CARNEGIE MELLON UNIVERSITY

DIETRICH COLLEGE OF HUMANITIES AND SOCIAL SCIENCES DISSERTATION

Submitted in Partial Fulfillment of the Requirements
For the Degree of DOCTOR OF PHILOSOPHY

Title: “Searching for the visual components of object perception”

Presented by: Daniel Demeny Leeds

Accepted by: The Center for the Neural Basis of Cognition
August 19, 2013

Thesis Committee:
Michael Tarr, Chair
Robert Kass
Tom Mitchell
David Sheinberg (external)

Searching for the visual components of object perception

Daniel Demeny Leeds

July 2013

Dietrich College of Humanities and Social Sciences
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Michael Tarr, Chair
Robert Kass
Tom Mitchell
David Sheinberg

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.*

Keywords: Realtime fMRI, optimization, object recognition, computational modeling, intermediate feature representation

Abstract

The nature of visual properties used for object perception in mid- and high-level vision areas of the brain is poorly understood. Past studies have employed simplistic stimuli probing models limited in descriptive power and mathematical underpinnings. Unfortunately, pursuit of more complex stimuli and properties requires searching through a wide, unknown space of models and of images. The difficulty of this pursuit is exacerbated in brain research by the limited number of stimulus responses that can be collected for a given human subject over the course of an experiment.

To more quickly identify complex visual features underlying cortical object perception, I develop, test, and use a novel method in which stimuli for use in the ongoing study are selected in realtime based on fMRI-measured cortical responses to recently-selected and displayed stimuli. A variation of the simplex method [7] controls this ongoing selection as part of a search in visual space for images producing maximal activity — measured in realtime — in a pre-determined $1\ cm^3$ brain region. I probe cortical selectivities during this search using photographs of real-world objects and synthetic "Fribble" objects [76]. Real-world objects are used to understand perception of naturally-occurring visual properties. These objects are characterized based on feature descriptors computed from the scale invariant feature transform (SIFT, [36]), a popular computer vision method that is well established in its utility for aiding in computer object recognition and that I recently found to account for intermediate-level representations in the visual object processing pathway in the brain [35]. Fribble objects are used to study object perception in an arena in which visual-properties are well defined a priori. They are constructed from multiple well-defined shapes, and variation of each of these component shapes produces a clear space of visual stimuli.

I study the behavior of my novel realtime fMRI search method, to assess its value in the investigation of cortical visual perception, and I study the complex visual properties my method identifies as highly-activating selected brain regions in the visual object processing pathway. While there remain further technical and biological challenges to overcome, my method uncovers reliable and interesting cortical properties for most subjects — though only for selected searches performed for each subject. I identify brain regions selective for holistic and component object shapes and for varying surface properties, providing examples of more precise selectivities within classes of visual properties previously associated with cortical object representation [24, 63, 71]. I also find examples of "surround suppression," in which cortical activity is inhibited upon viewing stimuli slightly deviation from the visual properties preferred by a brain region, expanding on similar observations at lower levels of vision [22, 73].

Acknowledgments

I am greatly indebted to the many individuals who have provided me with support, guidance, insight, and inspiration throughout my years of doctoral study at Carnegie Mellon. I came to Carnegie Mellon for the opportunity to pursue both my passion for computer science and my passion for the study of the brain. I depart with gratitude to all those who ensured my experiences here well surpassed the grand plans with which I arrived.

My thanks go to Mike Tarr, who welcomed me into his lab and directed me to the exciting challenges of merging computational modeling and realtime fMRI. Mike has taught me invaluable lessons in science and research, in management and communication. I am deeply grateful for the countless opportunities I have received through my work with Mike to expand my intellectual and professional horizons around Pittsburgh and around the country.

My thanks go to the members of my thesis committee — Rob Kass, Tom Mitchell, and David Sheinberg — all of whom have provided me valuable guidance in my research and in my career. By the time Mike arrived at Carnegie Mellon and I joined the Tarr lab, Rob and Tom already had given me years of great insights and inspiration in computational neuroscience. I have had the good fortune to benefit even further from their perspectives as they joined my committee. David's spirited and probing discussions on my work, from low-level socket implementation to high-level evaluation of the complex visual properties uncovered by my methods, have served as key guideposts in a little-explored field. I am honored by his in-person attendance of my thesis defense. The lessons I have learned from Mike and from all of the members of my committee will continue to be a compass for me in science, engineering, academia, and life in general.

My thanks go to all my colleagues at the Center for Neural Basis of Cognition (CNBC) and in the Perceptual Expertise Network (PEN). From my labmates' anchoring in neuroimaging and psychology to my fellow students' expertise in a breadth of fields in computational neuroscience, my days on campus and in meetings around the country continuously have been fueled by exciting new ideas and by steady practical guidance. I am grateful for the encouragement and good company of my many friends in the Carnegie Mellon community, from Smith Hall to the Mellon Institute. While his stay in the Mellon Institute was brief, I have had the great fortune to work with and learn from Darren Seibert, who I know will continue to inspire as he completes his own doctoral studies at MIT. My terabytes of data are rivaled by my terabytes worth of gratitude to Anna Hegedus for her perpetual, enthusiastic support in all things computational and otherwise.

My thanks and love go to my friends and family who have sustained me with their advice, support, and inspiration. While there are countless friends to whom I owe my gratitude, I hold a special place for Daniel and Jen Golovin and for Kathryn Rivard for their invaluable direction in my doctoral journey. I am grateful to those who have borne with me in recent months as the lion's share of my time has shifted to preparation for my defense. I owe deep love and appreciation to my parents who have guided my personal and intellectual growth throughout my life. My love goes to my grandmother whose own path in academia from Budapest to New York has inspired me from my youth to this day.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The process of visual object recognition typically associates visual inputs — commencing with an array of light intensities falling on the retina — with semantic categories, for example, “cow,” “car,” or “face.” Nearly every theory or computational system that attempts to implement or account for this process, including the biological visual recognition system realized in the ventral occipito-temporal pathway of the human brain, assumes a feedforward visual processing hierarchy in which the features of representation progressively increase in complexity as one moves up in a feedforward manner [48] — the ultimate output being high-level *object representations* that allow the assignment of category-level labels. Within this framework, it is understood that there are levels of *intermediate* featural representations that, while less complex than entire objects, nonetheless capture important object-level visual properties [69]. Yet, for all the interest in uncovering the nature of such features in biological vision, they remain remarkably elusive. At present there is little empirical data on the neural representations of visual objects employed between input image and object representation. The goal of my research is to develop a new method for the exploration and identification of visual properties used to encode object information along the ventral pathway — the neural regions most associated with visual object processing.

1.1 Prior work

The issue of constituent features for object representation has been somewhat sidestepped in neuroimaging's focus on feature codes realized in "category-selective" regions within the ventral-temporal cortex. Most investigations of these regions — for example, the "fusiform face area" (FFA), associated with the detection and discrimination of faces [17, 20], the "parahippocampal place area" (PPA), associated with scene processing [12], or the lateral occipital complex (LOC), associated with the processing of objects more generally [16] — emphasize specific object-level experiential factors or input characteristics that lead to their recruitment, but never establish the underlying compositional properties that form the basis of the nominally category-specific representations. Studies of the visual properties that lead to the recruitment of these class-specific, functionally-defined brain regions largely have focused on the effects of spatial transformations and of the alteration of simple domain-specific features [68]. For example, images of objects from within a given class often elicit similar neural responses when scaled, rotated, or moved to different locations in the visual field; although in the case of picture-plane inversion or 3D rotation, there is typically some change in neural activity [19, 45]. To the extent that viable models of neural representation have been developed, they often have relied on the statistical analysis of the input space within a restricted object domain. For example, "face spaces," nominally capturing the featural dimensions of human face representation, can be defined using principal component analysis (PCA) on face images or using parameterized models that are generative for constructing, what appear to be, realistic new face stimuli [5, 14]. Alternatively, the featural dimensions of representation are sometimes made more explicit, as in Kravitz et al. [31] who found that the encoding of scenes in human visual cortex can be understood in terms of an underlying set of intuitive properties, including "open/closed" and "natural/artificial" [31]. These properties may be understood in light of Ullman et al.'s more general proposal that intermediate features may be construed as image fragments most-informative to a visual encoding/recognition task [69]. Further supporting the effectiveness of this sort of approach, there is some neurophysiological evidence consistent with the fragment framework laid out by Ullman and colleagues [18].

Current computational models commonly applied to biological object recognition tend to make only weak assumptions regarding the nature of intermediate, compositional features[1]. For example, almost all models employ variants of Gabor filterbanks, detecting local edges in visual stimuli, to explain selectivities in primary visual cortex (V1) [22]. Extending this approach, Kay et al., Freeman and Simoncelli, and Serre et al. propose hierarchies of linear and non-linear spatial pooling computations, with Gabor filters at the base, to model higher-level vision [13, 27, 57]. Kay et al. is well known for exploring how neural units coding for orientation and scale within human V1, V2, and V3 can be assembled to reconstruct complex images. Although the study provides an elegant demonstration of how modern fMRI methods may support more fine-grained analyses (and therefore inspiration for further investigation), it does not inform us regarding the nature of *intermediate features* past the already-established edge-statistic selectivities of V1 and V2. Indeed, we see this as the fundamental problem in any attempt to decode the features of "intermediate-level" object representation — the parameter space is extremely large and highly underspecified, therefore it is difficult to find effective models that fit the data.

This is not to say that studies of intermediate feature representation have not provided some more fine-grained data regarding the neural encoding of objects. For example, Tanaka explored the minimal visual stimulus that was sufficient to drive a given neuron at a level equivalent to the complete object [63]. He found that individual neurons in inferior temporal (IT) cortex were selective for a wide variety of simple patterns and shapes that bore some resemblance to the objects initially used to elicit a response from each neuron. Interestingly, Tanaka hypothesized that this pattern-specific selectivity is organized into a columnar structure that maps out a high-dimensional feature space for representing visual objects. Similarly, Yamane et al. and Hung et al. used a somewhat different search procedure employing a highly constrained, parameterized stimulus space to identify contour selectivity for individual neurons in primate visual

[1] The exception being Hummel and Biederman who made very strong assumptions as to the core features used in object representation. Unfortunately, in this model such strong assumptions work against any generality for the model [23].

cortex [24, 78]. They found that most contour-selective neurons in V4 and IT each encoded some subset of the parameter space. Moreover, each 2D contour within this space appeared to encode specific 3D surface properties and small collections of these contour-selective units were sufficient to capture the overall 3D appearance of an object or object part. Similarly, Cadieu et al. characterized V4 selectivities as sets of prefered and anti-prefered edges, defined in the context of the hierarchical biological model "HMAX" [4].

Dynamic selection of stimuli while recording and analyzing neural responses — pursued by Tanaka, Yamane et al., and Hung et al. — opens a promising direction in the study of intermediate visual feature representations. Drawing from the world of all visual objects, the potentially infinite images that can be displayed far outnumbers the trials available in any experiment. For a given neural unit, one would like to converge quickly on the visual properties of greatest interest and avoid undue exploration of properties having no effect on neural activity. This concern is particularly pressing when performing human (rather than primate) studies, in which subjects remain in the lab at most for a few hours over the course of several days, permitting the exploration of possibly one hundred stimuli rather than the thousands possible in animal recordings. A recent rise in realtime neuroimaging analyses sets the stage for dynamic stimulus selection in human imaging studies. Shibata et al. used neurofeedback from V1 and V2 to control the size of a circular stimulus displayed to subjects and Ward et al. explored realtime mapping of the early visual field using Kalman filtering [59, 74]. While focusing on early visual regions, these studies show the promise of incorporating realtime analysis and feedback into neuroimaging work.

1.2 Approach

I utilize realtime fMRI analysis in conjunction with approaches from optimization and computer vision to further address the question of features underlying object representation in the brain. Similar to Yamane et al. and Hung et al., I explore the response of selected regions in the ventral visual pathway in a parameterized space of stimuli, searching for stimuli producing maximal activity from the selected regions. Recent work guides my selection of candidate models for the structure of stimulus space and for the principles determining cortical response.

I study visual object perception in the brain using fMRI — functional magnetic resonance imaging. fMRI measures neural activity across the brain as reflected by blood oxygen level. Increased firing within a group of neurons results in increased blood flow to these neurons to provide sufficient oxygen to support their activity. The changes in blood oxygen level produce changes in the magnetic field employed and measured by fMRI. The technology, as used in my work, records activity at a spatial resolution of ˜2 mm and a temporal resolution of 1 s. fMRI provides benefits over electrode recordings used in animal studies above, as it is non-invasive — no surgery is required for placement of electrodes — and records activity across the full brain — spanning diverse potential regions of interest in the ventral pathway. Unfortunately, blood flow requires as many as 3 to 6 seconds to respond to neural activity, while firing occurs at periods of ˜10 ms, hampering our ability to understand temporal dynamics underlying object perception across the brain. However, it is well fit for the present task, measuring static selectivities of one cubic millimeter to one cubic centimeter brain regions. Alternative neuroimaging techniques, such as magnetoencephology (MEG), offer higher temporal resolution measurements of brain activity, but have much more poor spatial resolution.

The structure of the stimulus space to be used in my study is determined by the visual features chosen to characterize the stimuli. While strong sets of candidate features for cortical object perception are unclear, a variety of visual properties have shown promise. Simple two- and three-dimensional surface contours have provided insights into neural coding [24, 78] and representation of objects as the combination of simple component shapes has accounted for facets of perception [23]. Select models drawn from computer vision literature — incorporating diverse linear and non-linear operations on image properties to maximize machine performance in object recognition tasks — have been shown as strong proxy theories of features used in biological object representation in intermediate stages of the ventral pathway [35]. In particular, Leeds et al. associates the scale invariant feature transform, “SIFT” [36] with visual reprentation in the fusiform cortex. In my present work, I define two visual representation spaces — the first space based on SIFT features, computed from photographs of real-world objects, and the second space based on parametric descriptions of synthetic “Fribble” objects [76] constructed from simple textured component shapes.

A variety of approaches can be used to characterize the response of a cortical region to the properties described in a given stimulus feature space. Models may focus on prediction of individual voxel responses for viewed stimuli [27], prediction of stimulus groupings eliciting similar multi-voxel activities [35], successful classification of viewed stimuli based on voxel activity [27], or identification of stimuli producing maximal response in a recorded neural unit [24, 63, 78]. Each of these focii can benefit from dynamic stimulus selection, to maximally sharpen model accuracy over limited recording time. Active learning literature — studying the strategy for selecting a small number of examples from which to maximize the effects of supervised learning — particularly focuses on efficiently learning boundaries in feature space for optimal binary classification. In contrast, work in dynamic stimulus selection for studying intermediate features in vision has focused on optimization — searching for the stimulus that produces the highest response for a given neural unit [24, 78]. I pursue the latter approach, developing novel realtime analysis software to perform a search for the most-preferred stimulus for a given brain region.

I select four one cubic centimeter cortical regions of interest (ROIs) to study in the ventral pathway of each of twenty subjects — ten viewing real-world objects and ten viewing Fribble objects. For each ROI, I search in the associated feature space to identify the stimuli producing maximal activity, selecting new stimuli for the search in realtime based on ROI responses to recently displayed stimuli. Optimally, most selected stimuli will cluster around a location in feature space corresponding to the visual properties for which the ROI is most selective. I assess the performance of my realtime search method, in addition to studying the resulting findings of complex featural selectivities.

While searches for many ROIs failed to converge to reveal clear featural selectivitivies, my method uncovers reliable and interesting cortical properties for a subset of regions in most subjects. I identify brain regions selective for holistic and component object shapes and for varying surface properties, providing examples of more precise selectivities within classes of visual properties previously associated with cortical object representation [24, 63, 71]. I also find examples of "surround suppression," in which cortical activity is inhibited upon viewing stimuli slightly deviated from the visual properties preferred by a brain region, expanding on similar observations at lower levels of vision [22, 73]. Stimuli producing the highest responses for an ROI often were distributed across multiple areas of visual feature space, potentially reflecting multiple distinct neural populations with distinct selectivities included within the one cubic centimeter ROI.

1.3 Contributions

My work on realtime fMRI analysis and the search for complex visual feature selectivities underlying cortical object perception explores myriad theoretical and technical questions with impact on multiple fields.

- *Realtime neural data processing* is novel in studies of perception and particularly in neuroimaging. Standard preprocessing methods for fMRI data incorporate information across hours of scanning and employ computations that can require many minutes to perform. Further processing is required to isolate a representation of cortical response from a selected cortical region. I introduce and assess adaptations of these methods for efficient, and more cursory, analyses that can be completed sufficiently quickly to provide the information needed to intelligently select new stimuli to show a subject based on his or her cortical responses to past stimuli.

- *Realtime communication among computers and programs* is essential to pass data quickly among the fMRI scanner, signal processing programs, programs intelligently selecting new stimuli based on recent and past cortical responses to stimuli, and programs displaying new stimuli to the subject. I use a collection of shared files and inter-program sockets for these communications and assess their performance.

- *A space of visual objects* must be established as the context in which to search for complex visual properties maximally activating a selected cortical region of interest. I presume objects are organized in the brain based on composition visual features, but the identity of these features — and of optimal candidate stimuli to illustrate these features — remains much in question. For example, in the study of IT neurons Hung et al. pursued an implicit space of synthetic stimuli generated from a medial axis representation [24]; however, it is unclear such a representation is a strong model for voxel-level encoding. Furthermore, the synthesized monochromatic blob stimuli used by Hung et al. may produce neural activity that generalizes poorly to neural responses to real-world objects. As discussed above, I define, use, and assess two spaces of visual objects, the first based on my recent work linking voxel-level coding and computer vision representations of real-world objects — particularly focusing on SIFT [36] (Chap. 2) — and the second based on a set of synthesized "Fribble" objects [76] with manually-defined axes of variability for textured component shapes.

- It also is unclear what is the optimal *search method* used to quickly identify stimuli and visual properties producing maximal activity from a selected cortical. Search is further complicated by the noise included in each measured cortical response to previously-viewed stimuli. I adapt and assess the performance of a version of the simplex method incorporating uncertainty through simulated annealing [7].

- *Regions of interest* for study of intermediate feature coding can be drawn from across the ventral pathway. Indeed, the ability of fMRI to record neural activity at high spatial resolution across the brain is one of the central benefits for its use in my study of human visual coding. Unfortunately, the anatomical areas on which to focus are uncertain when moving past V1 and V2, and the desirable expanse of these regions is similarly unclear. I select and assess 125-voxel cube ROIs identified by a searchlight method inspired by my recent work identifying voxel regions linked with computer vision representations of objects (Chap. 2). ROIs for each subject are identified based on data collected in a scan prior to the realtime scans.

1.4 Thesis organization

The rest of the thesis is organized as follows. In Chap. 2, I discuss my work modeling cortical visual representations with computer vision methods, further detailed in Leeds et al. [35]. This chapter introduces SIFT and representational dissimilarity analysis, which are important components in my realtime search. In Chap. 3, I discuss the methods used in my current search for intermediate featural selectivities. These methods address technical decisions made in fMRI study design, signal processing, software communications, image representations, and search technique. In Chap. 4, I present the performance of my searches, study results of processing decisions, and observe the stimuli producing strongest and weakest responses from the set regions of interest. In Chap. 5, I discuss the implications of my findings to the development of future realtime fMRI investigations and to the understanding of visual object perception in the brain. I propose further work in Chap. 6.

Chapter 2

Related work

My present study employs realtime search to identify complex visual properties used in the ventral pathway. Many of the methods contributing to the realtime search are drawn from my recent work evaluating computer vision methods as potential models for cortical object representation. In this recent work, I use a searchlight procedure [32] to select a contiguous group of voxels for each analysis, use *representational dissimilarity analysis* [33] to compare groupings of object stimuli based on their voxel and computer vision encodings, and identify the scale invariant feature transform, SIFT [36], as a strong model of visual representation in intermediate regions of the ventral object preception pathway. In Chap. 3, I discuss the use of voxel searchlights and representational dissimilarity analysis to identify regions of interest in which to perform realtime searches for complex feature selectivities; I also discuss the use of SIFT to parameterize visual properties to be searched. In the present chapter, I reproduce my paper currently in revision, Leeds et al. [35], to discuss the use of these analysis methods in my recent investigation, identifying links between computer vision models and cortical encoding in the ventral pathway.

2.1 Introduction

The process of visual object recognition typically associates visual inputs — commencing with an array of light intensities falling on the retina — with semantic categories, for example, “cow,” “car,” or “face.” Nearly every model, theory, or computational system that attempts to implement or account for this process, including the biological visual recognition system realized in the ventral occipito-temporal pathway of the human brain, assumes a feedforward visual processing hierarchy in which the features of representation progressively increase in complexity as one moves up in a feedforward manner [48] — the ultimate output being high-level *object representations* that allow the assignment of category-level labels. It goes almost without saying that within this framework, one presupposes levels of *intermediate* featural representations that, while less complex than entire objects, nonetheless capture important object-level visual properties [69]. Yet, for all the interest in uncovering the nature of such features with respect to biological vision, they remain remarkably elusive. At present there is little empirical data on the neural representations of visual objects in the netherworld between input image and object representation. The goal of our present study is to unravel how the human brain encodes object information along the ventral pathway — the neural “real estate” associated with visual object processing.

Given the paucity of data that bears on this question, how do we develop viable theories explicating the (compositional) features underlying the neural representation of objects? One possibility is to focus on feature codes realized in “category-selective” regions within the ventral-temporal cortex. However, most investigations of these regions — for example, the “fusiform face area” (FFA), associated with the detection and discrimination of faces [17, 20], the “parahippocampal place area” (PPA), associated with scene processing [12], or the lateral occipital complex (LOC), associated with the processing of objects more generally [16] — emphasize specific object-level experiential factors or input characteristics that lead to their recruitment, but never establish the underlying compositional properties that form the basis of the nominally category-specific representations. Most studies of the visual properties that lead to the recruitment of these class-specific, functionally-defined brain regions have focused on the effects of spatial transformations and of the alteration of simple domain-specific features [68]. For example, images of objects from within a given class often elicit similar neural responses when scaled, rotated, or moved to different locations in the visual field; although in the case of picture-plane inversion or 3D rotation, there is typically some change in neural activity [19, 45]. To the extent that viable models of neural representation have been developed, they have relied on the statistical analysis of the input space within a restricted object domain. For example, "face spaces," nominally capturing the featural dimensions of human face representation, can be defined using principal component analysis (PCA) on face images or using parameterized models that are generative for constructing, what appear to be, realistic new face stimuli [5, 14]. Alternatively, the featural dimensions of representation are sometimes made more explicit, as in Kravitz et al. [31] who found that the encoding of scenes in human visual cortex can be understood in terms of an underlying set of intuitive properties, including "open/closed" and "natural/artificial" [31].

This is not to say that studies of intermediate feature representation have not provided some more fine-grained data regarding the neural encoding of objects. For example, Tanaka explored the minimal visual stimulus that was sufficient to drive a given neuron at a level equivalent to the complete object [63]. He found that individual neurons in IT were selective for a wide variety of simple patterns and shapes that bore some resemblance to the objects initially used to elicit a response from each neuron. Interestingly, Tanaka hypothesized that this pattern-specific selectivity is organized into a columnar structure that maps out a high-dimensional feature space for representing visual objects. Similarly, Yamane et al. and Hung et al. used a somewhat different search procedure employing a highly constrained, parameterized stimulus space to identify contour selectivity for individual neurons in primate visual cortex [24, 78]. They found that most contour-selective neurons in V4 and IT each encoded some subset of the parameter space. Moreover, each 2D contour within this space appeared to encode specific 3D surface properties and small collections of these contour-selective units were sufficient to capture the overall 3D appearance of an object or object part. Within the human neuroscience literature, the study most often associated with feature decoding is that of Kay et al. who explored how neural units coding for orientation and scale within human V1, V2, and V3 can be assembled to reconstruct complex images [27]. Although Kay et al. provide an elegant demonstration of how modern fMRI methods may support more fine-grained analyses (and therefore inspiration for further investigation), their work does not inform us regarding the nature of *intermediate features* in that Kay et al. relied on well-established theories regarding the featural properties of V1 and V2. That is, they decoded features within a reasonably well-understood parameter space in which it is generally agreed that the particular brain regions in question encode information about the orientations and scales of local edges. Indeed, we see this as the fundamental problem in any attempt to decode the features of "intermediate-level" object representation — the parameter space is extremely large and highly underspecified, therefore it is difficult to find effective models that fit the data. As such, Ullman et al.'s proposal that intermediate features can be construed as image fragments of varying scale and location — leaving the content of said fragments entirely unspecified — is perhaps the strongest attempt yet at capturing task-relevant object information encoded within the human ventral pathway [69]. Supporting the effectiveness of this sort of approach, there is some neurophysiological evidence consistent with the fragment framework laid out by Ullman and colleagues [18].

Finally, we note that current computational models commonly applied to biological object recognition tend to make only weak assumptions regarding the nature of intermediate, compositional features[1]. For example, almost all models employ variants of Gabor filterbanks, detecting local edges in visual stimuli, to explain selectivities in primary visual cortex (V1) [22]. Extending this approach, both Kay et al. and Serre et al. propose hierarchies of linear and non-linear spatial pooling computations, with Gabor filters at the base, to model higher-level vision [27, 57]. One such hierarchical model, "HMAX" [4], partially predicts neural selectivity in the mid-level ven-

[1] The exception being Hummel and Biederman who made very strong assumptions as to the core features used in object representation. Unfortunately, in this model such strong assumptions work against any generality for the model [23].

tral stream (V4) for simple synthetic stimuli. However, HMAX imperfectly clusters pictures of real-world objects relative to clustering based on neurophysiological and fMRI data from IT [33].

To further address the question of the compositional features underlying neural object representation, we employed several models of visual representation drawn from machine vision — each provides a putative hypothesis regarding the features used in object perception. These representations incorporate diverse linear and non-linear operations on image properties to maximize machine performance in object detection and recognition tasks. As such, we are relying on these models as proxies for theories of features for biological object representation. Given this set of models, we collected data on human object processing using fMRI and a simple object perception task. We then correlated the resultant neural data with the object dissimilarity matrices predicted by each computer vision model, thereby establishing a correspondence between each model and patterns of neural activity in specific spatial locations within the brain. Consistent with the fact that these models make different assumptions with respect to object representation, we found that different models were associated with neural object encoding in different cortical locations. However, consistent with the overal visual nature of all of these representations, we observed that most of these associations lay within the ventral and dorsal visual cortices. Of particular interest, one popular machine vision represention, the scale invariant feature transform, “SIFT” [36], which encodes images using relatively simple local features, was the most strongly associated with measured neural activity in the brain regions typically associated with mid-level object perception (e.g., fusiform cortex). To better explicate how we arrived at this finding, we next define what is meant by “dissimilarity” with respect to both computational models and neural data.

2.1.1 Representational dissimilarity analysis

To assess model performance, neural stimulus representations as measured by fMRI and a given machine vision model were compared using representational dissimilarity analysis. For each set of voxels and for each model, a pairwise distance matrix was computed reflecting which sets of stimulus images were considered to be similar and which were considered to be different (more detail is given in Sec. 2.2.7). Model/neural matrices were more correlated when the two corresponding representations of the stimuli group the considered images in a similar manner. Kriegeskorte et al. demonstrated the advantages of dissimilarity analysis in observing and understanding complex patterns of neural activity — in their case, a collection of spatially contiguous voxels [33]. We similarly wished to understand object encoding across restricted volumes of voxels. The advantage of this approach is that it allows us to judge a model's descriptive power without requiring identification of the exact — most-likely non-linear — mapping between model and voxel responses. Indeed, O'Toole et al. and Kiani et al. pursued related cortical-computational dissimilarity analyses in studying visual perception, finding that the organization of object categories in IT is based, in part, on visual similarity [43] and, in part, on higher-order class information [28]. The ability of this method to bypass the issue of learning a direct mapping between model predictions and neural data provides particular benefit for fMRI studies in that it obviates the need to split rather limited datasets in order to cross-validate.

