heterodyne signals from tightly coupled contacts. *Journal of Applied Physics*, 41(4):1589–1593, 1970.

[64] O. V. Lounasmaa. Experimental principles and methods below 1K. 1974.

[65] T. Ryhänen, H. Seppä, R. Ilmoniemi, and J. Knuutila. SQUID magnetometers for low-frequency applications. *Journal of Low Temperature Physics*, 76(5-6):287–386, 1989.

[66] I. Gauthier, M. J. Tarr, and D. Bub. *Perceptual Expertise: Bridging Brain and Behavior*. Oxford University Press, USA, 2009.

[67] P. Jolicoeur, M. A. Gluck, and S. M. Kosslyn. Pictures and names: Making the connection. *Cognitive Psychology*, 16(2):243–275, 1984.

[68] J. W. Tanaka and M. Taylor. Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23(3):457–482, 1991.

[69] N. Sigala and N. K. Logothetis. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, 415(6869):318–320, 2002.

[70] W. De Baene, B. Ons, J. Wagemans, and R. Vogels. Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & Memory*, 15(9):717–727, 2008.

[71] M. J. Fenske, E. Aminoff, N. Gronau, and M. Bar. Top-down facilitation of visual object recognition: Object-based and context-based contributions. *Progress in Brain Research*, 155:3, 2006.

[72] I. Gauthier, M. J. Tarr, A. W. Anderson, P. Skudlarski, and J. C. Gore. Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects. *Nature neuroscience*, 2(6):568–573, 1999.

[73] H.P. Op de Beeck, C.I. Baker, J.J. DiCarlo, and N.G. Kanwisher. Discrimination training alters object representations in human extrastriate cortex. *The Journal of Neuroscience*, 26(50):13025–13036, 2006.

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

[75] J. W. Tanaka and T. Curran. A neural basis for expert object recognition. *Psychological Science*, 12(1):43–47, 2001.

[76] L. S. Scott, J. W. Tanaka, D. L. Sheinberg, and T. Curran. A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitive Neuroscience*, 18(9):1453–1465, 2006.

[77] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

[78] B. Rossion, T. Curran, and I. Gauthier. A defense of the subordinate-level expertise account for the N170 component. *Cognition*, 85(2):189–196, 2002.

[79] A. C. N. Wong, I. Gauthier, B. Woroch, C. Debuse, and T. Curran. An early electrophysiological response associated with expertise in letter perception. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3):306–318, 2005.

[80] E. Halgren, T. Raij, K. Marinkovic, V. Jousmäki, and R. Hari. Cognitive response profile of the human fusiform face area as determined by MEG. *Cerebral Cortex*, 10(1):69–81, 2000.

[81] L. S. Scott, J. W. Tanaka, D. L. Sheinberg, and T. Curran. The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*, 1210:204–215, 2008.

[82] K. Kveraga, A. S. Ghuman, and M. Bar. Top-down predictions in the cognitive brain. *Brain and cognition*, 65(2):145, 2007.

[83] F. G. Ashby, L. A. Alfonso-Reese, A. U. Turken, and E. M. Waldron. A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105(3):442–481, 1998.

[84] W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, and E. Donchin. A neural system for error detection and compensation. *Psychological Science*, 4(6):385–390, 1993.

[85] C. B. Holroyd and M. G. H. Coles. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4):679–708, 2002.

[86] ODoherty J. P. Dayan P. Koltzenburg M. Jones A. K. Dolan R. J. Friston K. J. Seymour, B. and R. S. Frackowiak. Temporal difference models describe high-order learning in humans. *Nature*, 429:664–667, 2004.

[87] C. B. Holroyd, N. Yeung, M. G. H. Coles, and J. D. Cohen. A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, 134(2):163, 2005.

[88] V. Bruce and M. Burton. Learning new faces. MIT Press, 2002.

[89] G. Barragan-Jason, F. Lachat, and E. J. Barbeau. How fast is famous face recognition? *Frontiers in Psychology*, 3(454), 2012.

[90] S. V. Stevenage. Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology*, 89(1):39–57, 1998.

[91] S. L. Fairhall and A. Ishai. Effective connectivity within the distributed cortical network for face perception. *Cerebral Cortex*, 17(10):2400–2406, 2007.

[92] A. Ishai. Let's face it: It's a cortical network. *NeuroImage*, 40:415–419, 2008.

[93] G. Van Belle, T. Busigny, P. Lefèvre, S. Joubert, O. Felician, F. Gentile, and B. Rossion. Impairment of holistic face perception following right occipito-temporal damage in prosopagnosia: Converging evidence from gaze-contingency. *Neuropsychologia*, 49(11):3145–3150, 2011.

[94] A. Ishai, C. F. Schmidt, and P. Boesiger. Face perception is mediated by a distributed cortical network. *Brain Research Bulletin*, 67(1-2):87–93, 2005.