2.2 Methods

2.2.1 Stimuli

A picture and word set comprised of 60 distinct color object photos displayed on 53% gray backgrounds and their corresponding basic-level names was used as stimuli (Fig. 2.1). The specific category of each object was selected to match the 60 objects used in Just et al. [26][2]. The photographic images used in our study were taken from web image searches; therefore, we do not have the rights to redistibute the actual images. The 60 objects included five examples from each of twelve diverse semantic classes, for example, tools, food, mammals, or body parts. Each object

[2] The particular images used in Just et al. were drawn from the "Snodgrass and Vanderwart" line-drawing image dataset [61].

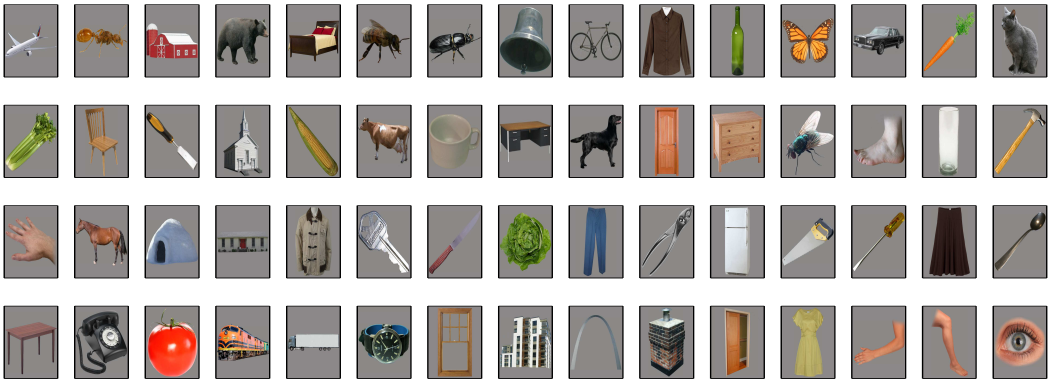


Figure 2.1: The 60 image stimuli displayed to subjects

was depicted by a single image. Although visual similarities among stimuli can be seen across semantic groups, such as knife and carrot (thin and slanted up to the right) or tomato and eye (circular in the image plane), objects within a semantic class were typically more similar to one another relative to their across-class similarities. Our use of real-world images of objects rather than the hand-drawn or computer-synthesized stimuli employed in the previously-discussed studies of mid-level visual coding, for example, Cadieu et al. [4] and Yamane et al. [78], is intended to more accurately capture the importance of the broad set of naturally-occuring visual features in object perception.

2.2.2 Subjects

Five subjects (one left-handed, one female, age range 20 to 24) from the Carnegie Mellon University community participated, gave written informed consent, and were monitarily compensated for their participation. All procedures were approved by the Institutional Review Board of Carnegie Mellon University.

2.2.3 Experimental Design

All stimuli were presented using Matlab [38] and the Psychophysics Toolbox [3, 44] controlled by an Apple Macintosh and were back projected onto a white screen located at the head end of the bore using a DLP projector (Sharp XG-P560W). Subjects viewed the images through a mirror attached to the head coil with object stimuli subtending a visual angle of approximately $8.3\,deg \times 8.3\,deg$. Each stimulus was displayed in the center of the screen for 2.0 s followed by a blank 53% gray screen shown for a time period randomly selected to be between 500 and 3000 ms, followed by a centered fixation cross that remained displayed until the end of each 10 s trial, at which point the next trial began. As such, the SOA between consecutive stimulus displays was fixed at 10 s. Subjects were instructed to press a button when the fixation cross appeared. The fixation onset detection task was used to engage subject attention throughout the experiment. No other task was required of subjects, meaning that our study addresses object perception under passive viewing conditions.

The 10 s SOA was chosen to minimize temporal overlap between voxel BOLD responses for multiple stimuli — a slow event-related design based on the assumption that the hemodynamic response in the ventral-temporal cortex has decreased to a sufficient degree in the 10–12 s after stimulus onset to minimize the noise in our measurements of the cortical responses.

The stimuli were presented in 24 six-minute runs, spread across three 1-hour scanning sessions and arranged to minimize potential adaptation and priming effects. Each scanning session included two sets of four runs. Each run contained 15 word and 15 picture stimuli, ordered such that the picture and the word corresponding to the same object were not viewed in direct succession and all stimuli were viewed exactly once in each four-run set to avoid priming and adaptation effects. Trials using the word stimuli were not analyzed or otherwise considered as part of our present study. Stimulus order was randomized across blocks and across subjects. Over the course of the experiment, each subject viewed each picture and each word six times; averaging across multiple repetitions was performed for each stimulus, described below, to reduce trial-by-trial noise.

The first session for each subject also included functional localizer scans to identify object selective cortex — namely, the Lateral Occipital Complex (LOC) — a functionally defined region [30] that we consider separately from the anatomically-identified lateral occipital cortex

(LO; although there is overlap between the two areas). For this localizer, 16 s blocks of common everyday objects were alternated with 16 s blocks of phase-scrambled versions of the same objects, separated by 6 s of fixation [16, 30]. Phase scrambling was achieved by taking the Fourier transform of each image, randomizing the resulting phase values while retaining the original frequency amplitudes, and reconstructing the image from the modified Fourier coefficients [53]. Within each block, 16 images, depicting 14 distinct objects, were shown for 800 msec each, each object being followed by a 200 msec gray screen. Two of the objects were sequentially repeated once during each block — to maintain attention, subjects were instructed to monitor for this, performing a one-back identity task in which they responded via a keypress whenever the same object image was repeated across two image presentations. Six blocks of both the intact and scrambled objects conditions were presented over the 282 s scan [47]. The object images used in the localizer scans were different from the object picture stimuli discussed in Sec. 2.2.1. LOC area(s) were identified as those brain regions more selective for intact versus scrambled objects. LOC areas included all regions containing spatially contiguous voxels (no minimum cluster size) for which beta weights for the block design had significance level of $p < .005$.

To provide anatomical information, a T1-weighted structural MRI was performed between runs within the first scanning session for each subject.

2.2.4 fMRI Procedures

Subjects were scanned using a 3.0 T Siemens Verio MRI scanner with a 32-channel head coil. Functional images were acquired with a gradient echo-planar imaging pulse sequence (TR 2 s, TE 26 ms, flip angle $90°$, $2\ mm \times 2\ mm \times 3\ mm$ voxels, field of view $192 \times 192\ mm^2$, 31 oblique-axial slices). Slices spanned the majority of the brain, to observe relevant stimulus representations beyond the visual streams (Fig. 2.2). An MP-RAGE sequence (flip angle $9°$, $1\ mm^3$ voxels, field of view $256 \times 256\ mm^3$, 176 sagittal slices) was used for anatomical imaging.

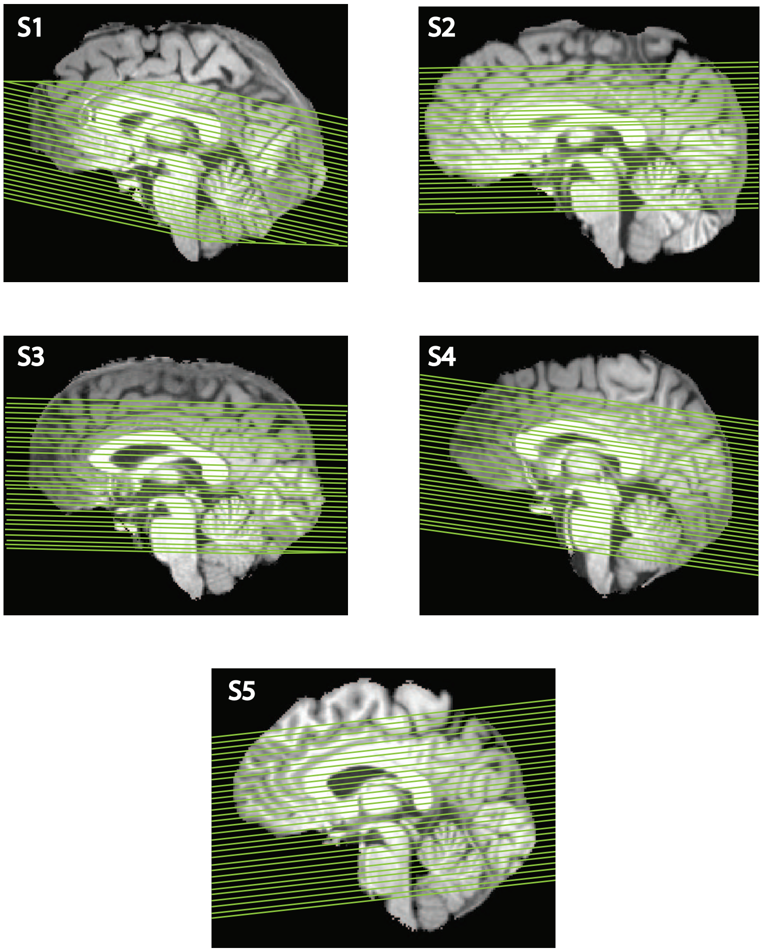


Figure 2.2: Slice coverage for all subjects.

2.2.5 Preprocessing

Functional scans were coregistered to the anatomical image and motion corrected using AFNI [46]. Highpass filtering was implemented in AFNI by removing sinusoidal trends with periods of half and full length of each run (338 s) as well as polynomial trends of orders one through three. The data then were normalized so that each voxel's time-course was zero-mean and unit-variance [26]. To allow multivariate analysis to exploit information present at high spatial frequencies, no spatial smoothing was performed [62].

For each stimulus repetition, the measured response of each voxel consisted of five data samples starting 2 s/1 TR after onset, corresponding to the 10 s between stimuli. Each five-sample response was consolidated into a weighted sum, intended to estimate the peak response. This was accomplished as one step in a "searchlight" process [32]: 123-voxel searchlight spheres — with radii of 3 voxels — were defined centered sequentially on every voxel in the brain. The average five-sample response of voxels across this sphere and across all stimulus presentations was computed. For a given searchlight, for each stimulus, each voxel was assigned a number based on the dot product of this average response and the voxel's mean response across all six repetitions of that stimulus. To the extent that hemodynamic responses are known to vary across cortical regions, this procedure allowed us to take into account a given voxel's local neighborhood mean-response shape. Fitting the local average response may provide a more accurate model of the relative activity of voxels across a sphere as compared to fitting a fixed response function across the whole brain.

2.2.6 Initial Voxel Selection

Data analysis was performed on the entire scanned brain volume, with subregions defined by the sequential searchlight. To distinguish the brain, in its entirety, from the surrounding skull and empty scanner space, a voxel mask was applied based on functional data using standard AFNI procedures. Voxels outside the full-brain mask were set to 0 at all time points; these 0 values were incorporated into searchlight analyses when performed close to the exterior of the brain. Because the inclusion of these null values was consistent across all stimuli, it did not affect the patterns of the dissimilarity matrices.

2.2.7 Representational Dissimilarity Measures

As discussed earlier, we employed *representational dissimilarity* as means for relating the neural representation of objects to the representation of the same objects within a variety of computer vision models. A *representational dissimilarity matrix* (RDM) D^m was computed for each encoding model m such that

$$D^m_{i,j} = d^m(s^i, s^j) \tag{2.1}$$

meaning the matrix element in the i^{th} row and j^{th} column contains the distance, or *dissimilarity*, between the i^{th} and j^{th} stimulus s^i and s^j in the model m. A given dissimilarity matrix captures which visual objects are clustered together by the corresponding representation. The searchlight procedure was then used to identify voxel clusters with D^ms similar to the RDMs of each computer vision model.

A 123-voxel searchlight sphere was defined centered on each voxel in the brain [32], with individual voxel responses to each stimulus computed as described in Sec. 2.2.5. For a given searchlight centered on voxel-location (x, y, z), each RDM entry $D^{srchlt_{x,y,z}}_{i,j}$ was defined as one minus the Spearman correlation between the voxel responses for stimuli i and j [33]:

$$d^{srchlt_{x,y,z}}(s^i, s^j) = 1 - r(v(s^i), v(s^j)) \tag{2.2}$$

The 123-element vector $v(s^i)$ represents the voxel responses for stimulus i averaged across all six blocks to compute the RDM. This averaging enhances the stimulus-specific response over the independent time-varying noise, providing a more stable estimate of the searchlight response to each stimulus.

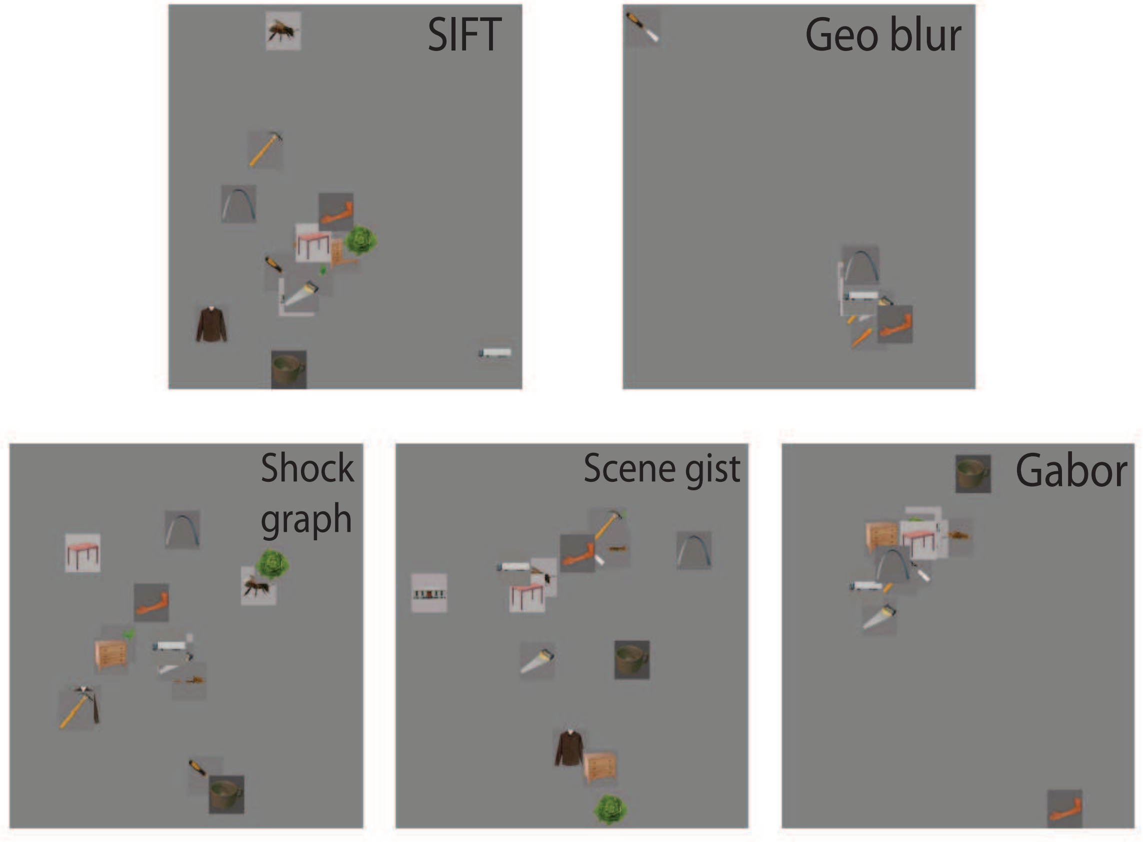


Figure 2.3: Multi-dimensional scaling visualization of the relative clustering of 15 of the stimulus pictures based on each computational model under analysis.

Five computational models of object representation were implemented for comparison with the neural representation of objects. Four of these methods were drawn from popular computer vision models with varied approaches to object representation, while the fifth was a standard computational model designed to account for neural responses relatively early in primate visual cortex. Distinct distance metrics $d^m(\cdots)$ were derived from each method. These models, ordered from relatively more local to more global feature representations, are described next.

1. **Gabor filterbank** The Gabor filter is a well-established model of cell and voxel-level selectivity in V1 [10]. Each filter identifies a specific local oriented edge in the stimulus. A bank of filters spans edge angle, position, and size. The first four levels of the filterbank

used in Kay et al. [27] were implemented and used to represent each image. The real-valued responses for each filter were recorded in a vector. Euclidean distance was used to measure the difference between the vectors associated with each pair of images.

2. **Geometric Blur** Geometric Blur uses local image properties at selected interest points. The relative locations of these interest points are included in the image encoding, thus incorporating more global geometric properties of each object. Feature vectors consist of pixel values regularly sampled in radiating circles around the interest point, with the starting point for sampling being determined by local image statistics. Pixel values are blurred over space, with increasing blur for higher-radius circles. This approach emphasizes precise details at each interest point and lower-resolution context from the surrounding region, similar to the decrease in spatial resolution away from the retina's focal point in early vision.

   Interest points were selected randomly from edges found by a Canny edge detector [6]. Features were extracted through an implementation of the algorithm described in Berg et al. [2]. For each pair of images, each interest point in one image (the image with fewer points) was matched with the point spatially closest in the second image. The dissimilarity for each pair of points was computed by taking the weighted sum of the negative correlation between the two feature vectors, the Euclidean distance between the points, and the change in circle orientation as defined in Berg et al. [2]. The final dissimilarity between images was found by summing the dissimilarities for all pairs of points. This incorporates both global geometric information and spatially-sampled local image statistics.

3. **Scale Invariant Feature Transform** Scale Invariant Feature Transform or "SIFT" features [36] have been widely used in computer vision systems for visual object recognition. This approach selects significant pixel patches of an image and captures associated visual properties that are invariant to a variety of common transformations, such as rotation, translation, image enlargement, and (potentially) changes in ambient lighting. More

specifically, for a given image, *interest points* are identified and a scaled, rotated frame is defined around each point. For each frame, a feature vector is computed to encode the local image properties, defined as coefficients of a pyramid of image gradients increasing in spatial scope. SIFT features were extracted from the 60 object stimuli using the VLFeat package for Matlab [70], with default settings when not otherwise specified.

A *bag of features* approach was used to compare SIFT features for pairs of images [40]. Conceptually, each of the SIFT feature vectors in each stimulus is categorized as one of 128 "words," where the words are consistently defined across all 60 images. Each image is then characterized by the frequency of each of the possible words. More specifically, k-means clustering is performed on the feature vectors from all interest points of all pictures, placing each vector into one of 128 groups. Assignment of multi-dimensional continuous-valued vectors to a small number of groups greatly reduces SIFT's representational complexity. A histogram is computed to find the frequency of each vector-group in each image and the histograms were normalized to sum to 1. For each image pair, the Kullback-Leibler (KL) divergence was used to measure the difference between the resulting two normalized histograms.

4. **Shock Graphs** The Shock Graph provides a complete and unique representation of a given visual object's external shape by constructing a modified form of Blum's medial axis [29] based on the object's silhouette. The graph is a set of vertices, edges, and *shock* labels, $\mathcal{G} = (V, E, \lambda)$. Each vertex represents a point or interval along the medial axis, edges connect spatially neighboring points/intervals, and each label specifies the curvature of the portion of the silhouette associated with the corresponding vertices:

   - $\lambda = 1$ when curvature is monotonic; object only widens or only narrows over an interval
   - $\lambda = 2$ when curvature reaches a local minimum at a point; object narrows prior to the point in the axis and widens after the point

- $\lambda = 3$ when curvature remains constant over an interval; object silhouette ends in a semi-circle or object is a circle
- $\lambda = 4$ when curvature achieves a local maximum at a point; object widens prior to the point in the axis and narrows after the point

Further details are provided by Siddiqi et al. [60]. The distance between graph pairs was computed using a graph-matching technique implemented by ShapeMatcher 5.2.1, which also was used to generate the graphs [37].

5. **Scene Gist** Although Scene Gist [41] is specially designed for recognition of scenes rather than objects, we included this model partly as a control for our assumptions about object representation and partly to explore whether global image encoding methods are applicable to biological object perception. In the Scene Gist model, each picture is represented as a weighted sum of bases, found through principal component analysis such that a small number of bases can be added together to reconstruct natural scene images with low error. The weights used in summing the bases to reconstruct an image serve as the features.

   A scene gist feature vector for each image was computed using Matlab code implemented by Torralba [67], and normalized to sum to 1. The distance between each image pair was calculated as the KL divergence between the corresponding normalized feature vectors.

After defining the distance metrics and calculating the representational dissimilarity matrix (RDM) entries for each of the five models, the resultant matrix for each model was compared to the matrix for each searchlight volume by converting the lower triangle of each 60×60 matrix into a 1770×1 vector and measuring correlations. When a model represents a set of image pairs as similar and a voxel sphere encodes the same pairs of images as similar, we may consider the voxels to be selective for the visual properties captured in the model. By comparing each computational representation with searchlights swept across the whole brain, we can identify which cortical regions, if any, have responses well described by each method's object/image representational approach.

Statistical significance values were computed at each searchlight location through permutation tests. The elements of the vectorized computer vision RDMs were permuted 500 times; the mean and variance of correlations for each searchlight position with each permuted RDM were computed to derive z values for the true correlation measures. The z values were converted into p values and a threshold was chosen such that the false detection rate was $q \leq .001$, following the method of Genovese et al. [15], and the regions above threshold were visualized over the subjects' anatomical surfaces. Surface maps were constructed using FreeSurfer [1] and SUMA [51].

2.3 Results

Our study was designed to illuminate how the human visual system encodes object information along the ventral pathway and, in particular, explicate the nature of intermediate neural object representations. To that end, we employed five computational models that make specific, and different, assumptions about the algorithms for recognizing content in visual images (Sec. 2.2.7). To the extent that there is a gap in our knowledge with respect to the nature of intermediate features in human vision, we adopted these models as proxy theories that each provide differing constraints on possible representations. Individual models were compared to our fMRI data by measuring the distance, or *representational dissimilarity*, between each pair of object stimuli for both the particular computational model and the neural encoding. A searchlight method was used to identify brain regions where the set of inter-stimulus distances, that is, the grouping of the stimuli, was similar to the grouping of the same stimuli produced by a given computational representation. Of note, in comparison to the limited functional regions identified by the LOC localization technique discussed in Sec. 2.2.3, we searched almost the entire brain to allow for the existence of brain regions selective for complex visual features beyond those regions often associated with object representation.

Given that all five of our included models rely on the same visual input as our fMRI experi-

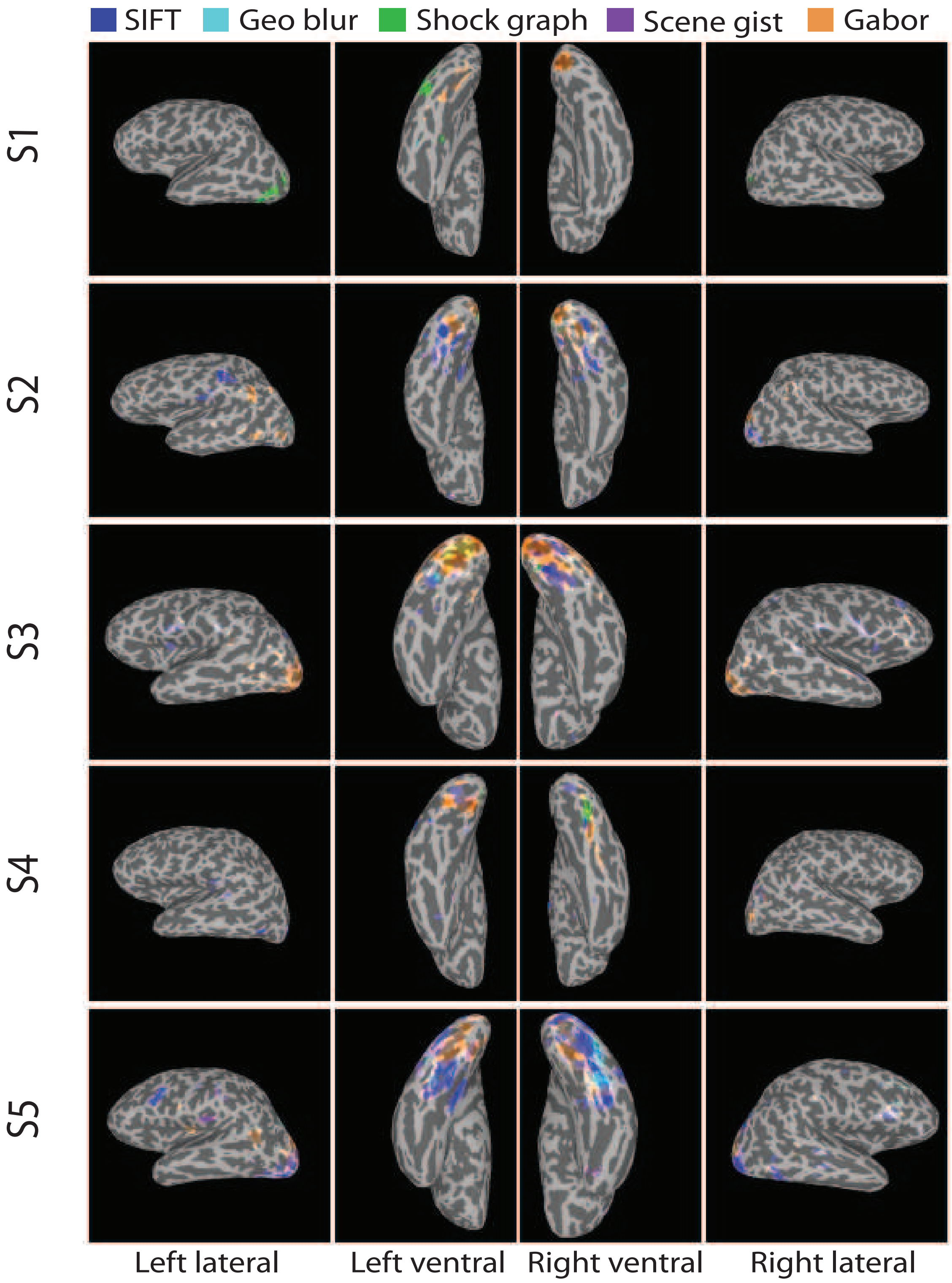


Figure 2.4: Cortical regions with a dissimilarity structure significantly correlated, $q < .001$, with the dissimiliarity structures of the five different models of visual feature coding. Colors are associated as follows: blue for *SIFT*, cyan for *Geometric Blur*, green for *Shock Graph*, purple for *Scene Gist*, and orange for *Gabor filterbank*. Color intensity proportional to correlation. Regions matching multiple models show the corresponding colors overlayed. Note first that although we illustrate these results on surface maps, the actual searchlights were run on brain volumes, and second, that color overlap sometimes forms misleading shades, for example, purple as the combination of blue and orange. Compare with Fig. 2.5 in cases of uncertainty.

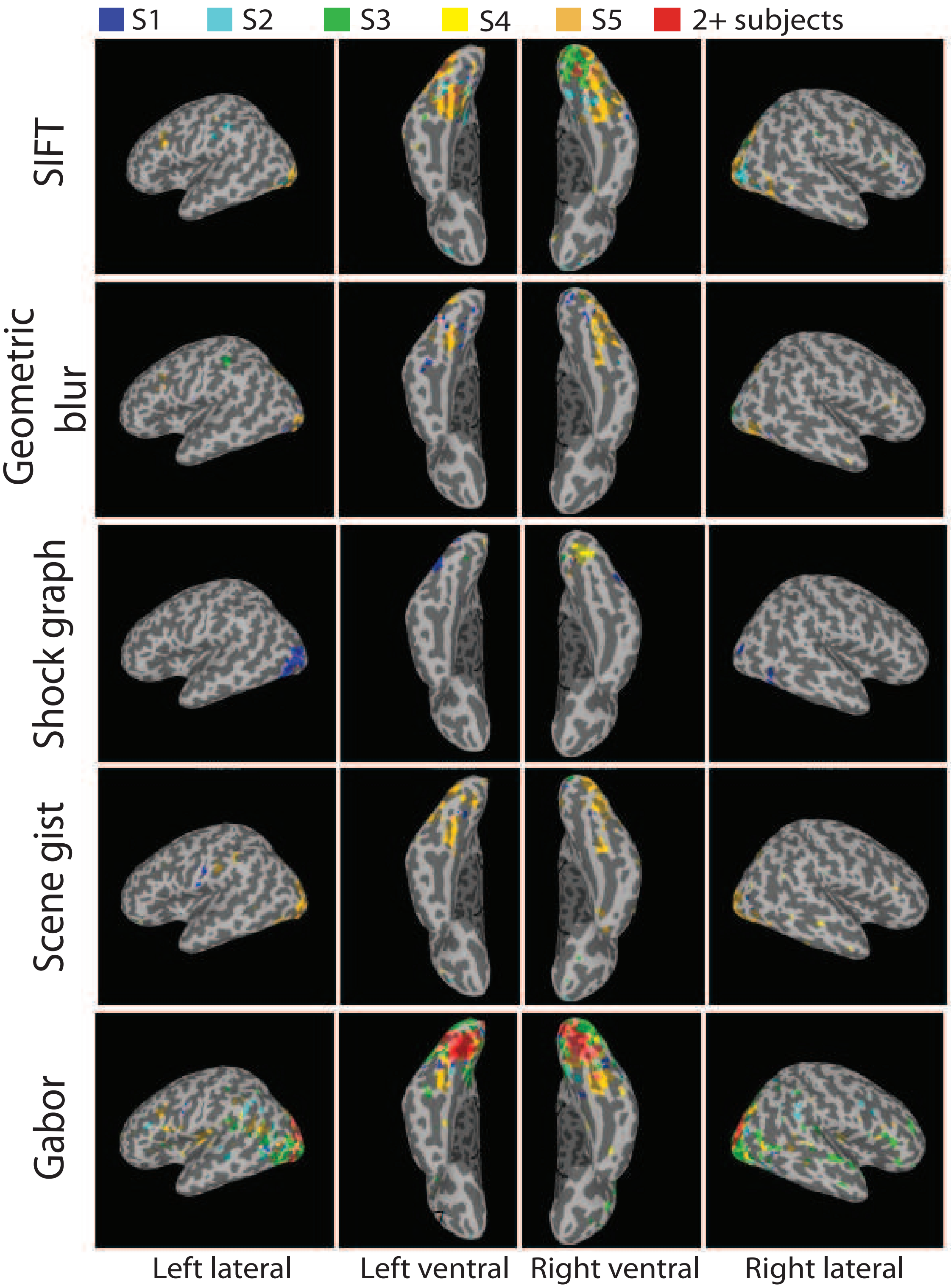


Figure 2.5: Cortical regions on Talairach brain with dissimilarity structure significantly correlated, $q < .001$, with the structures of computer visual models. Colors are associated with subjects as follows: blue for S1, cyan for S2, green for S3, yellow for S4, and orange for S5. Red denotes overlap between two or more subjects, with darker shades of red corresponding to increasing numbers of subjects overlapping with one another.

ment, it is not surprising, but still gratifying, that we observe significant correlations between our neural data and all five models. Fig. 2.4 depicts those brain areas with significant correlations ($q < .001$) between the distance matrices derived from each model and the neural responses within each area. Importantly, although we scanned across almost the entire brain, these correlated brain areas are focused in anatomical locations associated with low-, mid-, and high-level vision in both dorsal and ventral visual cortices, with limited spread to prefrontal cortex. Overall, the SIFT model most consistently matched the obtained stimulus representations in mid-level visual areas, while the Gabor filterbank model most consistently matched the obtained stimulus representations in low-level visual areas. The neuroanatomical locations for matches to the three other models were less consistent across subjects.

If we consider the underlying characteristics of each model, these results appear reasonable. First, the Gabor filterbank model encodes local oriented edges and has been used successfully to model receptive fields in early visual cortex [27]. Thus, the distance matrix correlations resulting from the Gabor filterbank model serve as a baseline to ensure that our overall approach is viable. As such, we expected a significant match between the activity observed in human V1 and this model. Moreover, including the Gabor filterbank model allows us to contrast these baseline correlations expected to be associated with earlier visual processing with any observed correlations arising in mid- and high-level visual areas. As illustrated in Fig. 2.4 in orange, S2, S3, and S5 all show a positive correlation between the RDMs from the Gabor filterbank model and neural activity in the left occipital pole, while all five subjects show a positive correlation in the right occipital pole. Somewhat surprisingly, the Gabor filterbank model also elicits significant positive correlations in mid-level visual areas, including the left fusiform (lFus) in all five subjects and the right fusiform (rFus) in subjects S2, S3, S4, and S5; subjects S2, S3, and S5 also exhibit positive correlations in left lateral occipital cortex (LO). We also see some correlation in anatomical regions often associated with higher-level visual processing, for example extending more anteriorly in the ventral temporal cortex for S1, S4, and S5. Finally, the Gabor filterbank model is correlated with activity beyond the ventral stream, including the inferior parietal (IP) region in the left hemisphere of S2, S3, and S4, and in the right hemisphere of S2; somewhat smaller correlations were also observed in left prefrontal cortex (PFC) of S2 and right PFC of S3 and S5. Least intuitive may be the small-area, weak correlation matches in left pre-central sulcus of S3 and S5. Fig. 2.5 emphasizes the most consistent match regions across subjects are in the bilateral occipital poles and early ventral stream.

In constrast with the Gabor filterbank model, the SIFT model encodes local visual statistics selected across points of interest in an image. The more restricted results observed for the SIFT model are consistent with this difference in representation. Positive correlations between the SIFT model and regions of neural activity are evident in subjects S2, S3, S4, and S5, as illustrated in Fig. 2.4 in blue. With respect to the SIFT model, our major finding is that these four subjects all show positive correlations in bilateral Fusiform. Subject S5 also shows a positive correlation in bilateral LO. In the dorsal stream, there is strong positive correlation for S2 in left IP. We also observed a positive correlation in left PFC for S5 and right PFC for S2 and S5. Fig. 2.5 illustrates the overlap of positively correlated regions across subjects in bilateral Fusiform and in the posterior right ventral stream.

The Geometric Blur model, much like SIFT, encodes local visual properties from selected points in each image, but also encodes more global information about object geometry. As illustrated in cyan in Fig. 2.4, all five subjects showed positive correlations with neural activity in mid-level visual areas; the breakdown by subjects being illustrated in Fig. 2.5. Subjects S1 and S5 exhibited positive correlations spanning bilateral Fusiform and posterior IT (pIT), with S5 exhibiting a more continuous region. More anteriorly in right IT, we observed spatially smaller positive correlation for S1 and S4. The right occipital pole also had small spatial regions showing positive correlations for S1, S2, S3, and S5, in addition to regions near the left occipital pole for S1 and S5. Within the ventral visual cortex, S5 also shows a positive correlation in bilateral LO. In the dorsal stream, there are small positive correlated areas in the parieto-occipital sulcus (POS) for S2. Finally, we observed a positive correlation in PFC for S5.

The Shock Graph model uniquely represents the silhouette shape of a given visual object, ignoring small-scale internal details critical to more local models such as SIFT and geometric blur. Positive correlations between neural activity and the Shock Graph model are illustrated in green in Fig. 2.4. These positive correlations are apparent for subjects S1, S3, S4, and S5. S1 exhibits positive correlations in bilateral LO and bilateral occipital poles. There are positive correlations for S3, S4, and S5 in rFus, as illustrated in Fig. 2.5.

The Scene Gist model encodes global image properties most commonly found in natural scenes, focusing on the two-dimensional spectrum across a given image. Positive correlations for the Scene Gist model are shown in purple in Fig. 2.4, with the most robust results being observed in S5, although, as illustrated in Fig. 2.5, there are also positive correlations in S1, S3, and S4. More specifically, S1 and S5 exhibit positive correlations in lFus. S5 also shows positive correlations in rFus, bilateral LO, and the bilateral pIT. S3 and S5 show positive correlations in the right occipital pole, with S5 also showing a positive correlation in the left temporal pole. Less robust effects are seen for S4 and S5 in a more anterior region of right IT; while S1 and S5 show positive correlations near left IP.

Taking a somewhat broader perspective, comparisons among these results indicate that some brain regions appear to consistently correlate with several of the computational models we considered. First, the Geometric Blur and SIFT models, both encoding local statistics of images, have overlapping regions on the ventral surfaces of S3 and S5 and in PFC of S5. Within the ventral surface, these regions tend to be in pIT. The greatest degree of overlap can be seen between SIFT and the Gabor filterbank model across subjects S2, S3, S4, and S5, largely along the posterior ventral surface. To some extent, this may be expected simply by chance, as these two methods produce the largest sets of model-cortical match regions. It also is worth noting SIFT is based on non-linear operations on selected Gabor filter responses, potentially tying the two methods together.

Another way of examining this data involves focusing on a specific functional region — in this case the area of the ventral stream most often associated with generic, high-level object processing — the LOC [16, 30]. Overlap between cortical regions differentially selective for

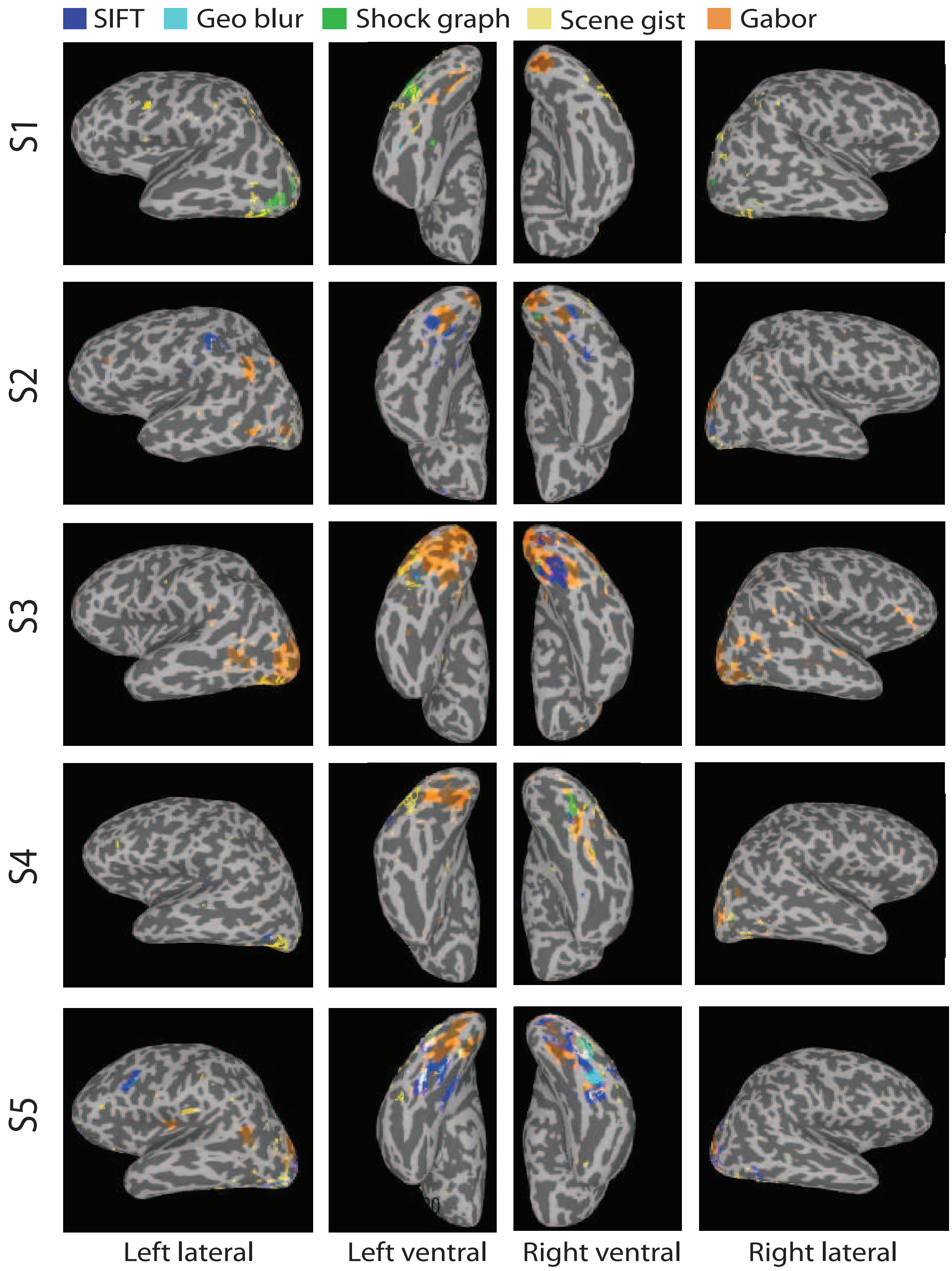
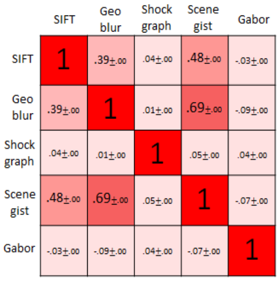


Figure 2.6: Cortical regions selected by LOC localizer and also found to have dissimilarity structure significantly correlated, $q < .001$, with the structures of computer vision models. Colors are associated as follows: blue for *SIFT*, cyan for *Geometric Blur*, green for *Shock Graph*, yellow for *Scene Gist*, orange for *Gabor filterbank*. Yellow countours show LOC localized regions.



	SIFT	Geo blur	Shock graph	Scene gist	Gabor
SIFT	1	.39±.00	.04±.00	.48±.00	-.03±.00
Geo blur	.39±.00	1	.01±.00	.69±.00	-.09±.00
Shock graph	.04±.00	.01±.00	1	.05±.00	.04±.00
Scene gist	.48±.00	.69±.00	.05±.00	1	-.07±.00
Gabor	-.03±.00	-.09±.00	.04±.00	-.07±.00	1

Figure 2.7: Distance matrix Spearman correlations among the five models. Mean and standard deviation correlations computed using leave-one-out method, leaving out 1 of the 60 stimuli for the distance matrices. Higher correlations in larger font and in darker red backgrounds.

objects, identified using the LOC "localizer" described above, and searchlight volumes found to be positively correlated with one or more of the five computational models are illustrated in Fig. 2.6. These overlap regions were spatially small as compared to the overall volumes identified by the searchlight process and varied in anatomical location depending on the particular computational model and the subject. For example, within the LOC, the anatomically-based left LO overlapped with a volume identified as correlated with the Gabor filterbank model in S3, while the lFus showed overlap with volumes associated with the Gabor filterbank model in S4. Further overlap within LOC was observed for Gabor filterbank volumes located in right pIT for S4, in a more anterior region of left IT for S1, and in left extrastriate cortex for S3 and S5. With respect to correlated searchlight volumes arising from the SIFT and Geometric Blur models, within LOC we observed overlap in right LO, pIT and more anterior IT for S5. Finally, the Geometric Blur model overlapped with LOC responses in anterior IT for S1.

To provide perspective on the similarities among the five studied computational models, we compared their respective stimulus distance matrices in Fig. 2.7. We compute correlations for distance matrices including 59 of the 60 rows and columns and observe the average and standard deviation for each model comparison. We observe that the correlation between models' stimulus grouping structures generally fails to act as a predictor of overlapping regions seen in Fig. 2.4, with the potential exception of the link between SIFT and Geometric Blur. Fig. 2.7 also illustrates that the models have notably low pairwise correlations, that is, representations, of the 60 stimuli.

Supporting this observation, for the most part, there are few overlapping regions across models in any of the five subjects.

A distribution of the model-neural activity positive correlation values, akin to a Gamma distribution, is above the FDR threshold for each subject and for each model. The nature of these distributions is illustrated in Fig. 2.8. Note that while the average significant correlations for each model are roughly the same, $r = 0.15$, the highest values provide a sense of ranking among computational representations in their abilities to account for neural responses. Most intuitively, the Gabor filterbank model, assumed to account for aspects of processing within primary visual cortex, shows the strongest matches with an average top correlation of roughly $r = 0.33$; analysis of individual subject correlations reveals the same pattern. SIFT exhibits the second highest set of correlations, with an average top correlation of roughly $r = 0.23$. The distribution of maximum correlations follows the same trend as the total area across all of the positively correlated regions for each model across all subjects; this is shown in Figs. 2.4 and 2.5. Fig. 2.8 also illustrates that there are significant positive correlations between every subject and every model. Certain matches are omitted from the discussion above because of their low correlations and their small surface spans, making them difficult to interpret.

2.4 Discussion

2.4.1 Computational models of vision as proxy theories of biological vision

Our goal in this project was to better elucidate the featural character of the ventral neural substrates supporting visual object processing. In contrast to our understanding of early visual processing (e.g., V1 through V4) and the high-level organization of visual cortex (e.g., the LOC, FFA, PPA, etc.), intermediate representation along the ventral pathway is poorly understood. To the extent that few theories account for this stage of visual object processing, we adopted a collection of theories drawn from computer vision to serve as proxies in that each theory makes

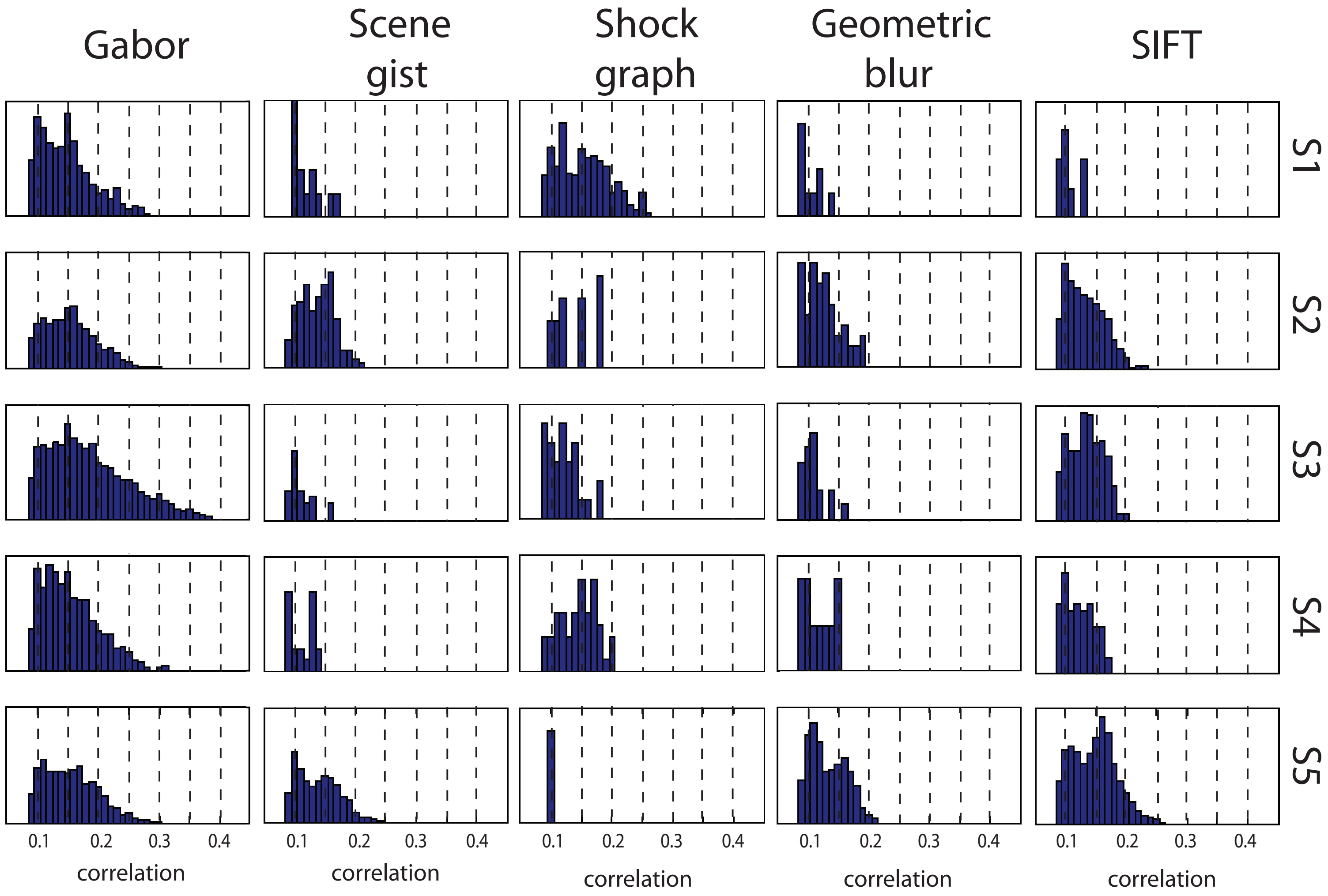


Figure 2.8: Histograms of significant correlations between model and searchlight RDMs.

specific, and different, assumptions regarding object representation.

To apply these theories to the neural representation of objects, we analyzed the pattern of similarity relationships between objects within the same collection of 60 objects as represented within the brain using fMRI and within each computational model. We then applied a searchlight analysis to uncover correlations between patterns of neural activity within brain subregions — sampled across the brain — and patterns within each computational model. This approach provided many regions where there was a reasonable correspondence between a given model and the observed neural activity. Importantly, almost all of these significant correlations occurred in brain areas associated with visual object processing, thereby providing a theoretical sanity check that our results are informative with respect to our question of interest. At one level, this general result should not be particularly surprising — all of our models relied on the same spatial input, images of objects, that were used as stimuli in the neuroimaging component of our study. Ideally, correlations at input should be reflected, at least to some degree, in correlations in representation of that input. On the other hand, the tested models each captured somewhat different linear and non-linear structures in their representation of objects (*e.g.*, [2, 8]). For example, the interest point frameworks used in the SIFT and Geometric Blur models provide a potential basis for parts-based perception — often assumed to be a critical element in the biological representation of objects [54, 79]. In contrast, the Shock Graph approach provides a compact encoding of an object's silhouette, supporting a parametric description of holistic representation [29]. Finally, Scene Gist is even more biased in representing global properties of an image, encoding the entire image structure of an object as well as the background [41].

Beyond the basic finding that our highest model-neural response correlations are observed within the visual system, we gain further confidence regarding the informativeness of our method from the observation that the strongest correlations between the Gabor filterbank model and neural activity are located early in the visual pathway, near the orbital pole and extrastriate cortex. This finding is consistent with a wide variety of studies characterizing early neural visual receptive fields as coding for local oriented edges [10, 22, 27]. The extension of these significant correlation regions into the higher-level bilateral fusiform and inferiorparietal has slightly less clear interpretations, but may support the hypothesis of Serre et al. [57] and Cadieu et al. [4] that later stages of the ventral visual stream employ a hierarchy of sometimes non-linear operations based on initial Gabor filter outputs. Beyond the operations specified in Serre et al. and Cadieu et al., SIFT represents a reframing of Gabor filter-like outputs for more complex recognition tasks, potentially accounting for the overlap in brain regions we observe between the correlations for the Gabor filterbank and SIFT models across subjects.

In summarizing the relative performance of the tested models, we find that both across and within subjects, the SIFT model appears to be the most promising of those tested for accounting for intermediate-level object representation in the human visual system. In particular, the SIFT model most strongly and consistently matched patterns of neural activity in rFus — an anatomical neighborhood associated with processing faces and other objects drawn from domains of expertise [17, 20, 65]. To a lesser extent, we also observed correlations for the SIFT model within left LO — a neuroanatomically-defined brain region also associated with object perception [16]. However, as shown in Fig. 2.6, the SIFT model rarely correlates with brain regions lying within the functionally-defined object-selective area referred to as LOC. Thus, it appears that the representation of objects in SIFT is similar to an intermediate encoding stage along the path to high-level object representation.

As a "proxy" model of intermediate feature representation, the preponderance of significant SIFT correlations in our results invites further reflection on its underlying algorithm. As discussed earlier, SIFT's interest point strategy is consistent with parts- or feature-based models of object perception. Notably, unlike Geometric Blur, our implementation of SIFT disregards the spatial locations of the local image regions it encodes, a characteristic that is consistent with the observation of invariance between intact images and their blockwise scrambled versions [71]. Similarly, SIFT incorporates aspects of the Gabor filterbank model which does a reasonable job at capturing characteristics of early visual processing; as such, this component of SIFT enhances its nominal biological plausibility. Finally, our "bag of words" implementation of the SIFT model [40] supports the learning of commonly-occurring local edge patterns as “visual words” — the use of such words allows the extraction of statistical patterns in the input similar to how vision scientists often characterize V1 receptive fields [42].

Our results also suggest that the Shock Graph model may be informative with respect to intermediate feature representation. Shock graphs describe objects in terms of their global shapes, capturing their axial structures and silhouettes. Thus, spatial information about the relative positions of shape features are preserved, but the local image statistics that may specify local features are not captured (e.g., texture). Our observation of correlations between ventral stream neural activity and the Shock Graph model supports the idea underlying shape-based encoding in intermediate-level neural representations [24, 63, 78]. To the extent that these correlations are confined to more posterior parts of the ventral stream, they are, however, somewhat inconsistent with Hung et al.’s [24] observation of shape-based representations in anterior IT in monkeys. At the same time, this observation should not be generalized to other models of global encoding, as we find that Scene Gist, encoding spatial frequencies across whole images, produces correlations in more anterior IT.

More generally, although our results are informative in some respects, it is doubtful that any established computational vision model accurately captures the neural representations instantiated in intermediate-level biological vision. Indeed, the best correlations between any model and the fMRI-derived cortical distance matrices (Fig. 2.8) fall below the majority of pairwise correlations observed between the model-derived distance matrices (Fig. 2.7). Nonetheless, the large majority of statistically significant ($q < .001$) model-fMRI correlations were found in visual brain areas, with some differentiation within these areas for different methods. Thus, we gain some sense of the properties for which given brain regions may be selective.

From a theoretical perspective, one potential concern with this interpretation is how we selected particular computational models for use in our study. In large part, our choices were based on each model’s success or popularity in the extant computational vision literature and on each model’s distinct encoding strategy with respect to intermediate feature representation —

an intuition validated by the fact that the models have measurably different stimulus dissimilarity matrices (Fig. 2.7). Of note, our present work does not include an analysis of the popular hiearchical model of biological vision known as “HMAX” [4, 48, 57]. HMAX employs a hierarchical series of summing and non-linear pooling operations to model mid-level visual regions such as V2 and V4. However, the HMAX model contains a variety of variables that must be fit either to the input stimulus set or to a set of experimental data [57]. In an additional experiment not presented here, we found the actual data set collected in our study using the 60 image stimuli was insufficient for reliable fitting of HMAX [56], even when limiting the model to layers S1 through C2, as in Cadieu et al. [4]. In contrast, the application of HMAX to the responses of individual neurons in monkeys [4] is more feasible, as data for 1,000s of trials can be acquired. At the same time, it is worth noting that neurophysiological recordings of IT do not correspond to HMAX predictions for stimulus grouping structure [33].

From an empirical perspective, a second potential concern is the degree of variability in the spatial location, or even the existence, of large high-correlation brain regions for each model within individual subjects. In some cases, as in SIFT and Gabor filterbank, the changes in anatomical positions across subjects were relatively slight, consistent with variability of functional region locations, such as LOC or FFA [34]. More qualitative variability, for example, across lobes or hemispheres, may reflect meaningful differences in our subjects’ cognitive and cortical approaches to object perception. For example, individuals may vary in the degree to which they attend to local versus global features or apply holistic mechanisms [77]. Beyond the potential strategic variation in how individuals perceive objects, noise in the hemodynamic signals may increase the variability of correlated brain regions across subjects. However, this latter possibility fails to explain why all subjects exhibit significant and consistent correlations within the visual pathway for several of the models.