[95] I. Gauthier, M. J. Tarr, J. Moylan, P. Skudlarski, J. C. Gore, and A. W. Anderson. The fusiform face area is part of a network that processes faces at the individual level. *Journal of cognitive neuroscience*, 12(3):495–504, 2000.

[96] J. Sergent, S. Ohta, and B. Macdonald. Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain*, 115(1):15–36, 1992.

[97] M. I. Gobbini and J. V. Haxby. Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1):32–41, 2007.

[98] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini. The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6):223–233, 2000.

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

[100] W. K. Simmons, M. Reddish, P. S. Bellgowan, and A. Martin. The selectivity and functional connectivity of the anterior temporal lobes. *Cerebral Cortex*, 20(4):813–825, 2010.

[101] J. A. Pyles, T. D. Verstynen, W. Schneider, and M. J. Tarr. Explicating the face perception network with white matter connectivity. *PloS One*, 8(4):e61611, 2013.

[102] J. DeGutis and M. DEsposito. Distinct mechanisms in visual category learning. *Cognitive, Affective, & Behavioral Neuroscience*, 7(3):251–259, 2007.

[103] J. DeGutis and M. D'Esposito. Network changes in the transition from initial learning to well-practiced visual categorization. *Frontiers in human neuroscience*, 3, 2009.

[104] S. Bentin, T. Allison, A. Puce, E. Perez, and G. McCarthy. Electrophysiological studies of face perception in humans. *Journal of cognitive Neuroscience*, 8(6):551–565, 1996.

[105] B. Rossion and C. Jacques. The n170: Understanding the time-course of face perception in the human brain. *The Oxford Handbook of ERP components*, pages 115–142, 2011.

[106] J. W. Tanaka, T. Curran, A. L. Porterfield, and D. Collins. Activation of preexisting and acquired face representations: The N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience*, 18(9):1488–1497, 2006.

[107] R. D. Pascual-Marqui. Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods & Findings in Experimental & Clinical Pharmacology*, 24(Suppl D):5–12, 2002.

[108] K. Uutela, M. Hämäläinen, and E. Somersalo. Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage*, 10(2):173–180, 1999.

[109] D. Wipf and S. Nagarajan. A unified Bayesian framework for MEG/EEG source imaging. *Neuroimage*, 44(3):947–966, 2009.

[110] A. Gramfort, D. Strohmeier, J. Haueisen, M. Hämäläinen, and M. Kowalski. Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. *NeuroImage*, 2013.

[111] F. Jiang, L. Dricot, J. Weber, G. Righi, M. J. Tarr, R. Goebel, and B. Rossion. Face categorization in visual scenes may start in a higher order area of the right fusiform gyrus: Evidence from dynamic visual stimulation in neuroimaging. *Journal of Neurophysiology*, 106(5):2720–2736, 2011.

[112] D. F. Nichols, L. R. Betts, and H. R. Wilson. Decoding of faces and face components in face-sensitive human visual cortex. *Frontiers in Psychology*, 1, 2010.

[113] K. Grill-Spector, N. Knouf, and N. Kanwisher. The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7(5):555–562, 2004.

[114] A. Nestor, J. M. Vettel, and M. J. Tarr. Task-specific codes for face recognition: How they shape the neural representation of features for detection and individuation. *PLoS One*, 3(12):e3978, 2008.

[115] M. A. Uusitalo and R. J. Ilmoniemi. Signal-space projection method for separating MEG or EEG into components. *Medical and Biological Engineering and Computing*, 35(2):135–140, 1997.

[116] J. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela. Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208, 1999.

[117] R. J. Itier, A. T. Herdman, N. George, D. Cheyne, and M. J. Taylor. Inversion and contrast-reversal effects on face processing assessed by MEG. *Brain Research*, 1115(1):108–120, 2006.

[118] Z. Gao, A. Goldstein, Y. Harpaz, M. Hansel, E. Zion-Golumbic, and S. Bentin. A magnetoencephalographic study of face processing: M170, gamma-band oscillations and source localization. *Human Brain Mapping*, 34(8):1783–1795, 2012.

[119] M. H. Tong, C. A. Joyce, and G. W. Cottrell. Why is the fusiform face area recruited for novel categories of expertise? A neurocomputational investigation. *Brain Research*, 1202:14–24, 2008.

[120] L. Reddy and N. Kanwisher. Coding of visual objects in the ventral stream. *Current Opinion in Neurobiology*, 16(4):408–414, 2006.

[121] D. Maurer, R. L. Grand, and C. J. Mondloch. The many faces of configural processing. *Trends in cognitive sciences*, 6(6):255–260, 2002.

[122] R. Amishav and R. Kimchi. Perceptual integrality of componential and configural information in faces. *Psychonomic Bulletin & Review*, 17(5):743–748, 2010.