2.5 Conclusion

Our study aims to connect the cortical encoding in mid- and high-level visual areas of human ventral stream and image representations as realized in several different computational models of object representation. Perhaps the most salient conclusion we can make is that the best biological/computational correspondence is observed for the Scale Invariant Feature Transform ("SIFT") model [36]. Although such results do not imply that SIFT-like computations are somehow realized in the human brain, they do suggest that the SIFT model captures some aspects of the visual stimulus that are likewise instantiated in human visual object processing. As this is one of the first attempts to directly connect extant computational models of object representation with the neural encoding of objects, there remains ample room to sharpen our observations and to further explore the space of possible biological vision representations. For example, the passive viewing task used in the neuroimaging component of our study could be replaced by an active object identification task, which, conceivably, might yield stronger neural signals and more robust results. Likewise, other computational vision models should be considered, for example, histograms of oriented gradients [9], the more biologically-inspired HMAX model (given that we first solve the problem of limited data using fMRI), or the biologically-motivated and hierarchical model described in Jarrett et al. [25]. In particular, SIFT's similarity to HMAX — both models rely on non-linearities to pool local edge information — indicates further pursuit of HMAX to describe high-level voxel encodings may prove fruitful course for future research. Finally, a more sophisticated approach to developing model-brain correspondences may be realized by combining the dissimilarity matrices for any group of representational methods with weights optimally learned to match the representation at any given brain region [66]. In sum, our present study provides a foundation for further exploration of well-defined quantitative models using dissimilarity analyses and points the way to methods that may help shed further light on the visual structures encoded in the human brain. I discuss my further exploration through realtime search in the following chapters.

Chapter 3

Methods

In my present work, I study cortical object perception by searching for the complex visual properties most activating pre-selected cortical regions of interest in the ventral pathway. Employing fMRI for my investigation, there are a limited number of stimulus display trials available to probe regional selectivities — roughly 300 displays per hour sampling from a near-infinite space of visual object properties. I develop, use, and assess novel methods to efficiently search the "space" of visual object stimuli to quickly identify those stimuli evoking the highest response from a pre-selected cortical region. These methods analyze fMRI signals in realtime to determine region responses to recently-displayed stimuli, and use these recent cortical responses to select new stimuli likely to produce higher activity.

I employ two sets of object stimuli and two corresponding definitions of visual properties to explore intermediate representations in the ventral pathway. The first stimulus set consists of photographs of real-world objects, assessing cortical perception of visual properties using images that can be encountered in ordinary life experience; these objects are characterized by a Euclidean feature space derived from the SIFT method [36], found to account for object representations in intermediate regions of the ventral pathway in Chap. 2. The second stimulus set consists of synthetic Fribble objects [76] constructed from simple textured shapes, providing careful control on the varying properties displayed to subjects; these objects are characterized by a Euclidean feature space in which each axis captures the degree of controlled manipulation to a corresponding component shape.

My work explores the effectiveness of my novel methods in realtime fMRI analysis and dynamic selection of stimuli, and my work uses these methods to explore the complex selectivities of intermediate regions in the cortical object perception pathway.

3.1 Realtime computational strategy and system architecture

In fMRI studies of human subjects, scanning time is limited to several hours across several days. During a given scan session, the slow evolution of the blood-flow dependent fMRI signal limits the frequency of stimulus displays to one display every 8 to 10 seconds. While number of display trials is small, the number of potential visual objects to show as stimuli is nearly infinite. Therefore, I develop, use, and assess methods for the dynamic selection of stimuli, choosing new images to display based on the response of the pre-selected brain region to previous images to try to maximize regional activity and to identify the associated complex featural selectivity. This approach effectively is a search through a stimulus space. The search requires realtime fMRI signal processing and analysis using an array of computer programs that execute in parallel and that interact with one another. Each brain region for study is chosen, or “pre-selected,” prior to the realtime analysis scanning session discussed here. Regions are chosen for each subject based their representation of visual objects as seen from data collected from a prior scanning session for the subject, as discussed in Secs. 3.3.6 and 3.4.5.

Three programs run in parallel throughout the realtime search for stimuli producing maximal regional activity:

- The **display program** sends visual stimuli to the display screen for subject to view while lying in the scanner.
- The **preprocessing program** converts recently-recorded raw full-volume scanner output into a single number corresponding to the response of a pre-selected cortical region to the

recently-displayed stimulus.

- The **search program** uses past stimulus responses, computed by the preprocessing program, to select the next stimuli to show the subject, to be sent to the screen by the display program. The stimuli are selected through a simplex "search" of visual feature space for the location/stimulus producing the highest response from a cortical region, described in Sec. 3.1.5.

3.1.1 Interleaving searches

To use scanning time most efficiently, four searches are performed studying four pre-selected brain regions during each scan. After a stimulus first appears in front of the subject in the scanner, 10–14 s[1] is required to gather the 10 s cortical response to the stimulus and an additional ˜10 s is required to process the response and to select the next stimulus for display. Before the next stimulus for a given search has been selected, the display program can rotate to another search, maximizing the use of limited scan time to study multiple brain regions. The display and analysis programs alternate in sequence among the four searches — i.e., `search 1` → `search 2` → `search 3` → `search 4` → `search 1` $\cdots$. Different classes of real-world and Fribble objects are employed for each of the four searches, as described in Secs. 3.3.1 and 3.4.1. Alternation among visually distinct classes is further advantageous to my study as it decreases the risk of cortical adaptation present if multiple similar stimuli were viewed in direct succession.

The preprocessing program evaluates cortical responses in blocks of two searches at a time — i.e., it waits to collect data from the current stimulus displays for `search 1` and `search 2`, analyzes the block of data, waits to collect data from the current stimulus displays for `search 3` and `search 4`, analyzes this block of data, and so on. This grouping of stimulus responses increases overall analysis speed. Several steps of preprocessing require the execution of AFNI [46] command-line functions. Computation time is expended to initialize and terminate each func-

[1] The 4 s beyond the duration of the cortical response accounts for communication delay between the fMRI scanner and the machine running the preprocessing and search programs.

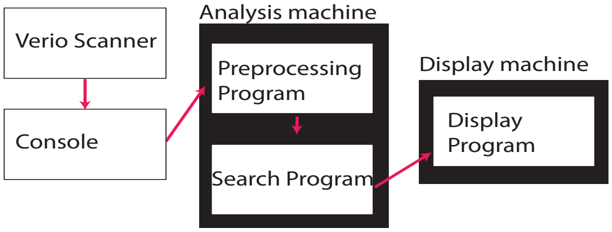


Figure 3.1: Diagram of communications between the console (which collects and sends fMRI data from the scanner), the "analysis machine," and the "display machine," as well as communications between the analysis programs. These elements work together to analyze cortical responses to object stimuli in realtime, select new stimuli to show the subject, and display the new stimuli to the subject.

tion each time it is called, independent of the time required for data analysis. By applying each function to data from two searches together, the "non-analysis" time across function calls is decreased.

3.1.2 Inter-program communication

Three programs run throughout each realtime search to permit dynamic selection and display of new stimuli most effectively probing the visual selectivity of a chosen cortical region. The programs — focusing on fMRI preprocessing, visual property search, and stimulus display tasks, respectively — are written and executed separately to more easily track, test, and debug each process and to more easily permit implementation and application of alternate approaches to each task. Furthermore, the display program runs on a separate machine from the other two processes, shown in Fig. 3.1, to ensure sufficient processor resources are dedicated to each task, particularly as analysis and display computations must occur simultaneously throughout each scan.

Due to the separation of tasks into three programs, each task relies on information determined by another program or machine, as indicated in Fig. 3.1. Below, I discuss the methods used to communicate information necessary for preprocessing, search, and stimulus display.

- **Preprocessing program input** The scanner console machine receives brain volumes from the fMRI scanner. I reconfigure console storage protocols such that each received volume is copied in realtime to a mounted drive on the analysis machine. The analysis machine runs the preprocessing and search programs[2], using the newly-recorded fMRI data to determine the responses of pre-selected cortical regions to the most recently-displayed stimuli, and using the responses to select the next stimuli to display to the subject in the scanner to probe the visual selectivities of the regions. The preprocessing program checks the shared disk every 0.2 seconds to determine whether all the volumes for the newest block of search results — the full 10 s cortical responses to two recently-shown stimuli, described further in Sec. 3.1.1 — are available for analysis. Once all the data is available, the preprocessing program uses the data, as discussed in Sec. 3.1.4, to compute one numbers to represent the response of each of two pre-selected brain regions to their respective stimuli. The program proceeds to write each response into a file labeled `responseN` and then creates a second empty file named `semaphoreN`, where $N \in 1, 2, 3, 4$ in each file is the number of the search being processed. The files are written into a pre-determined directory that is monitored by the search program, so the search program can find information saved by the preprocessing program. The creation of the `semaphoreN` file signals to the search program that the response of the brain region studied in the N^{th} search has been written fully to disk. This approach prevents the search program from reading an incomplete or outdated `responseN` file and acting on incorrect information.

- **Search program input** The search program alternates among four searches for the visual feature selectivities of four brain regions, i.e., searching for the stimuli containing features producing the most activity from a pre-selected cortical region. At any given time during a realtime scan, the search program either is computing the next stimulus to display for a search whose most recent cortical response has recently been computed, or is waiting for the responses of the next block of two searches to be computed. While waiting, the

[2] These programs are run on a separate machine from the console to ensure sufficient processing power is available for realtime analysis.

search program checks the pre-determined directory every 0.2 seconds for the presence of the semaphore file of the current search, created by the preprocessing program. Once the search program finds this file, the program deletes the semaphore file and loads the relevant brain region's response from the response file. The search program proceeds to compute the next stimulus to display, intended to evoke a high response from the brain region, as discussed in Sec. 3.1.5, and sends the stimulus label to the display program running on the display machine.

- **Display program input** Two methods were used for the transmission of stimulus labels between the search and display programs. (The display program controls what is shown to the subject at every moment of the scanning sessions.) For the first five subjects, studied using real-world objects, the search program sent each label to the display program by saving it in a file, `rtMsgOutN`, in a directory of the analysis computer mounted by the display computer. Immediately prior to showing the stimulus for the current search $N \in \{1, 2, 3, 4\}$ — alternating between four searches, as do the preprocessing and search programs — the display program looked for the corresponding file in the mounted directory. For the remaining subjects — using either real-world or Fribble objects — labels were passed over an open socket from the Matlab [38] instance running the search program to the Matlab instance running the display program. In the socket communication, the search program paired each label with the number identifier N of the search for which it was computed. Immediately prior to showing the stimulus for any given current search, the display program read all available search stimulus updates from the socket until it found and processed the update for the current search and then showed the current stimulus to display for the current search. Ordinarily, both techniques allowed the display program to present the correct new stimulus for each new trial, based on the computations of the search program. However, when preprocessing and search computations did not complete before the time their results were required for a new stimulus display, the two communication techniques between the search and display programs had differing behaviors. Message passing through a file ensured there always was a label to be read and used by the display program at display time, but sometimes the available label had been produced by the search program for use in the previous search iteration. Message passing through a socket ensured stimulus display in the display program did not occur until data intended for the current search iteration were available; however, waiting for the new data sometimes caused significant delays in stimulus display — occasionally delays of greater than 20 seconds — and sometimes updates were not computed for a search iteration for a given class. I study the effects of delayed preprocessing and search results on overall realtime search performance in Chaps. 4 and 5.

3.1.3 Stimulus display

All stimuli were presented using Matlab [38] and the Psychophysics Toolbox [3, 44] controlled by an Apple Macintosh and were back projected onto a white screen located at the head end of the bore using a DLP projector (Sharp XG-P560W). Subjects viewed the images through a mirror attached to the head coil with object stimuli subtending a visual angle of approximately $8.3\,deg \times 8.3\,deg$. During the realtime search scans, each stimulus was displayed for 1 s followed by a centered fixation cross that remained displayed until the end of each 8 s trial, at which point the next trial began. The 8 s trial duration is chosen to be as short as possible while providing sufficient time for the realtime programs to compute and return the next stimuli to display based on the previous cortical responses. Further experimental design details are provided for each scan in Secs. 3.3.4 and 3.4.3.

3.1.4 Preprocessing

Functional images were acquired with a Siemens Verio scanner using a 2 s TR. Further fMRI procedures are provided in Sec. 3.2. The preprocessing program analyzed all brain images in two-trial blocks, corresponding to the cortical responses to stimuli for displays for two consecutive searches. The preprocessing program computed a representation of the response of preselected cortical regions to displayed stimuli. Each display trial had a duration of 8 s and the measured hemodynamic (blood-flow) response for each stimulus had a duration of 10 s. Thus, each block considered by the preprocessing program spanned 18 s (containing 9 volumes), except for the first block which also contained the first 6 s of baseline activity prior to the first stimulus onset, and thus spanned 24 s (and 12 volumes). Because of the disparity between 8 s trials and 10 s hemodynamic responses, there was a 2 s (one volume) overlap between each pair of consecutively processed blocks.

Scans in each data block were registered to the first volume of the current run and motion corrected using AFNI. Polynomial trends of orders one through three were removed. The data then were normalized for each voxel by subtracting the average and dividing by the standard deviation, obtained from the current data block and from a previous “reference” scan session (described in Secs. 3.3.4 and 3.4.3), respectively, to approximate zero-mean and unit variance [26]. The standard deviation was determined from ˜1 hour of recorded signal from a previous scan session to gain a more reliable estimate of signal variability in each voxel. Due to variations in baseline signal magnitude across and within scans, each voxel’s mean signal value required updating based on activity in each block. To allow multivariate analysis to exploit information present at high spatial frequencies, no spatial smoothing was performed [62].

Matlab was used to perform further processing on the fMRI time courses for the voxels in the cortical region of interest for the associated search. For each stimulus presentation, the measured response of each voxel consisted of five data samples starting 2 s/1 TR after onset, corresponding to the 10 s hemodynamic response duration. Each five-sample response was consolidated into a weighted sum by computing the dot product of the response and the average hemodynamic response function (HRF) for the associated region. The HRF was determined based on data from an initial “reference” scan session[3] performed before the realtime scanning sessions, as described in

[3] In the reference scan session, 36 object stimuli were displayed multiple times over an hour session, in addition

Secs. 3.3.4 and 3.4.3. The pattern of voxel responses across the region was consolidated further into a single scalar response value by computing a similar weighted sum. Like the HRF, the voxel weights were determined from reference scan data. The weights correspond to the most common multi-voxel pattern observed in the region during the earlier scan, i.e, the first principal component of the set of multi-voxel patterns. This projection of recorded realtime responses onto the first principal component treats the activity across the region as a single locally-distributed code, emphasizing voxels whose contributions to this code are most significant and de-emphasizing those voxels with ordinarily weak contributions to the average pattern.

3.1.5 Search program

The search program chooses the next stimulus to display to probe the selectivity of a pre-selected cortical region based on the region's responses to recently displayed stimuli. The search chooses the next stimulus by considering a space of visual properties and probing locations in this space (corresponding to stimuli with particular visual properties) to most-quickly identify locations that will elicit maximal activity from the brain region under study. The visual spaces searched for the first group of ten subjects and the second group of ten subjects are defined in Secs. 3.3.2 and 3.4.1, respectively. Each stimulus i that could be displayed is assigned a point in space p_i based on its visual properties. The measured response of the brain region to this stimulus r_i is understood as:

$$r_i = f(p_i) + \eta \tag{3.1}$$

i.e., a function f of the stimulus' visual properties as encoded by its location in the space plus a noise term η, drawn from a zero-centered Gaussian distribution. The process of displaying an image, recording the ensuing cortical activity via fMRI, and isolating the response of the brain region of interest using the preprocessing program I model as performing an evaluation under

to other images. After the session was completed, cortical responses were processed to determine cortical regions of interest for study in the ensuing realtime scans and to measure signal properties of the voxels in these regions.

noise of the function describing the region's response. By performing evaluations in strategic points in visual space, each corresponding to a stimulus image, I seek to identify the location of the function's maximum — equated with the visual property selectivity of the brain region.

For simplicity, I assume my selected region has a selectivity function f that reaches a maximum at a certain point in the visual space and falls off with increasing Euclidean distance from this point. I assume the visual feature space, defined in Secs. 3.3.2 and 3.4.1 for each of the two subject groups, reasonably captures the representation of visual objects for the given brain region because each region was selected based on its representational match with the corresponding space as reflected by data recorded during viewing of object stimuli in the subject's earlier reference scan session, described in Secs. 3.3.6 and 3.4.5. I also expect a considerable amount of noise to be added to the underlying selectivity-driven response signal computed during realtime scans. Under these assumptions, I use the simplex method [39] as the foundation of my approach to finding the optimal stimuli in the space. More specifically, the search program uses a modified version of the simplex simulated annealing Matlab code available from Donckels [11], implementing the algorithm from Cardoso et al. [7]. I incorporate measurement uncertainty through partial resets of the simplex and through random perturbations to the measured cortical responses with magnitude determined by simulated annealing.

The searches for stimuli producing maximal activity for each of the four pre-selected brain regions are performed over the course of a 1.5 hour scanning session. Because subjects require time to rest their eyes and their attention across the scan time, stimulus display runs and the underlying searches selecting the stimuli are limited to 8.5 minute periods. The simplex for each cortical region is re-established at the start of each new run. At the beginning of the scanning session, the starting location for the simplex for each of the four brain regions is set either to the origin or to another pre-determined point in visual property space, as discussed in Sec. 3.5.2. The starting locations for the simplexes for each ensuing display run, i.e., the i^{th} run, is set to be the simplex point that evoked the largest response from the associated cortical region in the $(i-1)^{th}$ run. At the start of the i^{th} display/search run, each simplex is initialized with the starting point

$x_{i,1}$, as defined above, and D further points, $x_{i,d+1} = x_{i,1} + U_d \, v_d$, where D is the dimensionality of the space, U_d is a scalar value drawn from a uniform distribution between -1 and 1, and v_d is a vector with d^{th} element 1 and all other elements 0. In other words, each initial simplex for each run consists of the initial point and, for each dimension of the space, an additional point randomly perturbed from the initial point only along that dimension. The redefinition of each simplex at the start of each new run constitutes a partial search reset to more-fully explore all corners of the feature space while retaining a hold on the location from the previous run appearing to produce the most activity from the selected region. For the remainder of this section, I will focus on simplex updates in the search for the selectivity of one cortical region, although updates for four simplexes occur in an alternating fashion throughout each run, as discussed above.

After determining the initial points of the simplex, the simplex method operates as follows, seeking to identify new points (corresponding to stimuli) that evoke the highest responses from the selected cortical region:

1. Evaluate function at all points in the simplex
2. Hypothesize new point in space that will produce higher functional response than do current points in simplex
3. Evaluate function at new point
4. Based on value at new point
    - Replace point with smallest functional response in simplex with new point which contains higher functional response, return to step 2
    - Select a different new point to compare with all points in simplex, return to step 3
    - Adjust locations of all points in simplex, return to step 1

The spatial location of each new point selected by the simplex method for testing is modified to be the location of the nearest stimulus available to be shown, as there are a limited number of potential stimuli to choose for the next display and an infinite combination of potential spatial coordinates. Each "function evaluation" is achieved by display of the stimulus associated with the point in space by the display program, recording the cortical response by the fMRI machine, and fMRI signal processing by the preprocessing program, as described above. Over the search time, I expect the simplex to contract and the member points to move towards the location whose properties elicit maximum response in the brain region under study.

The presence of noise in the recorded fMRI signal potentially can reduce the observed response of a brain region to one of its most preferred stimuli, causing the observed response to lie below the responses associated with points already in the search simplex and leading to an incorrect rejection of the the new preferred stimulus from the simplex. Similarly, the measured response of a non-preferred stimulus can be inflated by noise and improperly accepted into the simplex. To counteract these effects, as the search progresses, the measured brain region response to each new potential simplex point r_{new}, corresponding to a visual stimulus viewed by the subject, is perturbed by subtracting a random scalar,

$$r'_{new} = r_{new} - T \; |ln(rand)| \tag{3.2}$$

where $rand$ is a value drawn from the uniform distribution between 0 and 1, and T is a scaling "temperature" value discussed below; the measured brain responses of the points currently in the simplex are perturbed by *adding* similar random scalars,

$$r'_j = r_j + |T \; ln(rand)| \tag{3.3}$$

for the purpose of comparing the current simplex points with the potential new point. These adding and substracting operations have a conservative effect, limiting newly-accepted points to those for which the brain region appears to show dramatically greater selectivity than the points currently in the simplex. The scaling temperature value T decreases over search time based on the equation:

$$T_{new} = \frac{T_{old}}{1 + \frac{T_{old} ln(1+\delta)}{3\sigma}} \tag{3.4}$$

where δ is a pre-determined cooling rate parameter and σ is the standard deviation of all brain region response values so-far measured. A larger δ value results in faster cooling, and a larger σ value — reflecting less convergence in brain region responses — results in slower cooling. As the search progresses, it is expected the simplex will focus on points in an increasingly narrow region of visual feature space — an area producing particularly high responses from the set cortical region — and more complete exploration of the smaller space is favored over cautious acceptance and rejection of new simplex points. Decreasing the temperature causes less perturbation of cortical region responses at each point, in Eqns. 3.2 and 3.3, relaxing the criteria for replacing current points in the simplex with new points and allowing freer movement of the simplex in the space. The strategy of decreasing random perturbations over time, and the method for decreasing perturbations through Eqn. 3.4, constitutes a form of "simulated annealing." The temperature is set to decrease when the span of the simplex in the visual feature space has narrowed sufficiently [7, 11]. However, due to the limited number of trials in each realtime scanning run for each search — each search run completes after 15 "function evaluations," corresponding to the number of stimulus display trials assigned to each search for each scanning run, as discussed in Secs. 3.3.4 and 3.4.3 — the reduced span criterion never is met, and the temperature never is decreased, in my present study.

Simulations were used to find δ and initial T values to maximize the chance of correctly identifying the spatial neighborhood eliciting maximum neural response using the simplex simulated annealing realtime search over six scan runs, with 15 searches steps in each run, similar to the conditions of the actual realtime scan searches. Simulated brain responses were computed following Eqn. 3.1, using a pre-determined selectivity center in the feature space and applying Gaussian noise $\eta = \mathcal{N}(0, s)$; the standard deviation of the simulated noise s was selected based on the statistics of ventral pathway responses to object stimuli recorded in my previous study described in Chap. 2. As a result of the simulations, the values in Eqn. 3.4 are set as $\delta = 5$ and $T = 10$.

3.2 fMRI Procedures

Subjects were scanned using a 3.0 T Siemens Verio MRI scanner with a 32-channel head coil. Functional images were acquired with a gradient echo-planar imaging pulse sequence (TR 2 s, TE 26 ms, flip angle $90°$, $2\ mm \times 2\ mm \times 3\ mm$ voxels, field of view $192 \times 192\ mm^2$, 31 oblique-axial slices). Slices spanned the majority of the brain, to all the possibility of future study of visual stimulus representations beyond the visual streams. An MP-RAGE sequence (flip angle $9°$, $1\ mm^3$ voxels, field of view $256 \times 256\ mm^3$, 176 sagittal slices) was used for anatomical imaging.

3.3 Real-world object search

In searching for complex visual feature selectivities in the ventral stream, I begin with a focus on the perception of real-world objects with visual features represented by the scale invariant feature transform (SIFT, [36]). Use of photographs of real objects — such as statues, cows, cars, and suitcases — provides a more realistic understanding of cortical activity while a person is interacting with the real world, rather than interacting with an artificial world of simplistic blob and pattern stimuli employed by most studies of complex visual properties used by the brain [4, 24]. Unfortunately, the optimal representation to capture the visual structure of real-world objects, particular as it is perceived in intermediate regions of the cortical ventral pathway, is unclear. My recent work, discussed in Chap. 2, indicates a SIFT-based representation of visual objects is a strong candidate to match representations used by voxel regions in mid- and high-level vision areas in the brain. SIFT incorporates established low-level biological features — capturing local edge information at selected interest point — and performs non-linear synthesis of statistics across the full image. These computational properties have contributed to SIFT's general success on object recognition tasks in the field of computer vision, and contribute to its association with intermediate cortical visual representations. I represent real-world object stimuli through coordinates in a SIFT-based space and search through this space to identify visual selectivities of regions in the ventral object perception pathway.

3.3.1 Stimuli

Stimuli were drawn from a picture set comprised of 400 distinct color object photos displayed on 53% gray backgrounds (Fig. 3.2). The photographic images were taken from the Hemera Photo Objects dataset [21]. My use of real-world images of objects rather than the hand-drawn or computer-synthesized stimuli employed in the studies of mid-level visual coding discussed in Chap. 1, e.g., Cadieu et al. [4] and Yamane et al. [78], is intended to more accurately capture the importance of the broad set of naturally-occurring visual features to object perception.

Four separate searches were performed in each realtime analysis scanning session, probing the visual property selectivities of four distinct pre-selected brain regions. Each search drew from a distinct class of visual objects — mammals, human-forms, cars, and containers. The images in each class were manually selected from the Hemera dataset — automatic grouping of stimuli was not possible as there was insufficient semantic information included in the dataset to assemble a class of sufficiently large size. The four manually-assembled classes varied in size from containing 68 to 150 objects.

The focus of each search within an object class limited visual variability across stimuli in the search. The remaining sources of variability, I hoped, would be relatively intuitively identified and easily associated with their influence on the magnitude of cortical region activity. Unfortunately, these hopes were not frequently fulfilled in our results, discussed in Chaps. 4 and 5. The cortical region for each search was selected based on the region's differentially high activation when the subject viewed objects within the search class, as reflected by data recorded from an

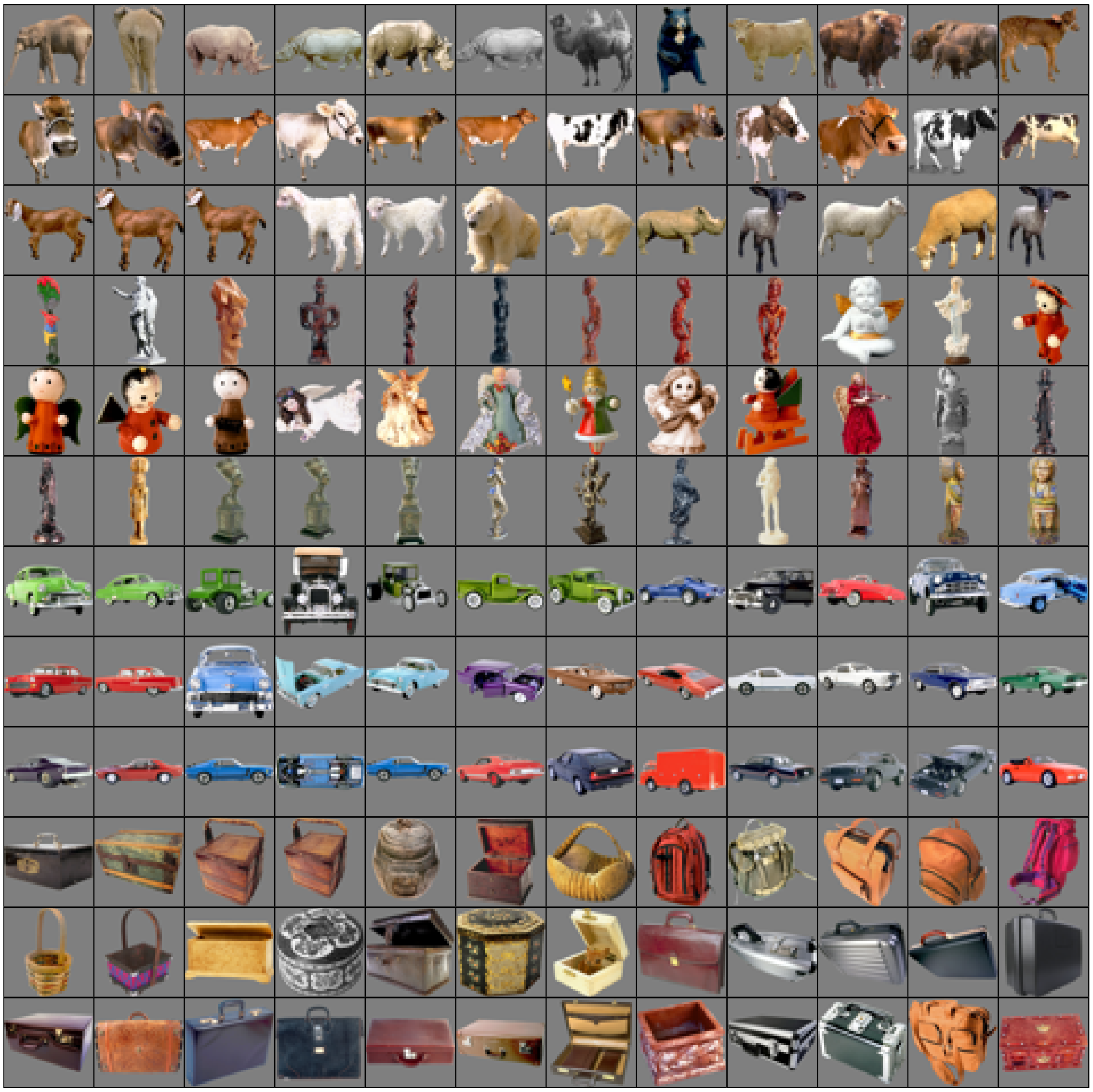


Figure 3.2: Example stimuli used in realtime search of real-world objects. Images were selected from four classes of objects — mammals, human-forms, cars, and containers.

earlier reference scan session. Use of a narrow object class to probe a region selective for the same class also was intended to produce strong cortical signal for analysis during the search scans, minimizing the effects of noise when computing the next stimulus to display based on regional response to the most recent stimuli.

3.3.2 Stimulus "SIFT" space

To identify visual properties producing highest activity in a pre-selected brain region, my simplex simulated annealing search program requires a Euclidean search space containing the object stimuli to display. While my recent work supports the use of SIFT as a model of representing visual objects in a manner similar to the grouping performed by voxel representations in intermediate regions of the ventral pathway, the SIFT measure of Leeds et al. does not directly generate a space that is easily searched. Entries to the pairwise distance matrix — the "representational dissimilarity matrix" — for pairs of object stimuli are computed based on non-linear Kullback-Leibler divergence comparison between histogram of visual words [35]. In my present work, I define a Euclidean space based on the distance matrix using Matlab's implementation of metric multidimensional scaling (MDS) [55]. MDS finds a space in which the original pairwise distances between data points — i.e., SIFT distances between stimuli — are maximally preserved for any given n dimensions.

Starting with a SIFT-based distance matrix for 1000 Hemera photo objects, the MDS method produced a space containing over 600 dimensions. Unfortunately, as the number of dimensions in a search space increases, the sparsity of data in the space can increase exponentially, making a clear conclusion about the underlying selectivity function increasingly more uncertain absent of further search constraints. Furthermore, the standard simplex method's average search time grows polynomially with the number of dimensions, and exponentially in the worst case [52], which poses a significant problem given limited subject time in the scanner. Therefore, I seek to use a small number of the most-representative dimensions for realtime search. Fig. 3.3 shows the

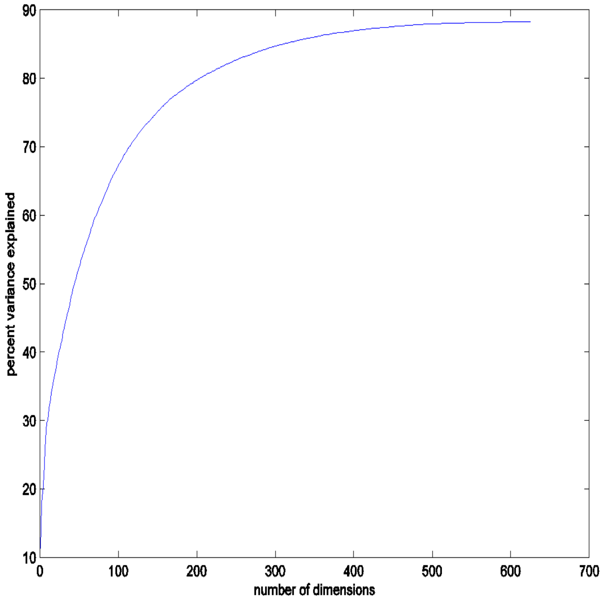


Figure 3.3: Percent variance explained of SIFT pairwise distance structure by multi-dimensional scaling (MDS) dimensions.

first four to eight dimensions of MDS space provide the most incremental benefit in capturing variance of the original SIFT-based pairwise distance matrix; inclusion of each further dimension adds important additional contributions to modeling the SIFT representation, but individual dimensions quickly diminish in descriptive power. To allow the search program to be able to converge, I limit the number of MDS dimensions to the top four. The potential shortcomings of a four-dimensional MDS SIFT space are evaluated in Chap. 5.

3.3.3 Subjects

Ten subjects (four female, age range 19 to 31) from the Carnegie Mellon University community participated, gave written informed consent, and were monitarily compensated for their participation. All procedures were approved by the Institutional Review Board of Carnegie Mellon University.

3.3.4 Experimental design

The study of featural selectivities in the perception of real-world objects was divided into an initial "reference" scanning session and two realtime scanning sessions for each subject. The reference session gathers cortical responses to the four classes of object stimuli. These responses are used to select the four brain regions — corresponding to the four object classes — to be further studied in the ventral pathway and to gather information about fMRI signal properties in these regions. The realtime scan sessions search for stimuli producing the maximal response from the four brain regions, dynamically choosing new stimuli to display based on the regions' responses to recently shown stimuli. Each realtime scan session begins with a distinct set of starting coordinates in the visual space, $x_{i,1}$ described in Sec. 3.1.5, corresponding to a distinct set of stimuli to display for the beginning of the four searches. After the completion of both sessions, I compare the visual feature selectivies identified in each session for each region to determine if search results are independent of starting conditions. This comparison constitutes one of my evaluations of my novel realtime processing and search methods, described in Sec. 3.5, which I study in addition to the scientific findings of ventral pathway regional feature selectivities.

Reference session

The reference session gathers cortical data needed to most-effectively pursue searches for real-world stimuli in the SIFT-based space defined in Sec. 3.3.2 and to perform realtime processing in the later scan sessions. In particular, the data are used to identify regions most differentially activated by each of the four stimulus classes, via a "class localizer," and are used to identify regions that group visual object images in a manner similar to SIFT, via a "SIFT localizer," described in Sec. 3.3.6. The overlap between these class and SIFT regions are used to define 125 voxel class/SIFT cubic regions of interest (ROIs) for study in the realtime scan sessions. Voxel-specific signal properties and multi-voxel pattern trends are learned from the data in the ROIs in the reference session and are used for realtime analysis, converting multi-voxel, multi-second responses from a given region to a given stimulus into a single number representing the region's response as discussed in Sec. 3.1.4.

Runs in the reference scan followed a slow event-related design, similar to that used in my recent work studying mid-level visual representations in cortical object perception [35]. Each stimulus was displayed in the center of the screen for 2.0 s followed by a blank 53% gray screen shown for a time period randomly selected to be between 500 and 3000 ms, followed by a centered fixation cross that remained displayed until the end of each 10 s trial, at which point the next trial began. As such, the SOA between consecutive stimulus displays was fixed at 10 s. Subjects were instructed to press a button when the fixation cross appeared. The fixation onset detection task was used to engage subject attention throughout the experiment. No other task was required of subjects, meaning that the scan assesses object perception under passive viewing conditions.

The 10 s SOA was chosen to minimize temporal overlap between voxel BOLD responses for multiple stimuli — based on the assumption that the hemodynamic response in the ventral-temporal cortex has decreased to a sufficient degree in the 10–12 s after stimulus onset to minimize the noise in my measurements of the cortical responses.

The stimuli were presented in four three-minute runs, spread across the one-hour scanning sessions and arranged to minimize potential adaptation and priming effects. Each run contained 36 object pictures, 9 objects from each of the four classes, ordered to alternate among the four classes similar to the realtime display design described in Sec. 3.1.1. Stimulus order was randomized across runs. Over the course of the experiment, each subject viewed each picture four times; averaging across multiple repetitions was performed for each stimulus, described below, to reduce trial-by-trial noise. I determined from data gathered in Leeds et al. that relatively little information is gained by averaging over more than four repetitions.

The session also included functional localizer scans to identify object selective cortex — namely, the Lateral Occipital Complex (LOC) — a functionally defined region [30] that we consider for comparison with the four "SIFT/object-class localizers" described above. For the LOC localizer, 16 s blocks of common everyday objects were alternated with 16 s blocks of phase-scrambled versions of the same objects, separated by 6 s of fixation [16, 30]. Phase scrambling was achieved by taking the Fourier transform of each image, randomizing the resulting phase values while retaining the original frequency amplitudes, and reconstructing the image from the modified Fourier coefficients [53]. Within each block, 16 images, depicting 14 distinct objects, were shown for 800 msec each, each object being followed by a 200 msec gray screen. Two of the objects were sequentially repeated once during each block — to maintain attention, subjects were instructed to monitor for this, performing a one-back identity task in which they responded via a keypress whenever the same object image was repeated across two image presentations. Six blocks of both the intact and scrambled objects conditions were presented over the 282 s scan [47]. The object images used in the localizer scans were different from the object picture stimuli discussed in Sec. 3.3.1. LOC area(s) were identified as those brain regions more selective for intact versus scrambled objects. LOC areas included all regions containing spatially contiguous voxels (no minimum cluster size) for which beta weights for the block design had significance level of $p < .005$.

To provide anatomical information, a T1-weighted structural MRI was performed between runs within the reference scanning session.

Realtime sessions

The realtime sessions displayed stimuli chosen in realtime to maximize the response of the four pre-selected ROIs under study, as discussed further in Sec. 3.1. The stimuli drawing the highest responses are considered to indicate the visual features selectivities for a given region.

Runs in the realtime analysis sessions followed a fast event-related design. Each stimulus was displayed centered on one of nine locations on the screen for 1 s followed by a centered fixation cross that remained displayed until the end of each 8 s trial, at which point the next trial began. As such, the SOA between consecutive stimulus displays was fixed at 8 s. For each trial, the stimulus center was selected to be +2.5, 0, or -2.5 degrees horizontally and/or vertically displaced from the screen center. The stimulus center changed with 30% chance on each new trial. Subjects were instructed to press a button when the image was centered on the same location as was the previous image. The one-back location task was used to engage subject attention throughout the experiment. This task was used instead of fixation detection employed in the reference scan because the one-back location task requires particular focus on each stimulus, which could potentially strengthen cortical activity above that elicited by passive viewing of the objects, aiding in the accurate computation of regional response in each trial. Unfortunately, the chosen design for the realtime scans also risks response variability resulting from slight changes in stimulus position and from maintaining the previous trial in memory as a strategy for comparing locations. In the reference session, there is greater liberty to pool responses over multiple runs in post-hoc analyses, and thus potentially weaker cortical signals were recorded for each trial using the fixation detection design while avoiding the potential confounds of the one-back location task.

The 8 s SOA was chosen instead of the 10 s SOA used in Sec. 3.3.4 to increase the number of objects viewed by the subject in each session. Concern about noise from temporally overlapping voxel responses — lasting 10–12 s after stimulus onset — is lessened because a stable response estimate can be obtained across the 10 s stimulus response signal by fitting each voxel signal to the average HRF for the ROI, learned during the reference scan described in Sec. 3.3.6. Furthermore, the design rotates among the four stimulus classes from trial to trial and each object class ROI was chosen to contain voxels responding selectively for stimuli in the associated class. Therefore, the response of a region for its given stimulus is expected to be relatively unaffected by the front and tail ends of the overlapping responses of the out-of-class stimuli displayed before and after it — responses that are presumably lower in overall amplitude. Notably, the "rapid" event related design is only 2 s faster per run than is the slow event related design used by the reference scan. The 8 s SOA is chosen to be short while still providing sufficient time for the realtime analysis and search programs to compute and return the next stimuli to display based on the previous cortical responses.

In each of the two 1.5-hour realtime scanning sessions, the stimuli were presented in four to eight 8.5-minute runs. Each stimulus was selected by the realtime search program based on ROI responses to the stimuli previously shown in the same category, as discussed in Sec. 3.1. Each run contained 60 object pictures, 15 objects from each class, ordered to alternate through the four classes. Interleaving the studies of four distinct visual object classes avoided adaptation, priming, and biasing effects.

Each realtime session began with an LOC localizer scan, following the design described in Sec. 3.3.4. This scan played an important role in the mechanics of the realtime search methods. The ROIs selected for study in the realtime scan sessions are defined as positions in the $96 \times 96 \times 31$ voxel volume returned by the fMRI scanner. However, the positions of the corresponding regions in the brain likely will differ between scan sessions, as the subject's brain will have a slightly different position and orientation in the fMRI volume each time he or she is placed into the scanner for the session. Because each voxel has its own assigned weight in each region, as a step towards multi-voxel pattern analysis discussed in Sec. 3.1.4, proper alignment between the reference scan and each realtime scan is important for each voxel. The first functional volume scanned for the LOC localizer was used to compute the spatial transformation between the brain's position in the current session and it's position in the reference session. This transformation was applied (in reverse) to the ROI locations computed from the reference scan data and the resulting corrected ROIs were applied in preprocessing analyses for the rest of the session. The initial estimation of the transformation matrix requires several minutes of computation — time that largely overlaps with the performance of the LOC localizer. Subject performance of the localizer task, rather than lying in the scanner with no task, decreases the risk for subject movement while the alignment computations to complete. Lack of motion within the session maximizes the alignment of the brain at any given point in scanning with the orientation of the corrected ROIs.

3.3.5 Preprocessing

Reference session

Functional scans during the reference scan session were coregistered to the anatomical image and motion corrected using AFNI [46]. Similar to the realtime preprocessing in Sec. 3.1.4, highpass filtering was implemented in AFNI by removing sinusoidal trends with periods of half and full length of each run (338 s) as well as polynomial trends of orders one through three. The data then were normalized so that each voxel's time course was zero-mean and unit variance [26]. To allow multivariate analysis to exploit information present at high spatial frequencies, no spatial smoothing was performed [62].

For each stimulus repetition, the measured response of each voxel consisted of five data samples starting 2 s after onset, corresponding to the 10 s between stimuli. Each five-sample response was consolidated into a weighted sum, intended to estimate the peak response. This sum took two forms, distinct from the method used in Sec. 3.1.4.

- For the "SIFT localizer," used to identify voxel regions that group stimuli in a manner similar to SIFT as in the analyses in Chap. 2, the sum over time points of each voxel's response was accomplished as one step in a "searchlight" process [32]. 123-voxel searchlight spheres — each with a radius of 3 voxels — were defined centered sequentially on every voxel in the brain. The average five-sample response of voxels across this sphere and across all stimulus presentations was computed. For a given searchlight, for each stimulus, each voxel was assigned a number based on the dot product of this average response and the voxel's mean response across all six repetitions of that stimulus. To the extent that hemodynamic responses are known to vary across cortical regions, this procedure allowed us to take into account a given voxel's local neighborhood mean-response shape. Fitting the local average response may provide a more accurate model of the relative activity of voxels across a sphere as compared to fitting a fixed response function across the whole brain. This "searchlight HRF-fitting" was previously employed successfully in my work on identifying models for intermediate-level visual representations in the ventral pathway, discussed in Chap. 2.

- For the “class localizer,” used to identify voxels responding more strongly to stimuli in one of the four object classes than to stimuli in the other three classes, the sum over time points for each voxel was simply an average of the middle three samples of its response [26]. Class preference analyses were performed on a per-voxel basis, rather than over a 123-voxel searchlight, providing less data from which to compute local hemodynamic response shapes used in the SIFT localizer. Thus, I employed an earlier peak estimation method, presuming response peaks across the brain will occur 4 to 8 s after stimulus onset.

Realtime session

At the beginning of each realtime session, an LOC localizer scan was run and AFNI was used in realtime to compute an alignment transformation between the initial functional volume of the localizer and the first functional volume recorded during the reference scan session. The ROIs selected for study in the realtime scan sessions (Sec. 3.3.6) were defined as positions in the voxel volume returned by the fMRI scanner during the reference scan. However, the actual positions of these object-specific regions are set in the anatomical reference frame of the brain. Changes in head position between reference and realtime scan sessions result in the brain, and its associated ROIs, moving to different locations in the scan volume. Because each voxel has its own assigned weight in each region, as discussed in Sec. 3.1.4, proper alignment between the reference scan and each realtime scan is important for each voxel. The transformation computed between the realtime LOC volume and the reference volume was applied in reverse to each voxel in the four ROIs. The resulting corrected ROIs were applied throughout the realtime search runs to extract signal from the voxels associated with each search.

Preprocessing during the realtime search runs was performed in realtime by the preprocessing program, as discussed in Sec. 3.1.4, and was used to compute region responses to recently display stimuli. The results of this preprocessing were used by the search program to select new stimuli to display to the subject to determine the visual properties evoking the highest activity from each of the selected regions. The results of preprocessing also were stored for later post-session analyses.

3.3.6 Selection of class/SIFT regions of interest

For each subject, functional activity recorded in the reference scan was used to identify brain regions most differentially activated by each of the four stimulus classes and that cluster images in a manner similar to SIFT. Regions of interest (ROIs) to study during the realtime scan sessions were selected manually from contiguous groups of voxels matching SIFT representation of objects and showing class-specific activity. Focus on these pre-selected (and non-overlapping) regions during realtime search sessions allowed me to target stimuli for each region — i.e., selecting stimuli limited to be within a defined real-world object class — in order to produce stronger fMRI signals for more reliable single-trial analyses in the presence of fMRI noise. The use of these pre-selected regions also maximized the effectiveness of using the SIFT-based space, defined in Sec. 3.3.2, to search for visual feature selectivities.

Class localizer

For each stimulus class S, selectivity s_c was assessed for each voxel by computing:

$$s_c = \frac{\langle r_c \rangle - \langle r_{\bar{c}} \rangle}{\sigma(r_c)} \tag{3.5}$$

where $\langle r_c \rangle$ is the mean response for stimuli within the given class, $\langle r_{\bar{c}} \rangle$ is the mean response for stimuli outside of the class, and $\sigma(r_c)$ is the standard deviation of responses within the class. I visually identified clusters of voxels with the highest relative responses for the given class, using a variable threshold and clustering through AFNI. Thresholds were adjusted by hand to find overlap with SIFT localizer results, rather than selecting thresholds based on significance tests.

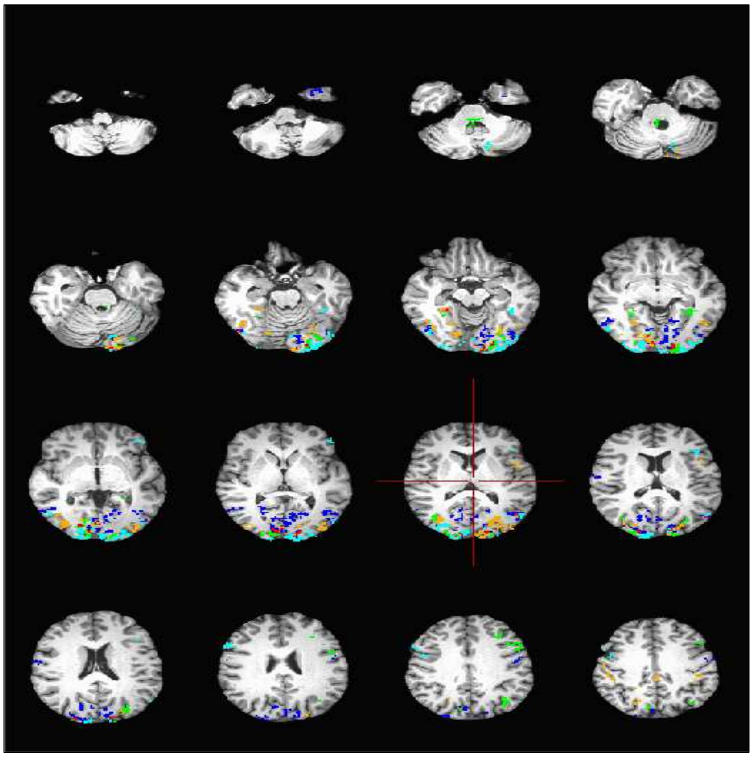


Figure 3.4: Class-selective regions for subject S9 in real-world object search. Colors are associated as follows: blue for human-forms, cyan for mammals, green for cars, orange for containers, red for overlap.

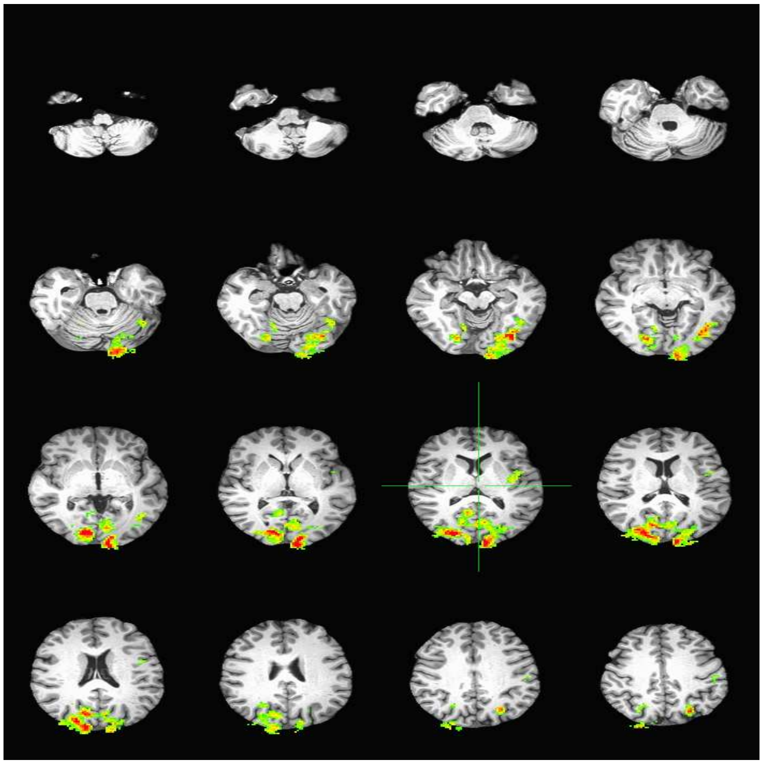


Figure 3.5: Cortical regions with dissimilarity structures highly correlated with the dissimilarty structure of the SIFT visual coding model for subject S9 in real-world object search. Comparisons follow the representational dissimilarity matrix-searchlight method discussed in Chap. 2.

Alternative approaches for merging class and SIFT localizer results are discussed in Chap. 6. In general, the highest-selectivity clusters of voxels appear within the visual processing streams, as we would expect, though further regions are apparent that may by associated with the semantic meaning of the class (Fig. 3.4). Chaps. 4 and 5 further assess the localizer results and their implications.

SIFT localizer

The representational dissimilarity matrix-searchlight method described in Chap. 2 was used to determine brain regions with neural representations of objects similar to the representation of the same objects by SIFT. Thresholds were adjusted by hand to find contiguous cluster with high voxel sphere z values, computed based on distance matrix correlations as described in Chap. 2;

full-volume significance tests were not performed. The regions showing highest matches between SIFT and voxel representations of the stimuli appear focused in the ventral visual stream, associated with visual object perception (Fig. 3.5), consistent with prior findings for SIFT [35].

Selection of ROIs

Visual inspection was used to find overlaps between the class-selective and SIFT-representational regions. For each class, a 125 voxel cube ROI was selected based on the observed overlap in a location in the ventral visual stream.

The use of relatively small cortical — one cubic centimeter — regions enables exploration of local selectivities for complex visual properties. Analyses were successfully pursued on similar spatial scales in our past work (Chap. 2 identifying 123-voxel searchlights showing significantly similar stimulus grouping structure to those of computer vision models and in past neurophysiological studies identifying neural columns showing selectivity to particular object shapes [63]. In potential future work, adjacent one cubic centimeter regions can be studied to mark the progression across the brain of featural selectivities in SIFT space.

ROI statistics

Once the ROIs were established, further statistics were computed for each region for use in realtime preprocessing. The average HRF was computed for each region, taken across all voxels and all real-world object stimulus displays. This HRF was used in the realtime sessions to compare with the time course of voxel responses to recently displayed stimuli. In computing region statistics, the original voxel time course responses for each reference scan stimulus display were consolidated into a weighted sum by computing the dot product of the response and the HRF. Proceeding with the consolidated voxel responses, principal component analysis was performed on the multi-voxel responses across runs and stimuli to identify the most common multi-voxel pattern. The first principal component — the most common pattern — was compared to multi-voxel pattern responses found in realtime. As indicated above, the HRF, the first principal component of the multi-voxel response, and the variance of each voxel's time courses were stored for use during realtime searches.

3.4 Fribble object search

While real-world object stimuli provide a more direct perspective on cortical object perception of regularly-observed objects, compared to the simplistic artificial stimuli often used in similar studies [4, 24], the broad variety of visual properties contained in such stimuli are difficult to capture in the small search space we can explore in a practical time frame with the simplex simulated annealing search. Indeed, Fig. 3.3 shows 10 dimensions of SIFT-based Euclidean space captures less than 35% of variance in the grouping structure of 1000 real-world objects, while my search can explore only four dimensions in the limited scanning time. Chaps. 4 and 5 report that realtime searches performed on cortical regions from the 10 subjects viewing classes of real-world objects purposely restricted to limit visual variability frequently failed to converge on clear visual feature selectivities (though some of the real-world object searches do show exciting results).

As a potential solution to the challenges of visual variability in the real-world objects, I pursue search for complex visual selectivities using a set of synthetic objects called “Fribbles” [76]. These stimuli are composed of colored, textured, three-dimensional shapes forming objects richer in structure than the gray blobs employed by Yamane et al. and Hung et al. [24, 78], but more controlled in appearance than the world of real objects. While the stimuli differ from the set discussed in Sec. 3.3, the majority of the methods are nearly identical for the dynamic selection of stimuli to identify preferred visual properties. The Fribble study provides a complementary perspective to the real-world object study on realtime search performance and on visual selectivities in the ventral pathway.

3.4.1 Fribble stimuli

Stimuli were generated based on a library of synthetic Fribbles [64, 76], and were displayed on 54% gray backgrounds as in Sec. 3.3.1. Fribbles are animal-like objects composed of colored, textured geometric shapes. They are divided into classes, each defined by a specific body form and a set of four locations for attached appendages. In the library, each appendage has three potential shapes, e.g., a circle, star, or square[4] head for the first class in Fig. 3.6, with potentially variable corresponding textures — thus, there are $3^4 = 81$ initial members of each class.

As in Sec. 3.3.1, four separate searches were performed in each realtime scanning session, probing the visual property selectivities of four distinct brain regions. Each search drew from a distinct class of Fribble objects, as shown in Fig. 3.6. While each search explored a relatively narrow visual space around a baseline body and configuration of appendages, the variability of appearance across Fribble classes allowed a broader perspective on selectivities across the world of objects (though perspective was considerably more constrained than it was on the set of real-world objects), and trends in the nature of visual features used across classes and brain regions.