[123] N. Hadjikhani, K. Kveraga, P. Naik, and S. P. Ahlfors. Early (N170) activation of face-specific cortex by face-like objects. *Neuroreport*, 20(4), 2009.

[124] N. Tsuchiya, H. Kawasaki, H. Oya, M. A. Howard III, and R. Adolphs. Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. *PLoS One*, 3(12):e3892, 2008.

[125] R. N. Henson, G. Flandin, K. J. Friston, and J. Mattout. A parametric empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. *Human brain mapping*, 31(10):1512–1531, 2010.

[126] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, pages 1–26, 1979.

[127] K. A. Heller and Z. Ghahramani. Bayesian hierarchical clustering. In *Proceedings of the 22nd international conference on Machine learning*, pages 297–304, 2005.

[128] S. Waldert, H. Preissl, E. Demandt, C. Braun, N. Birbaumer, A. Aertsen, and C. Mehring. Hand movement direction decoded from MEG and EEG. *The Journal of Neuroscience*, 28(4):1000–1008, 2008.

[129] W. Wang, G. P. Sudre, Y. Xu, R. E. Kass, J. L. Collinger, A. D. Degenhart, A. I. Bagic, and D. J. Weber. Decoding and cortical source localization for intended movement direction with MEG. *Journal of Neurophysiology*, 104(5):2451–2461, 2010.

[130] S. Behseta, R. E. Kass, D. E. Moorman, and C. R. Olson. Testing equality of several functions: Analysis of single-unit firing-rate curves across multiple experimental conditions. *Statistics in Medicine*, 26(21):3958–3975, 2007.

[131] R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[132] Y. Xu, K. Heller, and Z. Ghahramani. Tree-based inference for Dirichlet process mixtures. pages 623–630, 2009.

[133] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15(1):1–25, 2002.

[134] E. Maris and R. Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164:177–190, 2007.

[135] S. Taulu, J. Simola, and M. Kajola. Applications of the signal space separation method. *IEEE Transactions on Signal Processing*, 53(9):3359–3372, 2005.

[136] A. P. Georgopoulos, R. Caminiti, J. F. Kalaska, and J. T. Massey. Spatial coding of movement: A hypothesis concerning the coding of movement direction by motor cortical populations. *Experimental Brain Research*, 7(32):336, 1983.

[137] A. P. Georgopoulos, J. F. Kalaska, R. Caminiti, and J. T. Massey. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *The Journal of Neuroscience*, 2(11):1527–1537, 1982.

[138] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, 1986.

[139] G. P. Sudre, Y. Xu, R. E. Kass, D. J. Weber, and W. Wang. Cluster-based algorithm for roi analysis and cognitive state decoding using single-trial source MEG data. In *17th International Conference on Biomagnetism Advances in Biomagnetism–Biomag2010*, pages 187–190, 2010.

[140] P. C. Hansen, M. L. Kringelbach, and R. Salmelin. *MEG: An Introduction to Methods*. Oxford University Press, USA, 2010.

[141] A. M. Dale and M. I. Sereno. Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of cognitive neuroscience*, 5(2):162–176, 1993.

[142] C. Lamus, M. S. Hämäläinen, S. Temereanca, E. N. Brown, and P. L. Purdon. A spatiotemporal dynamic distributed solution to the MEG inverse problem. *NeuroImage*, 63(2):894–909, 2012.

[143] W. Ou, P. Golland, and M. Hämäläinen. A distributed spatio-temporal EEG/MEG inverse solver. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 26–34. 2008.

[144] A. K. Liu, J. W. Belliveau, and A. M. Dale. Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proceedings of the National Academy of Sciences*, 95(15):8945–8950, 1998.

[145] A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren. Dynamic statistical parametric neurotechnique mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26(1):55–67, 2000.

[146] C. M. Bishop. *Pattern recognition and machine learning*. springer New York, 2006.

[147] K. S. Scherf, M. Behrmann, K. Humphreys, and B. Luna. Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Developmental Science*, 10(4):F15–F30, 2007.

[148] K. S. Scherf, B. Luna, G. Avidan, and M. Behrmann. "what" precedes "which": Developmental neural tuning in face- and place-related cortex. *Cerebral Cortex*, 21(9):1963–1980, 2011.

[149] A. S. Ghuman, M. Bar, I. G. Dobbins, and D. M. Schnyer. The effects of priming on frontal-temporal communication. *Proceedings of the National Academy of Sciences*, 105(24):8405–8409, 2008.

[150] B. M. Lake, R. Salakhutdinov, J. Gross, and J. B. Tenenbaum. One shot learning of simple visual concepts. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 2011.

MACHINE LEARNING
DEPARTMENT

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA15213

Carnegie Mellon