A Euclidean space was constructed for each class of Fribble objects. In the space for a given Fribble class, movement along an axis corresponded to morphing the shape of an associated appendage. For example, for the purple-bodied Fribble class, the axes were assigned to 1) the tan head, 2) the green tail tip, and 3) the brown legs, with the legs grouped and morphed together as a single appendage type. Valid locations on each axis spanned from -1 to 1 representing two end-point shapes for the associated appendage, e.g., a circle head or a star head. Appendage appearance at intermediate locations was computed through the morphing program Norrkross MorphX [75] based on the two end-point shapes.

For each Fribble class, stimuli were generated for each of 7 locations — the end-points -1 and 1 as well as coordinates -0.66, -0.33, 0, 0.33, and 0.66 – on each of 3 axes, i.e., $7^3 = 343$ locations. Rather than generate each stimulus individually, the 7 appearances for each appendage

[4] Square heads were not used in this study, as discussed below.

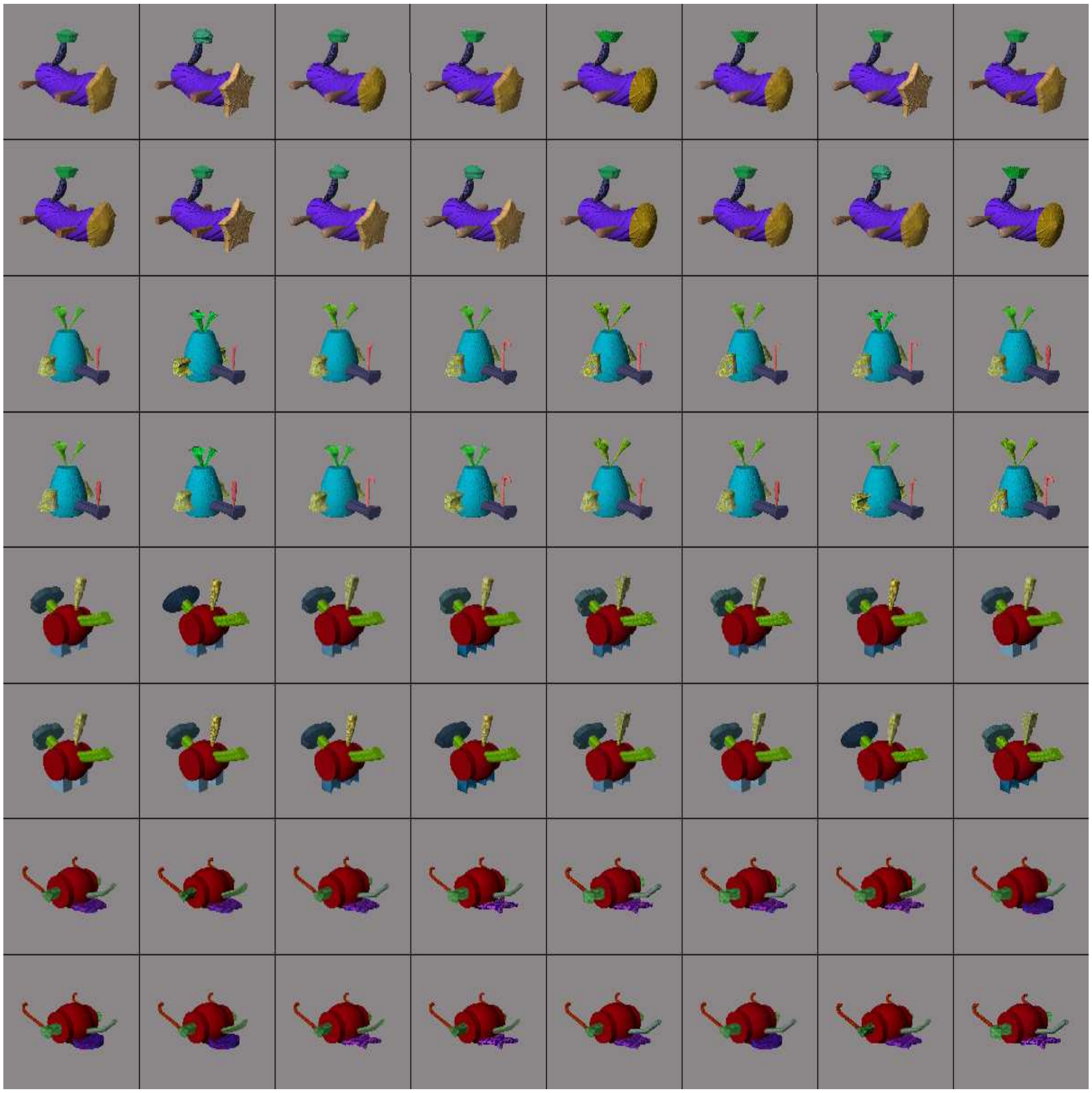


Figure 3.6: Example stimuli used in realtime search of Fribble objects. Images were selected from four synthesized classes, shown in rows in rows 1 and 2, 3 and 4, 5 and 6, and 7 and 8, respectively.

were generated separately and then assembled together into the full Fribble object for each of the 343 coordinates in the space.

While the coordinates in the space can be understood as visual edit instructions (through morphing) from a baseline shape at the origin, distances between object pairs in Fribble space are distinct from edit distances. Edit distances count the number of discrete changes between objects. In contrast, Fribble distances represent each appendage morph along a continuum of values, from 0 to 2, and combine changes to different appendages using Euclidean distance — the sum of the **squared** distances along each axis.

3.4.2 Subjects

Ten subjects (six female, age range 21 to 43) from the Carnegie Mellon University community participated, gave written informed consent, and were monitarily compensated for their participation. All procedures were approved by the Institutional Review Board of Carnegie Mellon University.

3.4.3 Experimental design

As with the real-world objects (Sec. 3.3.4), the search for Fribble objects producing the highest activity in pre-selected cortical regions — indicating visual featural selectivities — was pursued through an initial reference scan session and two realtime scan sessions. The reference session gathered cortical responses to the four classes of object stimuli. These responses were used to select the four brain regions — corresponding to the four Fribble classes — to be further studied in the ventral pathway and to gather information about fMRI signal properties in these regions. The realtime scan sessions searched for stimuli producing the maximal response from the four brain regions, dynamically choosing new stimuli to display based on the regions' responses to recently shown stimuli.

Reference session

The reference session design was almost identical to that of the reference session for real-world objects, described in Sec. 3.3.4. 36 stimuli, 9 stimuli from each of the four Fribble classes, were passively viewed in the context of a fixation onset detection task spread over four three-minute runs. The data gathered during these runs were used to identify Fribble class specific ROIs for study in the realtime sessions. Functional localizer scans to identify the object selective cortex — lateral occipital cortex, or "LOC" — were included for comparison with the four Fribble class localizers. To provide anatomical information, a T1-weighted structural MRI was performed between runs within the reference scanning session.

Realtime sessions

The realtime sessions displayed stimuli chosen in realtime to maximize the response of the four pre-selected ROIs under study. The stimuli drawing the highest responses are considered to indicate the visual features selectivities for a given region. The methods for stimulus selection are discussed in Sec. 3.1.

Runs in the realtime analysis session followed a fast event-related design. Each stimulus was displayed in the center of the screen for 1 s followed by a centered fixation cross that remained displayed until the end of each 8 s trial, at which point the next trial began. On any trial there was a 10% chance the stimulus would be displayed as a darker version of itself — namely, the stimulus' red, green, and blue color values each would be decreased by 50 (max intensity 256). Subjects were instructed to press a button when the image appeared to be "dim or dark." The dimness detection task was used to engage subject attention throughout the experiment. This task was used instead of fixation detection employed in the reference scan because the dimness detection task requires particular focus on each stimulus, which could potentially strengthen cortical responses above those elicited by passive viewing of the objects, aiding in the accurate computation of regional response in each trial and leading to more reliable choices in the dynamic selection of future stimuli. The one-back location task was employed for this purpose in the real-world object searches. However, the one-back location task may have caused subjects to recall past stimuli in memory when viewing new stimuli, to compare the location of the two stimuli to correctly determine if the stimuli were centered in different locations; this recollection potentially could contribute to limited interpretable results for this search method, as seen in Chaps. 4 and 5.

At the beginning of each realtime session, an LOC localizer scan was performed. The first volume from this localizer was used to compute the spatial transformation between the subject's brain position in the scanner during the current scan and the position during the initial reference scan session. This transformation was used to correctly align the Fribble class-specific ROIs identified from reference session data to the brain position in the current scan session, ensuring the correct voxels are studied throughout the realtime search runs. Sec. 3.3.5 further discusses the need for proper ROI alignment.

Further scan details were the same as those for the real-world object study, discussed in Sec. 3.3.4.

3.4.4 Preprocessing

Preprocessing of fMRI signals for reference and realtime scan sessions largely followed the same methods as did preprocessing for the real-world object search, discussed in Sec. 3.3.5.

Reference sessions functional scans were motion corrected, detrended, and filtered using AFNI. The time courses were further normalized. For the Fribble class localizer, the five-sample response of each voxel to each stimulus display was condensed to a single number representing response magnitude, computed using the searchlight HRF-fitting method used by the SIFT localizer.

In the realtime sessions, preprocessing was performed in realtime. The first volume of the LOC scan was used to align the ROIs computed based on the reference session to the subject's brain position in the current session. During realtime search runs, the preprocessing program computed region responses to recently displayed stimuli, as discussed in Sec. 3.1.4.

3.4.5 Selection of Fribble class regions of interest

I organize each class of Fribble object stimuli in a visual space that groups objects together based on morphs to specific shapes and textures (Sec. 3.4.1). The simplicity of this representational model makes it easy to study and search. However, its simplicity and specificity also risks the inability to properly characterize actual visual representations used in the ventral pathway. By performing the representational dissimilarity matrix-searchlight procedure used to identify cortical regions modeled by SIFT for the real-world objects (and regions modeled by other computer vision methods) in Chap. 2, I was able to identify cortical areas reasonably whose visual representations are well characterized by each simple Fribble space. ROIs were selected manually from these areas for study during the realtime scan sessions. In these regions, I could search effectively for complex featural selectivities using the associated Fribble space.

The RDM-searchlight method described in Chap. 2 is used to compare neural and Fribble space representations of stimuli. The distance matrix (RDM) for each Fribble space was constructed from pairwise distances between stimuli based on their degree of morphing for three appendages, each associated with an axis in the space. Thresholds for the maximum acceptable correlation between Fribble space and voxel-searchlight RDMs were adjusted by hand to find contiguous clusters with high voxel sphere z values, computed as in Chap. 2; full-volume significance tests were not performed. The regions showing highest matches appear in the ventral visual stream, associated with visual object perception, but also in a variety of areas less-typically associated with vision (Fig. 3.7). As Fribble object classes have visual similarities to animals and tools, semantic associations with class shapes may explain Fribble class associations in non-visual cortical areas.

125 voxel cube ROIs were selected for each object class based on visual inspection of searchlight results in the ventral visual stream. As in Sec. 3.3.6, the HRF, the first principal component

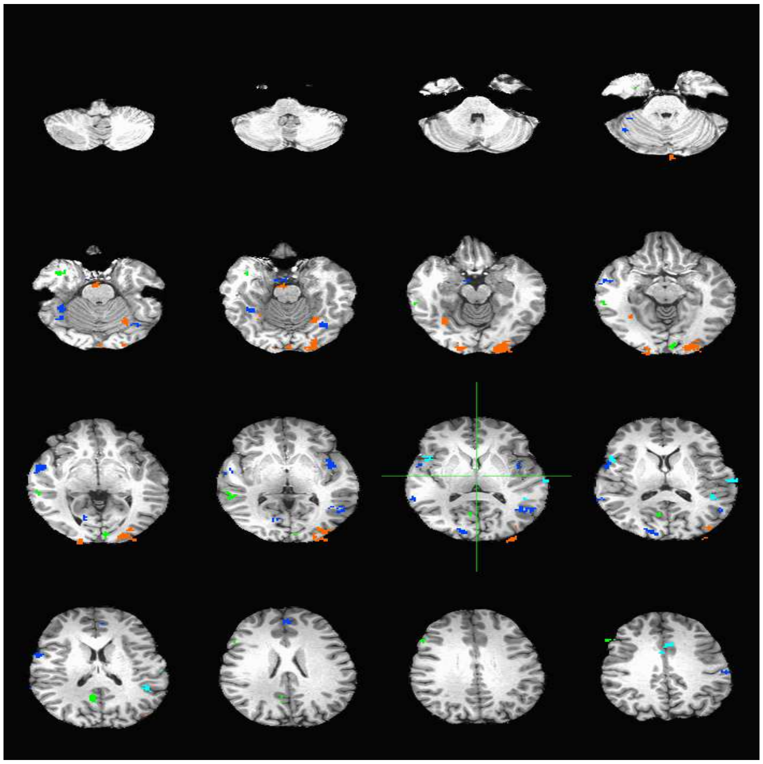


Figure 3.7: Cortical regions with a dissimilarity structure highly correlated with the dissimilarty structure of the space for each Fribble class for subject S1 in Fribble object search. Colors are associated as follows: blue for class 1, cyan for class 2, green for class 3, orange for class 4.

of the multi-voxel response, and the variance of each voxel's time courses were computed and stored for each ROI for future use during realtime search runs. These class-specific patterns were used in comparison with response signals recorded across multiple time samples and multiple voxels to derive a single numeric measurement of response for each stimulus display; this number was used to inform the search for region visual feature selectivity.

3.5 Assessment of search performance

The realtime search methods developed for and used in this study rely on a variety of assumptions about visual representations employed in the ventral visual stream and about the featural selectivities of small one cubic centimeter neural populations within that stream. Presuming, after the preprocessing employed in this work, the measured response of a selected ROI is characterized by a function with a unique maximum in the associated feature space, the simplex simulated annealing method will:

- **be consistent**, i.e., identify the stimuli in the area of feature space producing maximal response, regardless of the starting point of the search and
- **be convergent**, i.e., mostly investigate stimuli near the location producing maximal response, expending little time to investigate more distant stimuli that will evoke lower region responses.

Consistency provides confidence in the reliability of search results while **convergence** indicates advantage of strategic selection of stimuli over a limited scan time compared to random selection from the full pool of potential images. Metrics were defined for both properties and applied to all search results. Given the limited number of existant methods for the study of complex visual selectivities and for realtime fMRI analysis, assessment of my novel combination of methods is important to identify promising directions for further investigative approaches in the future.

Due to the variability of cortical responses and the noise in fMRI recordings, analyses were focused on stimuli that were visited three or more times. The average response magnitude for stimuli visited multiple times is more reliable for conclusions of underlying ROI selectivity. When subjects did not see the correct stimulus at the proper time for a trial, which happened on infrequent occasions discussed in Chap. 4, their ROI responses for those trials were excluded from analysis.

3.5.1 Convergence

For a given class, convergence was computed based on the feature space locations of the visited stimuli S, and particularly the locations of stimuli visited three or more times, $\mathbf{S_{thresh}}$. The points in $\mathbf{S_{thresh}}$ were clustered into groups spanning no more than d distance in the associated space based on average linkage, where $d = 0.8$ for Fribble spaces and $d = 0.26$ for SIFT space.[5] The result of clustering was the vector $\mathbf{clusters_{S_{thresh}}}$, where each element contained the numeric cluster assignment (from 1 to N) of each point in $\mathbf{S_{thresh}}$. The distribution of cluster labels in $\mathbf{clusters_{S_{thresh}}}$ was represented as $\mathbf{p_{clust}}$, where the n^{th} entry $p_{clust}(n)$ is the fraction of $\mathbf{clusters_{S_{thresh}}}$ entries with the cluster assignment n.

Conceptually, convergence is assessed as follows based on the distribution of points, i.e., stimuli visited at least three times:

- If all points are close together, i.e., in the same cluster, the search is considered to have converged.
- If most points are in the same cluster and there are a “small number” of outliers in other clusters, the search is considered to have converged sufficiently.
- If points are spread widely across the space, each with its own cluser, there is no convergence.

[5] The distance thresholds were chosen based on empirical observations of clusterings across regions and subjects in each space.

Set as an equation, the convergence metric is

$$metric(\mathbf{S}) = ||\mathbf{p_{clust}}||_2 - .1||\mathbf{p_{clust}}||_0 \tag{3.6}$$

where $||\mathbf{p_{clust}}||_2 = \sqrt{p_{clust}(1)^2 + \cdots + p_{clust}(N)^2}$ and $||\mathbf{p_{clust}}||_0$ is the number of non-zero entries of $\mathbf{p_{clust}}$. The metric awards higher values when $\mathbf{p_{clust}}$ element entries are high (most points are in a small number of clusters) and the number of non-zero entries is small (there are few clusters in total). Eqn. 3.6 pursues a strategy related to that of the elastic net, in which $\ell 2$ and $\ell 1$ norms are added to award a vector that contains a small number of non-zero entries, all of which have small values [80].

3.5.2 Consistency

For each subject and each stimulus class, search consistency was determined by starting the realtime search at a different location in feature space at the beginning of each of the two search scan sessions. In the first scan session, the starting position was set to the origin for each class, as stimuli were distributed in each space relatively evenly around the origin. In the second scan session, the starting position was manually selected to be in a location opposite from the regions with stimuli frequently visited and producing the highest magnitude responses. If a given dimension was not explored in the first session's search, a random offset from the origin along that axis was selected for the beginning of the second session. If the second search returns to the locations frequently visited by the first search, despite starting distant from those locations, the search method shows consistency across initial conditions.

The metric for determining consistency of results across search sessions was a slight modification of the convergence metric. The locations of the stimuli visited three or more times in the first and second searches were stored in $\mathbf{S^1_{thresh}}$ and $\mathbf{S^2_{thresh}}$, respectively. The two groups were concatenated into $\mathbf{S^{both}_{thresh}}$, taking note which entries came from the first and second searches. Clustering was performed as above and labels were assigned into the variable

$\mathbf{clusters_{Sboththresh}}$. The distribution of cluster labels was represented as probabilities $\mathbf{p_{clustBoth}}$.

To measure consistency, the final metric in Eqn. 3.6 was applied only to entries of $\mathbf{p_{clustBoth}}$ for which elements of $\mathbf{S^1_{thresh}}$ **and** $\mathbf{S^2_{thresh}}$ were present

$$metric(\mathbf{S^{both}}) = ||\mathbf{p_{clustBoth}}(i \in \mathbf{B})||_2 - .1||\mathbf{p_{clustBoth}}(i \in \mathbf{B})||_0 \tag{3.7}$$

where $\mathbf{B}$ is the set of indices i such that cluster i contains at least one point from $\mathbf{S^1_{thresh}}$ and from $\mathbf{S^2_{thresh}}$. The metric awards the highest values for convergence if there is one single cluster across search sessions. A spread of points across the whole search space visited consistently between sessions would return a lower value. Complete inconsistency would leave no $\mathbf{p_{clustBoth}}$ entries to be added, returning the minimum value of 0.

3.5.3 Testing against chance

As the convergence and consistency metrics above are not well established, it is not clear what values should be considered sufficiently high to indicate desirable search performance and what values would arise by chance. A variant of the permutation test is used to assess the metric results. The null hypothesis is that the convergence or consistency measure computed for a given search or pair of searches, based on clustering of the k stimuli visited three or more times during the search(es), would be equally likely to be found if the measure were based on clustering of a **random** set of k stimuli; this random set is chosen from the stimuli visited **one** or more times during the same search(es). The group of stimuli visited one or more times is considered a conservative estimate of all stimuli that could have been emphasized by the search algorithm through frequent visits. In the permutation test, the designation “displayed three or more times” is randomly reassigned among the larger set of stimuli displayed one or more times to determine if a random set of stimuli would be considered similarly convergent or consistent as the set of stimuli frequently visited in my study. More specifically, indices are assigned to all points visited in search 1 and search 2, S^1 and S^2, respectively, the indices and recorded number of visits are randomly permuted, and $metric(\mathbf{S^1})$, $metric(\mathbf{S^2})$ and $metric(\mathbf{S^{both}})$ are computed based on the locations randomly assigned to each "frequently-visited point." For each subject and each search, this process is repeated 500 times, the mean and standard deviation are computed, and the Z score for the original search result metrics are calculated. Based on visual inspection, searches with $z \geq 1.8$ are considered to mark notably non-random convergence or consistency.

Chapter 4

Results

My study was designed to explore complex visual properties utilized for object perception by the ventral pathway in the brain. To this end, I implemented and employed a collection of techniques in realtime fMRI analysis and in dynamic stimulus selection to identify the visual feature selectivities of pre-selected voxel regions. Dynamic stimulus selection was pursued to most effectively explore the space of visual properties in limited scan time and to most quickly identify objects that produce the highest responses from each brain region under study.

Several of my methods are novel and, as they dynamically interacted with newly recorded cortical activity, required “testing in the field” through execution of my realtime study. Below, I examine the performance of the programs I have written for 1) realtime fMRI signal analysis used to determine region response to recently viewed stimuli, 2) selection of new stimuli to display to subjects, and 3) display of the newly-chosen stimuli. I confirm the programs generally work as expected, while identifying areas for future improvement, e.g., in proper stimulus display, and areas challenging my initial assumptions about visual encoding in intermediate brain regions, e.g., in the observed results from exploring visual feature spaces.

To better understand cortical object perception, I examine the results of realtime analysis for evidence of the selectivities of pre-selected brain regions in the ventral pathway. I study the distribution of recorded ROI responses in the visual feature space, defined in Sec. 4.1, as well as comparing responses recorded for anatomically proximal ROIs. I visually inspect stimuli producing high responses from individual ROIs to gain intuition about the properties most exciting to these regions. Unfortunately, results from the analyses of many subject brain regions are inconclusive. However, in several ROIs, one can observe one or a few sets of visual properties evoking high responses, and slightly-different visual properties evoking extreme negative responses, reminiscent of surround suppression seen in lower levels of the visual system [22, 73]. Salient visual properties are seen to include holistic object shape, shapes of component parts, and surface textures.

Realtime analyses were performed on two groups of subjects using two types of object stimuli. 10 subjects viewed photographs of real-world objects, capturing perception of visual properties as they appear in the world, and 10 subject viewed images of synthesized Fribble objects [76], capturing visual properties carefully controlled in the creation of the stimulus set. The performance of my methods and the selectivities they revealed are reported below for each group of objects.

4.1 Feature spaces and a tool for their visualization

Dynamic selection of a new stimulus to display is performed in the context of a simplex search of a visual feature space, as described in Sec. 3.1.5. The Euclidean space used for each search was constructed to represent complex visual properties by spatially grouping stimuli that are considered similar according to a selected visual metric. When choosing stimuli within a class of real-world objects — specifically, mammals, human-forms, cars, or containers — the search space is defined from a SIFT-based similarity metric, as discussed in Sec. 3.3.2. When choosing stimuli within a Fribble class, the search space is defined based on visual morphs to components of an example object from the class, as discussed in Sec. 3.4.1.

Each space contains a low number of dimensions — four dimensions for SIFT and three dimensions for each Fribble class — to allow the searches for visual selectivities to converge in

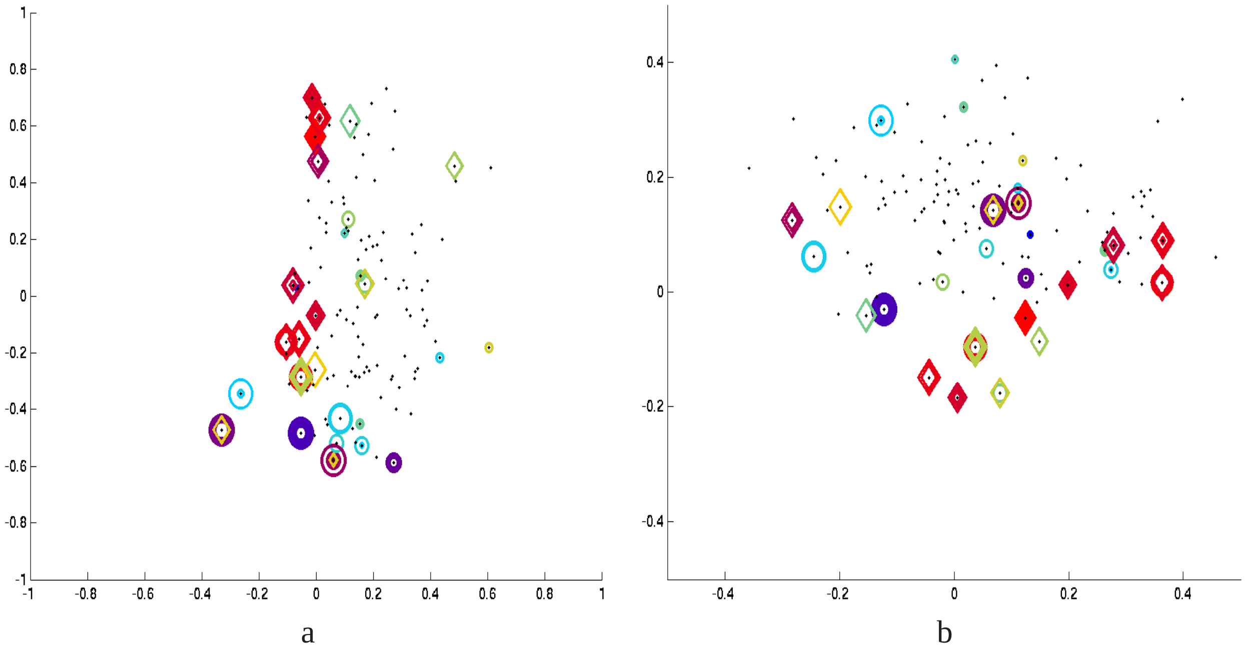


Figure 4.1: Search results for S3, class 2 (human-forms), shown in (a) first and second SIFT space dimensions and (b) third and fourth dimensions. Location of all potential stimuli in space shown as black dots. Results from realtime scan session 1 are circles, results from realtime scan session 2 are diamonds. For stimuli visited three or more times, colors span blue–dark blue–dark red–red for low through high responses; for stimuli visited one or two times, colors span cyan–yellow–green for low through high responses. Size of shape corresponds to time each point was visited in search, with larger shapes corresponding to later points in search. Note axes for (a) are from -1 to 1 and for (b) are from -0.5 to 0.5.

the limited number of simplex steps that can be evaluated over the course of a scanning session. These low dimensional spaces also permit visualization of search activity over each scan session and visualization of general ROI response intensities across the continuum of visual properties represented by a given space. I display this information through a colored scatter plot. For example, representing each stimulus as a point in feature space, Fig. 4.1 shows the locations in SIFT-based space visited by the search for human-form images evoking high activity in the pre-selected SIFT/"human-form" region of subject S3, and shows the regional response to each of the displayed stimuli. The four dimensions of SIFT-based space are projected onto its first two and second two dimensions in Figs. 4.1a and b, respectively. Stimuli "visited" during the first and second realtime sessions are shown as circles and diamonds, respectively, centered at the stimuli's corresponding coordinates in the space. (Black dots correspond to the locations of all stimuli in the human-form class that were available for selection by the search program.) The magnitude of the average ROI response to a given visited stimulus is reflected in the color of its corresponding shape. For stimuli visited three or more times, colors span blue–dark blue–dark red–red for low through high average responses; for stimuli visited one or two times, colors span cyan–yellow–green for low through high responses. The average responses for stimuli visited three or more times are more reliable reflections of regional response — data from my previous study (Chap. 2) indicates noise effects are greatly reduced through averaging over responses to three or more viewings of the same stimulus. Furthermore, repeated visits to a location by the search method indicates the method "expects" stimuli drawn from points near this location will evoke high responses from the brain region under analysis. Initial inspection of the red-or-blue shapes shows two clusters of stimuli along the y-axis in Fig. 4.1a evoking high responses, indicating multiple distinct featural selectivities in the selected ROI. Furthermore, stimuli corresponding to nearby locations in space can produce extreme high and low responses together, indicating the ROI can suppress its activity due to slight changes in visual features. Similar findings are observed for other subjects and searches in Sec. 4.4.

Intuition about search method behavior is further provided by the colored scatter plot display.

Color coding helps visually distinguish between frequently and infrequently probed stimuli. The size of each shape in the plot reflects the time each location was visited in the scan session, with larger shapes corresponding to later points in the search. Examination of the locations of large and small shapes provides visual intuition of the temporal evolution of the simplex search in the feature space. Reviewing the first two dimensions in Fig. 4.1a, one can observe stimuli visited early in the scan are more likely to be more distant from the main clusters of stimuli frequently displayed by the end of the scan. More objective measurements of search performance, confirming the trend visually indicated in Fig. 4.1, are explored in Sec. 4.3.3.

Sec. 4.4 contains further examples showing that low-dimensional visual feature spaces — both for real-world and Fribble objects — and their corresponding scatter plot visualizations provide powerful new means to understand complex feature selectivities of regions in the ventral object perception pathway and to evaluate the performance of my novel realtime search methods.

4.2 Selection of regions of interest

Realtime searches for cortical visual selectivities dynamically measured and incorporated the responses of pre-selected brain regions to recently-displayed stimuli to determine new stimuli to display. For each subject, four searches were performed, each exploring a distinct class of visual objects and an associated ROI. The four brain regions were selected prior to the realtime sessions using data collected from a previously-performed “reference” scan session. Analysis of cortical activity recorded in this earlier session identified cortical areas that were characterized by the SIFT search space, or by each Fribble class search space, in the representation of visual properties; for real-world objects, analysis of cortical activity also was performed to identify areas that were more highly activated when viewing objects in a given class. Selection of ROIs from these brain areas strengthened the validity of assumptions about the correct form of the search space employed and, for real-world objects, likely resulted in stronger ROI responses to stimuli used within the associated searches, lessening the effects of background noise.

Both for subjects viewing real-world objects and subjects viewing Fribble objects, ROIs containing cubes of 125 voxels were manually selected for each of four stimulus classes searched. Beyond incorporating voxels most highlighted by reference scan analyses reviewed above, the four regions for each subject were selected to be non-overlapping and to lie within the ventral pathway, with a preference for more anterior voxels, presumably involved in higher levels of object perceptions. With this selection approach in mind, consideration of the anatomical locations of the chosen ROIs provides perspective on the span of areas using SIFT-like and "Fribble-morph-like" representational structures across subjects, and the distribution of areas most strongly encoding each of the four studied object classes across subjects. We also gain perspective on the range of brain areas across subjects and searches studied for complex visual selectivities.

ROIs used for real-world object searches are distributed around and adjacent to the fusiform cortex, while ROIs used for Fribble object searches are distributed more broadly across the ventral pathway.

4.2.1 Real-world objects search

ROIs selected for study using real-world objects, organized using a SIFT-based feature space, were distributed across the ventral pathway, as shown in Fig. 4.2. Regions largely were centered in or near fusiform cortex, with limited anterior and lateral spread of centers. This anatomical focus reflects the strong matches between SIFT and multi-voxel code representations of object stimuli in fusiform cortex, not present in other areas associated with mid- and high-level vision. In Sec. 3.3.6, Fig. 3.5 shows the findings of the "SIFT localizer" for subject S9, indicating the presence of SIFT-associated regions in fusiform cortex and around primary visual cortex. Similar results are observed across subjects when comparing SIFT and multi-voxel encodings in my previous work discussed in Chap. 2. To study visual properties used in higher level vision, I selected ROIs beyond the primary visual cortex, resulting in ROIs focused around fusiform

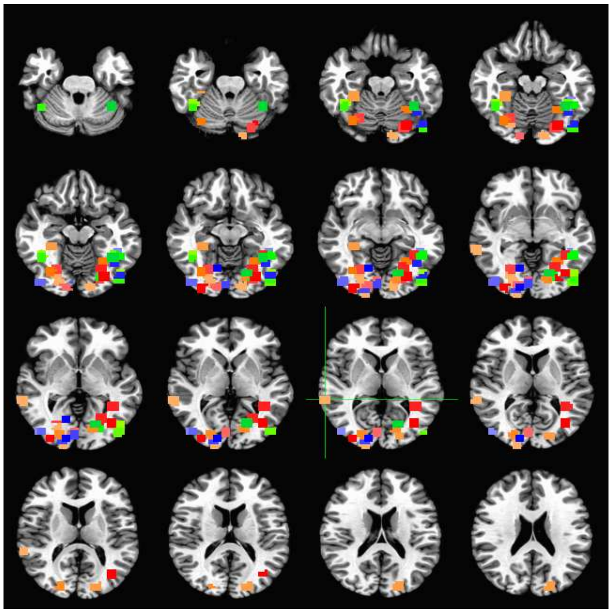


Figure 4.2: Class-selective regions for 10 subjects in real-world objects search, projected onto the Talairach brain. Colors are associated as follows: blue for mammals, green for human-forms, orange for cars, red for containers, overlaps shown as single color. Each subject is assigned a shade of each of the four colors.

cortex. SIFT encodes images through a small number of non-linear operations on selected edge statistics. Thus, it may serve as a model of cortical visual processing in areas anatomically and computationally close to the "edge detectors" of primary visual cortex, but less closely predict representations in higher-level brain areas [35].

The locations of ROIs specific to each of the four real-world object classes did not form any clear patterns across subjects. Fig. 4.2 shows regions for each class — assigned to the four colors blue, green, orange, and red — were centered in both hemispheres in varied positions — anterior and posterior, medial and lateral. Inspection of spatial ROI clustering in the Talairach brain, shown in Fig. 4.3, similarly indicates no clear grouping across subjects of class-specific regions. For example, while human-forms (class 2) regions for S4, S8, and S9 (corresponding to d, h, and i in the dendrogram labels) are grouped together, their group also contains two ROIs for mammals (class 1) and one ROI for containers (class 4). The remaining human-forms regions are spread into various other clusters. These observations, reflecting "class localizer" results,

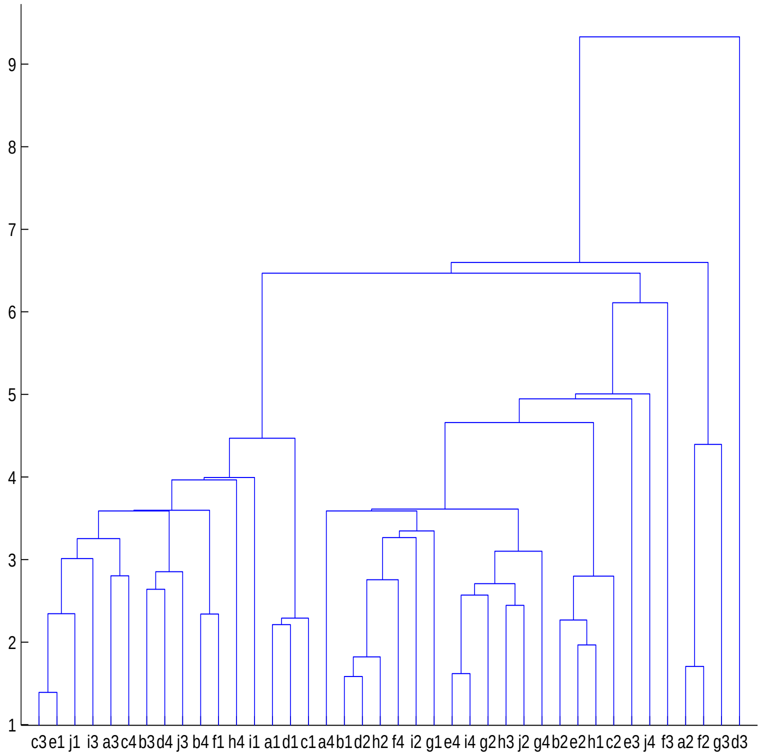


Figure 4.3: Clustering of Talairach-coordinate centers for class-selective regions for 10 subject in real-world objects search, shown as a dendrogram. Height of links between subtrees indicates shortest distance between members of the two trees as number of voxels in Talairach brain ($54 \times 64 \times 50$ voxels). Regions are labeled as nM, where $n \in \{a, \ldots, j\}$ corresponds to the subject numbers S$\{1, \ldots, 10\}$ and $M \in \{1, 2, 3, 4\}$ is the region number, corresponding to mammals, human-forms, cars, and containers.

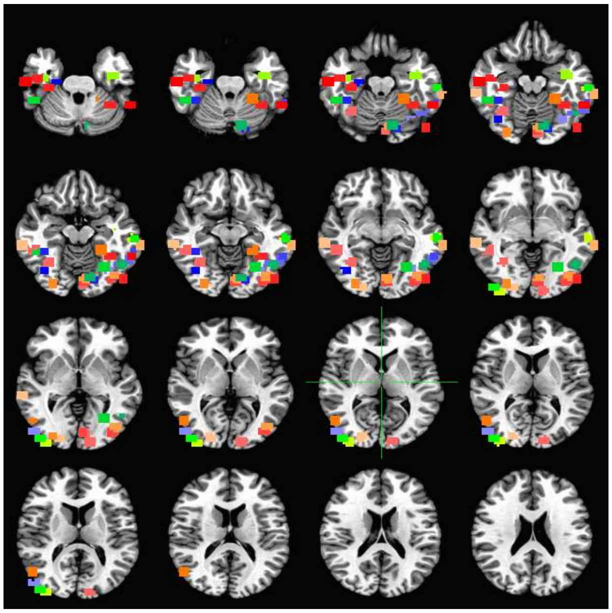


Figure 4.4: Class-selective regions for 10 subject in Fribbles search, projected onto Talairach brain. Colors are associated as follows: blue for class 1, green for class 2, orange for class 3, red for class 4, overlaps shown as single color. Each subject is assigned a shade of each of the four colors.

underscore cross-subject variability as we delve into more narrow areas in object perception.

4.2.2 Fribble objects search

ROIs selected for study using Fribble objects, organized using feature spaces defined on morphs to Fribble components, were distributed broadly across the ventral pathway, as shown in Fig. 4.4. Regions were centered in areas from fusiform to lateral occipital to anterior inferotemporal cortex, in addition to posterior areas above the occipital pole. This spread is notably more broad than that of ROIs selected for real-world objects, shown in Fig. 4.2. The morphing operations performed to shape the space for each Fribble class operate on the forms, colors, and textures of whole component shapes — such as circle or star heads for purple Fribbles in Fig. 3.6 — potentially constituting a "higher-level" process than the non-linear fusion of localized edge statistics computed by SIFT. This increased complexity may account for the recruitment of more anterior (and perhaps more lateral) areas beyond SIFT's fusiform regions in the ventral pathway.

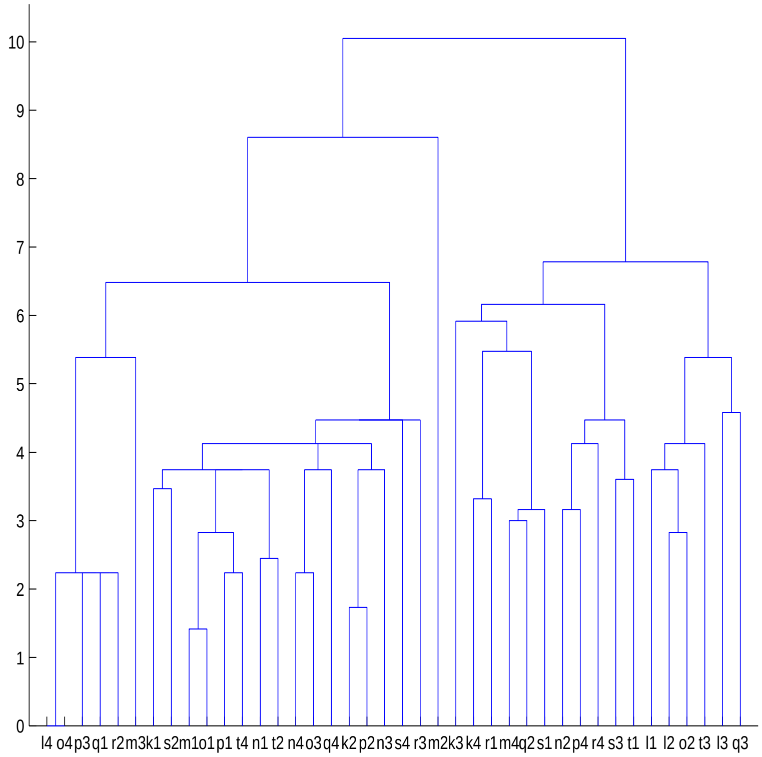


Figure 4.5: Clustering of Talairach centers for class-selective regions for 10 subject in Fribbles search, shown as a dendrogram. Height of links between subtrees indicates shortest distance between members of the two trees as number of voxels in Talairach brain ($54 \times 64 \times \times 50$ voxels). Regions labeled as nM, where $n \in \{k, \ldots, q\}$ corresponds to the subject numbers S$\{11, \ldots, 20\}$ and $M \in \{1, 2, 3, 4\}$ is the region number, corresponding to the four classes illustrated in Fig. 3.6.

However, more posterior regions still are selected for many subjects and Fribble classes as well.

As in Sec. 4.2.1 for the study of real-world objects, the locations of ROIs specific to each of the four Fribble object classes did not form any clear patterns across subjects. Fig. 4.4 shows regions for each class — assigned to the four colors blue, green, orange, and red — were centered in both hemispheres in varied positions — anterior and posterior, medial and lateral. Inspection of spatial ROI clustering in the Talairach brain, shown in Fig. 4.5, similarly indicates no clear grouping across subjects of class-specific regions. Perhaps the group nearest to a cluster of same-class regions is the 8-element group containing 5 class 1 regions — for S11, S13, S14, S15, and S16 — and one region from each other class. These observations underscore cross-subject variability as we delve into more narrow areas in object perception.

4.3 Realtime search performance

To search for complex visual feature selectivities using stimuli evoking maximal activity from pre-selected brain regions, I designed and used a collection of three programs, introduced in Sec. 3.1 — 1) the display program, 2) the preprocessing program, and 3) the search program. Together, these programs work in realtime 1) to display object stimuli to a subject, 2) to measure the ROI responses to these stimuli, and 3) to select further stimuli to display and further probe regional selectivities. As I developed the programs for the present study, and my realtime fMRI stimulus search explored uncharted waters in methodology, I study the behavior of my code over the course of scan sessions to confirm its generally successful execution and to understand technical and scientific challenges to overcome in future work.

4.3.1 Display program behavior

The display program continuously interacts with the preprocessing and search programs to properly execute the search of ROI activities in visual stimulus spaces. At any given point during the realtime scan, the search program determines a new stimulus to show to a subject based on the subject's ROI responses to recently-shown stimuli. These responses are extracted from set times intervals in the fMRI signal by the preprocessing program, presuming the stimulus associated with each response was viewed by the subject at the intended time. From the perspective of my realtime search, **the display program's central task is to display each intended stimulus** (chosen by the search program) **at its intended time** (at the beginning of its associated 8 s trial, described in Sec. 3.1.3).

Unfortunately, in the course of each realtime session, challenges periodically arose to the prompt display of the next stimuli to explore in each realtime search[1]. The computations required

[1] Four realtime searches were performed during each realtime run for four distinct object classes and four distinct

to determined ROI response to a recent stimulus and to determine the next stimulus to display occasionally did not complete before the time required by the display program to show the next search selection. When the new stimulus choice was not made sufficiently quickly, the stimulus displayed to the subject could be shown seconds delayed from its intended onset time or could incorrectly reflect the choice made from the previous iteration of the search, depending on the stimulus update method used by the display program.

Two stimulus update methods were used by the display program, as explained in Sec. 3.1.2. **Update method 1:** For five of the subjects viewing real-world objects, the display program received the search program's next stimulus choice by reading a file in a directory shared between the machines respectively running the display program and the search program. Delays in the search program computations and in directory updates over the network sometimes resulted in the display program showing the stimulus from the previous search step. To circumvent potential delays in shared directory updates, the display program ran a forced directory update prior to reading the latest copy of the chosen stimulus file for subjects S9 and S10. This directory update sometimes caused noticeable delays in stimulus display time and thus was discontinued for S6, S7, and S8.

Update method 2: For the remaining subjects, five viewing real-world objects and ten viewing Fribble objects, the display program received the search program's next stimulus choice through a dedicated socket connection. The display program waited to receive a message from the search program before proceeding, thus leading to noticeable display delays when the search program's computations required extra time for the given block of data. The search program also sometimes skipped simplex computations for a given class of objects at a given step, e.g., if a new point had been accepted into the simplex for one class but multiple point evaluations were left before simplex acceptance for other classes, c.f., [39]. If a class was skipped, no information was written over the socket, and the display program waited until the next time new

brain regions. All programs alternated between computations for each of the four searches, i.e., `search 1` → `search 2` → `search 3` → `search 4` → `search 1` $\cdots$, as discussed in Sec. 3.1.1.

information was written before displaying stimuli. Thus, stimulus displays on some occasions occured at 20 second delays, followed in succession by the displays of the other stimuli whose trials had passed during the time waiting. This problem did not occur often, but requires further code development for future versions of the realtime search study. Notably, use of direct socket communication significantly reduced the number of displayed stimuli that were not the current choices of the search program for the given object class.

Generally, the display program showed the correct stimulus at the intended time across subjects, sessions, and stimulus groups. Below, I report the infrequent **late** and **wrong** stimulus displays. The first five subjects scanned were the only ones for whom searches had display errors for more than 10% of trials. For these subjects, S6, S7, S8, S9, and S10, all viewing real-world objects, stimuli were updated through checking of a file in a mounted directory — therefore, this update approach was not used for the remainder of the study. For all subjects and sessions, trials in which there was a delay of 0.5 s or more or in which the wrong stimulus was shown are removed from consideration in analyses beyond those in the present section.

Real-world objects search

The number of displays that appeared late or showed the wrong stimulus for subjects viewing real-world objects is shown in Table 4.1 for each subject, object class, and scan session. Stimulus presentations were considered delayed if they were shown 0.5 s or more past the intended display time.

When updates for display stimuli were performed through inspection of shared files, for S6, S7, S8, S9, and S10, showing of incorrect stimuli dominated the display errors. S6, S7, and S8 were shown incorrect stimuli for 15 to 42% of trials for search 1 and search 3, corresponding to the mammal and car classes. Among these three subjects, incorrect displays for search 2 and search 4 only were observed in session 2 for S6. S9 was shown no incorrect stimuli; S10 was shown incorrect stimuli on ˜10% of trials for all searches in session 1 and no incorrect stimuli in

Subject$_{\text{session}}$	late1	late2	late3	late4	wrong1	wrong2	wrong3	wrong4	# trials
$S1_1$	4	3	3	1	0	0	0	0	80
$S1_2$	0	0	0	0	0	0	0	0	96
$S2_1$	3	1	1	1	0	0	0	0	96
$S2_2$	3	1	1	0	0	0	0	0	112
$S3_1$	8	4	9	7	0	3	0	1	112
$S3_2$	3	2	2	2	3	0	3	0	112
$S4_1$	4	3	4	2	0	0	0	1	112
$S4_2$	5	4	4	1	0	0	0	1	112
$S5_1$	6	3	4	4	0	0	0	0	112
$S5_2$	7	2	3	6	0	0	0	0	112
$S6_1$	0	0	0	0	23	0	18	0	112
$S6_2$	0	0	0	0	41	25	47	23	112
$S7_1$	0	0	0	0	32	0	24	0	112
$S7_2$	0	0	0	0	21	0	30	0	112
$S8_1$	0	0	0	0	36	0	24	0	112
$S8_2$	0	0	0	0	30	0	25	0	112
$S9_1$	5	3	3	3	0	0	0	0	112
$S9_2$	5	1	1	2	0	0	0	0	112
$S10_1$	3	1	2	3	12	13	12	11	112
$S10_2$	6	2	3	3	0	0	0	0	112
total	62	30	40	35	198	41	183	37	

Table 4.1: Number of delayed and incorrect display trials for real-world objects searches for each object class and each subject. Delayed trials were those shown 0.5 s or more past the intended display time. Results tallied separately for realtime sessions 1 and 2 for each subject, and tallied separately for each stimulus class. Class numbers correspond to mammals, human-forms, cars, and containers, respectively. Number of search trials per class varied in each session, as seen in final column.

session 2.

The preprocessing and search programs evaluate cortical responses and compute new stimuli to display in blocks of two searches at a time — examining and acting on the fMRI data for search 1 and search 2, then examining and acting on the fMRI data for search 3 and search 4 — as explained in Sec. 3.1.1. Ordinarily, the new stimuli to be shown for the next block of two searches were computed 1 s or more before the first stimulus for this block is shown by the display program. The second stimulus of the block then is shown at the start of the following trial, 8 s later. When preprocessing is slowed (by factors considered below), the first stimulus of the block may be chosen 0.5 s prior to display time or several seconds after display time. Even when stimuli are chosen 1 s prior to display time, updates through the shared files read over the mounted folder may require as much as 3 s to complete, resulting in the display program reading and acting on old stimulus choices. These sources of typically 1 to 5 s delay past display time in conjunction with the block processing method result in the strong discrepency in incorrect display frequency of search 1 and search 3 — whose updates may not arrive to the display computer by the required time — compared with that of search 2 and search 4 — whose updates usually arrive at least 3 s before they are needed. Despite the frequency of incorrect stimuli displayed in searches for object classes 1 and 3, it is important to note that even in the worst case, correct stimuli were displayed on at least 55% of trials.

Variability in the frequency of display errors reflects variability in the speed of realtime fMRI signal processing and in update speeds for the directory mounted and read by the display computer. The preprocessing program removes low-frequency drift and motion effects from the scanner signal prior to computing ROI responses. Depending on subject motion and scanner magnet behavior during each scan session, the computation and extraction of these signal components can require seconds of extra processing time. Other programs running on the “analysis machine” — the machine running the preprocessing and search programs — also can unexpectedly take up processor resources, slowing down realtime analysis and communication of updates to the directory mounted by the display machine. While I initiate no extra programs on the analysis machine during realtime sessions, I also do not reconfigure the machine to suspend potentially unnecessary background processes.

When updates for display stimuli were performed through inspection of shared files, and an update was performed on the directory containing the files, display errors also included a limited number of delayed displays. S9 and S10 had delayed stimulus displays for 1 to 6% of trials, with delay on at least one trial for every session and for each of the four searches. In most cases, there were more delays for search 1 than for any of the other searches. These delays likely resulted from the directory update performed by the display program prior to reading the file from the directory containing the stimulus choice for the current search. The update operation usually executes in a fraction of a second, but occasionally runs noticeably longer. Chances of a longer-duration update are greater when the operation has not been performed recently, such as at the start of a realtime search run following a ˜2 minute break between runs. As search 1 starts every run, it may be slightly more likely to experience display delays

When updates for display stimuli were performed through a socket, for S1, S2, S3, S4, and S5, display delays dominated the errors in display program performance. Most subjects had delayed stimulus displays for 1 to 9% of trials, with delay on at least one trial for every session and for each of the four searches. However, the second session for S1 showed no delayed displays, nor did search 4 for the second session for S2. The number of delays for search 1 was greater than or (occasionally) equal to the number of delays for any of the other searches, except for session 1 for S3 for which search 3 had the most delays. Across the five subjects, search 3 had the second, or sometimes first, highest number of delayed displays. The discrepency in display error frequency between the first searches of each processing “block” as described above, i.e., search 1 and search 3, and the second searches of each processing block, search 2 and search 4, are significantly less pronounced than they were for the frequency of incorrect stimuli for S6, S7, S8, S9, and S10, though the pattern remains weakly observable. For S1, S2, S3, S4, and S5, display delays can result from delays in completing processing of cortical responses for the block of two recently viewed stimuli — causing a greater number of delays for search 1 and search 3, as described above. Display delays also can occur when the search program refrains from exploring a new simplex point for a given stimulus class at a given iteration, as described above, thus refraining from sending a stimulus update over the socket. This lack of communication causes the display program to pause several seconds because it will not display any stimulus until it receives new information over the socket.

A limited number of incorrect stimulus displays also occurred when updating display stimuli through a socket. S3 and S4 were shown incorrect stimuli on 1 to 3% of trials for one or two searches in each scan session. The source of these errors was not determined, though they may have resulted from skipped evaluations in the simplex search. These errors did not occur using socket updates for searches of Fribble object stimuli reported below.

Far fewer display errors occured when updates for display stimuli were performed over a socket than when they were performed through inspection of a shared file. Indeed, the socket update approach was introduced to improve communication speed between the search program and the display program and, thereby, to decrease display errors. (S1, S2, S3, S4, and S5 were studied after the five other subjects viewing real-world object stimuli.) Reflecting on the increased performance caused by use of sockets, I employ only socket communication for the Fribble objects searches.

Fribble objects search

The number of displays that appeared late for subjects viewing Fribble objects is shown in Table 4.2 for each subject, object class, and scan session. Stimulus presentations were considered delayed if they were shown 0.5 s or more past the intended display time. There were no displays showing the wrong stimuli, because the display program waited for updates to each stimulus over an open socket with the search program before proceding with the next display.

All subjects had delayed stimulus displays in each scan session in one or more of the four searches. Across subjects, a total of ˜70% of searches showed delayed displays, with errors

Subject$_{\text{session}}$	late1	late2	late3	late4	# trials
S11$_1$	4	0	3	3	96
S11$_2$	5	0	2	4	80
S12$_1$	4	0	1	3	80
S12$_2$	11	5	5	9	96
S13$_1$	6	2	2	1	96
S13$_2$	6	2	2	1	96
S14$_1$	3	0	0	0	96
S14$_2$	5	0	0	1	80
S15$_1$	6	2	3	1	64
S15$_2$	5	0	0	2	80
S16$_1$	5	0	0	0	80
S16$_2$	5	0	0	0	80
S17$_1$	3	0	0	0	96
S17$_2$	6	1	0	2	96
S18$_1$	6	2	2	1	80
S18$_2$	6	0	0	1	80
S19$_1$	2	1	1	1	96
S19$_2$	3	2	2	1	80
S20$_1$	3	0	0	2	96
S20$_2$	5	1	1	3	64
total	99	18	24	36	

Table 4.2: Number of delayed display trials for Fribble searches for each subject. Delayed trials were those shown 0.5 s or more past the intended display time. Results tallied separately for realtime sessions 1 and 2 for each subject, and tallied separately for each stimulus class. Class numbers correspond to object classes. Number of search trials per class varied in each session, as seen in final column.

occuring in 1 to 10% of trials. The number of delays for search 1 was greater than the number of delays for any of the other searches; across subjects, search 1 had roughly three times as many errors as any of the other classes. Delayed displays were produced by delays in the completion of fMRI signal preprocessing and by skipped simplex search evaluations, as discussed for subjects viewing real-world objects.

The first block processed for each run requires slightly extra time for processing than does any other block, because the first block contains six extra volumes, corresponding to the cortical activity prior to the start of the first display trial. Often, this extra processing time causes a delay for the first update of search 1. This slow start to preprocessing also contributes to the larger number of delayed displays for search 1 observed in subjects viewing real-world objects, shown in Table 4.1. Variability in the frequency of display errors likely results from variable fMRI signal properties, requiring differing periods of time for processing, and from other programs competing with the preprocessing and search programs for processor resources, as discussed above.

Overall, display program performance was quite good for subjects viewing Fribble stimuli. Correct stimuli were displayed on at least 90% of trials, and usually more, for each subject, session, and search.

4.3.2 Preprocessing program behavior

The preprocessing program monitors the fMRI output throughout the course of each realtime scan and computes the responses of pre-selected ROIs to stimuli recently shown by the display program. To rapidly convert raw fMRI signal to ROI response values, a small set of preprocessing methods were used to remove scanner and motion effects from small blocks of fMRI data, followed by methods for extracting and summarizing over selected voxel activities. In more typical, i.e., non-realtime, analysis, a larger array of preprocessing methods would be employed over data from the full session to more thoroughly remove signal effects irrelevant to analysis.

While the approach used by my preprocessing program enables performance of realtime analysis, realtime stimulus selection, and realtime search of visual spaces, truncated preprocessing may lead to inaccurate measures of brain region responses, misinforming future search choices. To investigate this potential concern, I compare the correlation between computed ROI responses computed using preprocessing employed during the realtime sessions (Sec. 3.1.4) and the computed responses using "offline" preprocessing considering all runs in a scan session together, and following the drift and motion correction as well as normalization methods of Sec. 3.3.5.

Correcting for subject motion in the scanner is a particular challenge in preprocessing that may be affected by my methodological choices. My preprocessing program aligned fMRI volumes in each time block to the first volume of the current 8.5-minute run, rather than to the first volume recorded in the scanning session. To extract brain region responses for each displayed stimulus, voxel selection is performed based on ROI masks aligned to the brain using the first volume recorded in the scan session (Sec. 3.3.5), under the assumption voxel positions will stay relatively fixed across the session. Significant motion across the scan session could potentially place voxels of interest outside the initially-aligned ROI mask as the session procedes, or cause voxels to be misaligned from their intended weights used in computing the overall ROI stimulus response (Sec. 3.1.4). In my analysis of preprocessing program performance, I track subject motion in each scan session and note its effects on the consistency between responses computed in realtime and offline.

While there were some inconsistencies between responses computed by the two methods, particularly under conditions of greater subject motion, I find realtime computations generally to be reliable across subjects and sessions. This reliability is particularly strong for subjects viewing Fribble objects rather than real-world objects, for reasons considered below.

Subject$_{\text{session}}$	max motion	corr1	corr2	corr3	corr4	average
$S1_1$	8.5	0.56	-0.22	0.63	0.03	0.25
$S1_2$	1.7	0.44	0.21	0.82	-0.19	0.32
$S2_1$	2.2	0.43	-0.06	0.79	0.17	0.33
$S2_2$	1.1	0.41	0.23	0.48	0.47	0.40
$S3_1$	2.1	0.39	0.55	0.71	-0.43	0.31
$S3_2$	9.6	0.63	0.44	0.33	-0.17	0.31
$S4_1$	2.2	0.91	-0.24	-0.59	0.34	0.11
$S4_2$	1.1	0.82	0.23	-0.74	0.20	0.13
$S5_1$	2.0	0.59	-0.37	0.54	0.08	0.21
$S5_2$	1.2	0.71	0.35	0.77	0.20	0.51
$S6_1$	2.3	0.39	0.57	0.16	-0.09	0.26
$S6_2$	2.7	0.69	0.33	-0.07	-0.62	0.08
$S7_1$	3.1	0.09	-0.15	0.74	-0.09	0.15
$S7_2$	2.2	0.64	-0.05	0.62	-0.09	0.28
$S8_1$	2.9	0.19	-0.04	0.77	0.61	0.38
$S8_2$	2.1	0.10	0.10	0.55	0.04	0.20
$S9_1$	2.0	0.70	0.34	0.24	0.10	0.35
$S9_2$	2.2	0.26	0.45	0.55	-0.06	0.30
$S10_1$	1.2	0.40	0.11	0.40	0.34	0.31
$S10_2$	2.1	0.76	0.42	0.63	0.38	0.55

Table 4.3: Motion effects on ROI computed responses for real-world objects searches. Correlation between computed responses for each of four class ROIs using preprocessing on full scan session versus preprocessing on small time blocks within single runs. Average column shows average correlation results across the four ROIs for a given subject and session. Maximum motion magnitude among the starts of all runs also included, pooled from x, y, z (in mm) and yaw, pitch, roll (in degrees).

Real-world objects search

Consistency between ROI responses computed in realtime and responses computed offline for subjects viewing real-world objects are shown in Table 4.3 for each subject, object class, and scan session. Consistency was measured as the correlation between responses of the two methods for each display of each trial.

Correlation values were low but generally positive. 50% of searches produced correlations of 0.3 or above, and 20% produced correlations of 0.5 or above. While realtime and offline processing results are not perfectly consistent, the realtime methods capture desired ROI response trends across each session — generally indicating which stimul evoked particularly high and low cortical activities. Notably, 5 of the 17 searches producing negative correlations showed values below -0.3, pointing to a marked negative trend between the two methods. Consistent misalignment of positive and negative voxel weights when combining voxel activity to form a single regional response to a stimulus may consistenly invert the sign of the computed realtime response. Effects of this inversion on search behavior are considered in Chap. 5.

Correlation values can vary dramatically within a given subject and session across ROIs, e.g., $S6_2$. At first consideration, this within-session variability is quite surprising, as all regions presumably are affected by the same subject movement and scanner drift. However, brain regions differ in the form of the multi-voxel patterns that constitute their response. Patterns the are more broad in spatial resolution, with voxels responding similarly to their neighbors, are less affected in their appearance if subject movement shifts the ROI ˜2 mm from its expected location. High-resolution patterns, in which neighboring voxels exhibit opposite-magnitude responses to a stimulus, are harder to analyze correctly when shifted. Significant angular motion also could produce differing magnitudes of voxel displacement for ROIs closer and farther from the center of brain rotation.

I propose brain motion as a prominent potential source of inconsistency between computed responses. Table 4.3 shows the maximum motion, pooled across translational and rotational di-

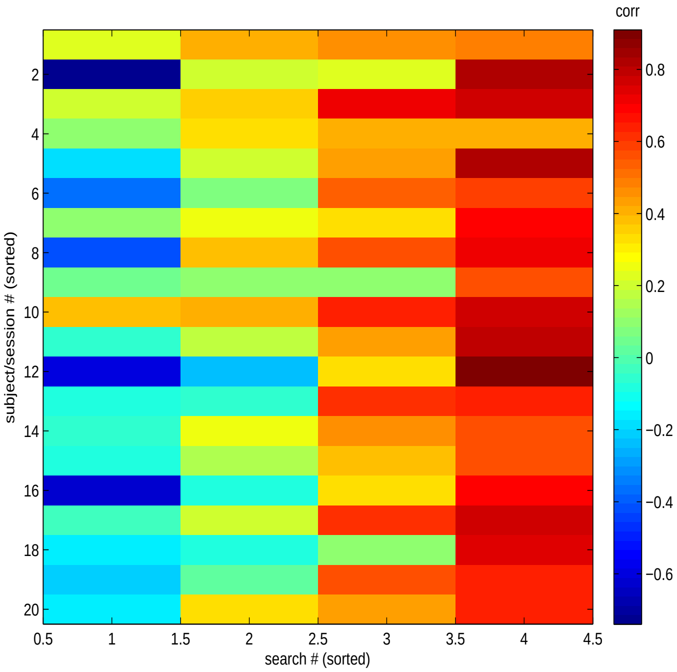


Figure 4.6: Motion effects on ROI computed responses for real-world objects searches, as in Table 4.3. Rows are sorted from lowest to highest corresponding maximum motion magnitude (values not shown), and columns within each row are sorted from lowest to highest correlation values.

mensions, between the start of the scan session and the start of each scan run. I compare only between the starts of scan runs because motion correction computations in the realtime preprocessing program account for further motion between the start and middle of each scan run. For most subjects and sessions (12 of 20), maximum motion falls between 2 and 3 millimeters/degrees in a given direction, while the motion along other directions is usually less than 1 millimeter/degree (non-maximum motion data not shown). Thus, by the end of each session, true ROI locations often shift from their expected locations by a voxel's width in a certain direction. The significant overlap between starting and ending ROI positions lessens my concerns about motion effects, though high-resolution multi-voxel response patterns can produce response computation inconsistencies even from this slight motion, as discussed above.

I expected increased motion would cause increased inconsistency between realtime and offline computations. The expected pattern is weak but apparent when viewing correlation values sorted by subject motion, shown in Fig. 4.6. In this figure, sessions with the least motion are in the top rows and sessions with the most motion are in the bottom rows; colors correspond to correlation values and are sorted from lowest to highest in each row for ease of visualization. Sessions containing two to three searches with low correlation values, corresponding to green and cyan colors, are predominantly seen when there is greater subject motion. However, all sessions contain searches with high correlations, and the search with the most motion, $S3_2$, contains three high-correlation searches.

Regardless of their root cause, inconsistencies in realtime and offline preprocessing are worth noting, and potentially can motivate future development in realtime scanning and search methods. At the same time, responses computed in realtime are relatively reliable across subjects and sessions, following my rather conservative correlation metric. While my method for correlation expresses the consistency between realtime and offline preprocessing results on a trial-by-trial basis, the consistency of computed cortical responses considered for study of ROI selectivity likely is higher. As discussed in Sec. 4.1, ROI responses across the associated visual space are examined only for stimuli shown three or more times. Responses for each of these stimuli are averaged across displays to reduce variability from noise. This noise removal may mimic offline preprocessing effects, increasing the correlation between the two methods' results.

Fribble objects search

Consistency between ROI responses computed in realtime and responses computed offline for subjects viewing Fribble objects are shown in Table 4.4 for each subject, object class, and session. Consistency was measured as the correlation between responses of the two methods for each display of each trial.

Correlation values were low but generally positive, and higher than those observed in the real-times objects searches. 75% of searches produced correlations of 0.2 or above, and more than 50% produced correlations above 0.45. While realtime and offline processing results are

$\text{Subject}_{\text{session}}$	max motion	corr1	corr2	corr3	corr4	average
$S11_1$	1.2	0.50	0.49	-0.55	0.51	0.24
$S11_2$	0.7	0.54	0.48	-0.40	0.12	0.19
$S12_1$	4.8	0.31	0.24	-0.08	-0.04	0.11
$S12_2$	1.2	0.87	0.64	0.67	-0.59	0.40
$S13_1$	2.5	0.56	0.70	0.45	-0.17	0.39
$S13_2$	1.6	0.51	0.68	0.62	-0.10	0.43
$S14_1$	2.4	0.60	0.65	-0.10	0.57	0.43
$S14_2$	1.2	0.39	0.74	-0.01	0.44	0.39
$S15_1$	1.2	0.44	0.53	-0.54	0.23	0.17
$S15_2$	7.0	0.34	-0.07	-0.01	-0.15	0.03
$S16_1$	2.7	0.60	0.72	0.50	0.20	0.51
$S16_2$	1.4	0.84	0.65	0.50	0.20	0.55
$S17_1$	0.7	0.46	0.75	0.37	0.56	0.54
$S17_2$	2.7	0.57	0.71	0.44	0.48	0.55
$S18_1$	2.7	0.59	0.62	0.19	-0.57	0.21
$S18_2$	1.9	0.47	0.54	0.20	-0.67	0.14
$S19_1$	2.0	0.60	0.70	0.69	0.29	0.57
$S19_2$	2.6	0.74	0.60	0.62	0.27	0.56
$S20_1$	1.7	0.59	0.57	-0.14	-0.57	0.14
$S20_2$	1.0	0.62	0.32	0.22	-0.60	0.14

Table 4.4: Motion effects on ROI computed responses. Correlation between computed responses for each of four class ROIs using preprocessing on full scan session versus preprocessing on small time blocks within single runs. Average column shows average correlation results across the four ROIs for a given subject and session. Maximum motion magnitude among the starts of all runs also included, pooled from x, y, z (in mm) and yaw, pitch, roll (in degrees).

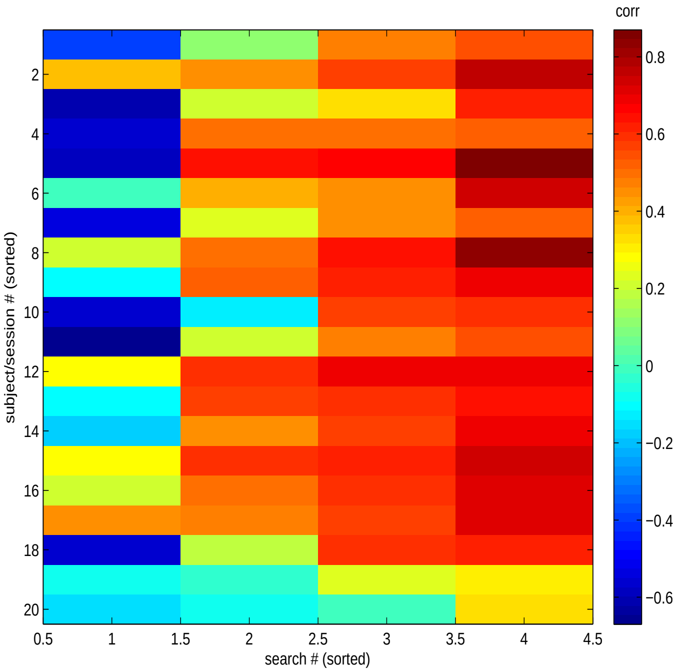


Figure 4.7: Motion effects on ROI computed responses, as in Table 4.4. Rows are sorted from lowest to highest corresponding maximum motion magnitude (values not shown), and columns within each row are sorted from lowest to highest correlation valuesd.

not perfectly consistent, the realtime methods capture desired ROI response trends across each session — generally indicating which stimul evoked particularly high and low cortical activities. 7 of the 17 searches producing negative correlations showed values equal to or below -0.4, pointing to a marked negative trend between the two methods. The mechanism for a consistent inversion in the sign, e.g., +3 becomes -3, of the computed ROI responses is discussed above for subjects viewing real-world objects.

Correlation values can vary dramatically within a given subject and session across ROIs, e.g., S10_2. Potential sources for this variability are discussed above for subjects viewing real-world objects. However, within-session variation is notably less pronounced for subjects viewing Fribble objects, as seen in Fig. 4.7. 12 of the 20 sessions, with each session corresponding to a row in the figure, contain three or four searches with consistently high realtime–offline result correlations.

I propose above that brain motion is a prominent potential source of inconsistency between computed responses. Table 4.4 shows the maximum motion, pooled across translational and rotational dimensions, between the start of the scan session and the start of each scan run. Motion for subjects viewing Fribble objects is generally reduced from that of subjects viewing real-world objects shown in Table 4.3. For 11 of 20 Fribble sessions, maximum motion falls under 2 millimeters/degrees in a given direction, while the motion along other directions is usually less than 1 millimeter/degree. In contrast, 5 of 20 real-world object sessions fit this description. Thus, by the end of each Fribble-viewing session, true ROI locations usually stay within a voxel-width's distance of their expected locations. This decreased motion may be due to the differing tasks performed for the two object types. For real-world objects, subjects were asked to perform a one-back location task in which they were to judge the relative location of consecutively-displayed objects (Sec. 3.3.4). For Fribble objects, subjects were asked to perform a dimness-detection task in which they were to judge whether the object, always displayed in the same central location, was dim (Sec. 3.4.3). Slight movement of real-world objects around the screen may have encouraged slight head motion during stimulus viewings.

Comparing between real-world object and Fribble object viewing groups, there appears to be a relation between subject motion and consistency for realtime and offline computations. Fribble subjects, who moved less as a whole, showed a much higher number of searches with high correlation values, as well as more pronounced negative correlation values for several searches. To consider motion effects within the Fribble sessions, we study correlation values sorted by subject motion, shown in Fig. 4.7. In this figure, there is no clear smooth transition from high (red) to low (green/cyan) correlations with increasing motion (moving from higher to lower rows). However, the two sessions with unusually high motion, $S12_1$ and $S15_2$, contain searches with consistently lower realtime–offline result correlations as shown in the bottom two rows. Even these two sessions contain at least one search with a correlation value above 0.3.

For subjects viewing Fribble stimuli, responses computed in realtime are relatively reliable across subjects and sessions, even under a modest amount of subject motion.

4.3.3 Search program behavior

The search program dynamically selects stimuli to show subjects based on the responses of preselected brain regions to recently viewed stimuli. This stimulus selection process is central to my study of complex visual feature selectivity. Using the simplex simulated annealing method [7] and the spaces I have defined to capture complex visual properties relevant to object perception (Sec. 4.1), the search program is designed to explore the space of visual properties and quickly identify those properties producing the highest response from a pre-selected ROI. These high-response properties correspond to the region's visual selectivity. The search tests each selected location in space by showing a corresponding picture to the subject and then recording and acting on the resulting cortical response.

Presuming each brain region is selective for a single location in its associated search space, and its activity decreases on viewing stimuli drawn at increasing distance from this location, the search program should display two properties:

- **consistency**, i.e., identifying the stimuli in the area of feature space producing maximal response, regardless of the starting point of the search and
- **convergence**, i.e., mostly investigating stimuli near the location producing maximal response, expending little time to investigate more distant stimuli unlikely to evoke a high region response.

Consistency provides confidence in the reliability of search results while **convergence** provides advantage of strategic selection of stimuli over a limited scan time compared to random selection from the full pool of potential images. Sec. 3.5 defines metrics for these two properties in Eqns. 3.6 and 3.7, respectively, and presents a form of permutation test to assess whether these measures would reach their computed values by chance. Beyond the definition of convergence above as focusing on one (or a few) select regions in space across each session, the search is expected to begin by probing broadly around the space and to narrow to its focused location over time as it identifies the maximum-response region in the space. I define two metrics below to measure the temporal evolution of locations explored by each realtime search.

Using the four defined metrics, I observe searches of real-world objects, using SIFT-based space, follow expected behavior for a very limited number of subjects and stimulus classes. Searches of Fribble objects, using their four corresponding search spaces, follow expected behavior significantly more often — in ˜25% of searches executed — though there remain many searches that do not show strong consistency or convergence. Given the relative simplicity of assumptions made about the structure of visual feature spaces (Secs. 3.3.2 and 3.4.1) and about the presence of a single-maximum selectivity for each brain region (Sec. 3.1.5), these results nonetheless constitute a strong start for realtime fMRI search methods exploring complex visual properties used by the brain.

Temporal evolution metrics

I studied movement of the search simplex across space for each stimulus class search and each session by comparing the distribution of locations visited during the first and second half of the session. I characterized these distributions by their mean and variance.

To assess the changing breadth of visual space examined across a search session, I divided the stimulus-points into those visited in the first half of the session and those visited in the second half:

$$\Delta\text{var} = \sum_j \left(\sigma^2(X_2^j)\right) - \sum \left(\sigma^2(X_1^j)\right) \tag{4.1}$$

where $sigma^2(\cdot)$ is the variance function and X_i^j is the set of coordinates on the j^{th} axis for the i^{th} half of the session. Δvar pools variance across dimensions by summing. More fine covariance structure is ignored as the measure is intended to test for overall contraction across all dimensions rather than changes in the general shape of the distribution.

To assess the changing regions within visual space examined across a search session, I again compared points visited in the first half of the session with those in the second half of the session:

$$dist = \sqrt{\sum_j \frac{\left(X_1^j - X_2^j\right)^2}{s_j^2}} \tag{4.2}$$

where X_i^j are as defined for Eqn. 4.1 and $s_j^2 = \frac{\left[\sigma_j^2\right]_1 + \left[\sigma_j^2\right]_2}{2}$ is the mean variance along the j^{th} dimension of the point locations visited in the two halves of the search session. $dist$ measures the distance between the mean location of points visited in the first and second halves of the search session, normalized by the standard deviation of the distributions along each dimension — similar to the Mahalanobis distance using a diagonal covariance matrix. A shift of 0.5 on a dimension with variance 0.1 will produce a larger metric value than a shift of 0.5 on a dimension with variance 1.0.

Real-world objects search

Convergence of realtime searches, i.e., the focus of searches on one or a small number of locations across a session, is shown for real-world object searches in Table 4.5 for each subject, object class, and session. Convergence is assessed based on its Z score; inspection of clustering seen in scatter plot displays (e.g., Sec. 4.1) and in dendrogams led me to set a threshold of $Z \geq 1.8$ for Z scores above chance.

Above-threshold convergence occurred for only 9 of 80 searches performed across all sessions and object classes. 8 of the 9 converged searches were performed for stimulus classes 2 (human-forms) and 4 (containers), with 4 performed for each class. The three programs contributing to realtime search processed stimuli in blocks of two, as discussed in Sec. 3.1.1 and more recently in Sec. 4.3.1, computing in quick succession the next stimuli to display for search 1 and search 2, and then, after a delay, computing in quick succession the next stimuli to display for search 3 and search 4. As a result, when calculation of stimulus choices for a block required more time than expected, they were much more likely to adversely affect the proper display of

$\text{Subject}_{\text{session}}$	z1	z2	z3	z4
$S1_1$	-0.36	1.29	-0.34	**2.14**
$S1_2$	-0.82	1.26	.01	-0.68
$S2_1$	-0.09	0.15	-0.43	0.39
$S2_2$	0.01	0.38	-0.75	0.67
$S3_1$	-1.34	0.77	-0.41	-1.01
$S3_2$	-0.87	**2.60**	0.60	-0.32
$S4_1$	0.30	0.71	-0.35	**2.27**
$S4_2$	-.49	-1.04	-0.08	-0.45
$S5_1$	0.35	-0.08	-1.23	-0.95
$S5_2$	0.52	1.14	-0.32	-0.88
$S6_1$	-0.57	**2.77**	0.79	**2.37**
$S6_2$	-0.01	-1.43	-0.20	**2.58**
$S7_1$	-0.57	**1.95**	-1.01	1.00
$S7_2$	0.11	**1.91**	-0.54	1.30
$S8_1$	**2.23**	0.36	0.07	-0.37
$S8_2$	-1.26	0.14	1.23	0.83
$S9_1$	0.20	0.20	-1.38	-0.93
$S9_2$	-0.15	-0.80	0.05	-0.42
$S10_1$	-1.35	-0.34	-0.42	-1.07
$S10_2$	-0.69	-0.21	-0.18	0.14

Table 4.5: Convergence for searches of real-world objects as measured by Z score metric discussed in Sec. 3.5. Z scores of 1.8 and above in bold.

stimuli for search 1 and search 3, which occured 8 s prior to display for search 2 and search 4. The greater success of searches appearing second in each processing block indicates a long-term advantage to proper stimulus displays throughout the search, as discussed further in Sec. 5.2. However, it is worth noting there are a large number of display errors for search 4 of S6_2, as seen in Table 4.1, despite its high convergence. Motion and preprocessing factors underlying the rare above-threshold convergence results are not apparent. Indeed, only in one session, S6_1, does high convergence occur for two different searches.

Below-threshold convergence Z values ranged widely. Several searches showed values $Z < -1.3$, seeming to indicate that a "random" set of stimuli was markedly more convergent than the stimuli actually visited frequently. To some extent, this phenomenon may point to an unexpected feature of my significance test, defined in Sec. 3.5. Convergence measures the clustering of stimuli visited by the search three or more times, while stimuli visited one or two times are ignored. For my permutation test, I randomly reassign each "frequently-visited" label to one of the stimuli visited any number of times by the search. This approach was intended to judge the convergence of frequently visited stimuli in light of the distribution of stimuli that were visited but not considered sufficiently close to the ROI selectivity center to be re-visited. However, if several stimuli are nearby in space and close to the location producing highest cortical response, their neighborhood may be visited many times but each stimulus visited only visited once or twice. This non-frequently visited clustering may be indicated by extreme negative Z values. However, it is also worth noting convergence Z values did not fall below -2, while the majority of above-threshold values were greater than 2.

Consistency of realtime searches, i.e., the focus of a search on the same location or locations in visual space when initialized at two different points in the space in two different scan sessions, is shown for real-world object searches in Table 4.6 for each subject and object class. Consistency is assessed based on its Z score; inspection of cross-session clustering in scatter plot displays and in dendrograms led me to set a threshold of $Z \geq 1.8$ for Z scores above chance.

Above-threshold consistency occurred for only 2 of 40 searches performed across all subjects

$\text{Subject}_{\text{session}}$	z1	z2	z3	z4
S1	-1.02	**1.80**	-0.39	-0.59
S2	0.34	-1.40	-0.21	-1.78
S3	-1.91	-0.82	1.44	0.04
S4	-0.92	0.10	-1.35	0.44
S5	-1.12	**2.19**	-0.71	0.41
S6	0.20	-0.67	0.86	-0.83
S7	0.21	0.74	0.60	0.21
S8	-0.49	-0.53	1.79	1.35
S9	1.69	-0.33	0.65	-0.91
S10	-1.54	-0.59	.09	-1.36

Table 4.6: Consistency between searches of real-world objects as measured by Z score metric discussed in Sec. 3.5. Z scores of 1.8 and above in bold.

and object classes. The searches were performed for stimulus class 2 (human-forms). From the two above threshold-results, no clear pattern for successful convergence could be deduced. Motion and preprocessing factors underlying the rare above-threshold convergence results are not apparent. Neither of the two subjects, S1 and S5, showed above-threshold convergence for class 2 searches. The lack of consistency for searches with above threshold convergence — particularly for search 2 for S7, which converged in both session but shows a consistency score of $Z = 0.74$ — indicates the potential presence of multiple regions in SIFT-based space producing high responses from a given ROI. Consideration of further sources of difficulty for search performance of real-world objects are discussed in Sec. 5.2.

Below-threshold consistency Z values ranged widely. 6 searches showed values $Z < -1.3$, seeming to indicate a "random" set of stimuli selected for each of two sessions would be markedly more consistent than the stimuli actually visited frequently by the search. Reasons for extreme low Z scores are discussed above in the context of convergence results.

The change in the distribution of locations visited by realtime searches, as reflected by change in the distribution's mean (dist) and variance (∆var), is shown for real-world object searches in Table 4.7 for each subject, object class, and session.

Change in the variance of locations explored from the first half to the second half of each

Subject$_{\text{session}}$	Δ var1	Δ var2	Δ var3	Δ var4	dist1	dist2	dist3	dist4
S1$_1$	0.01	0.01	0.02	-0.01	1.81	0.42	0.69	1.20
S1$_2$	0.01	-0.00	0.01	0.00	0.87	1.11	1.90	0.86
S2$_1$	-0.01	0.00	-0.00	-0.01	1.11	0.74	0.64	0.70
S2$_2$	0.00	-0.02	-0.00	-0.03	0.75	1.08	**2.28**	**2.48**
S3$_1$	-0.00	-0.03	0.01	0.03	1.61	1.14	1.39	**2.29**
S3$_2$	-0.02	0.02	0.01	-0.00	1.14	0.85	1.40	1.72
S4$_1$	0.02	0.01	0.02	0.01	1.49	**2.29**	0.92	1.22
S4$_2$	-0.01	0.03	0.01	-0.01	0.60	1.19	1.61	1.21
S5$_1$	-0.01	0.03	0.00	0.00	0.97	1.55	1.00	1.70
S5$_2$	-0.01	-0.03	0.00	0.02	1.22	**2.55**	1.06	1.45
S6$_1$	-0.01	-0.02	0.02	-0.01	1.37	1.47	1.40	0.65
S6$_2$	0.02	0.04	-0.02	-0.01	1.83	**2.36**	**2.29**	1.65
S7$_1$	-0.00	-0.01	-0.00	0.03	1.20	0.74	1.44	1.29
S7$_2$	-0.01	0.02	-0.01	0.01	1.62	0.80	1.37	1.29
S8$_1$	-0.01	-0.00	-0.00	-0.03	1.40	1.56	1.00	1.36
S8$_2$	-0.01	0.01	-0.03	-0.00	0.80	1.31	**2.54**	0.96
S9$_1$	0.01	0.03	-0.02	-0.01	0.45	0.48	1.12	0.60
S9$_2$	-0.01	0.01	0.00	0.01	1.45	1.57	1.65	1.23
S10$_1$	-0.02	-0.01	-0.03	0.02	1.35	1.04	1.60	1.58
S10$_2$	-0.01	0.01	0.01	0.02	1.65	1.15	**2.61**	1.65

Table 4.7: Temporal evolution of real-world objects searches. $\Delta\text{var}n$ and $\text{dist}n$, corresponding to the change in variance and mean of locations visited in the first and second half of each scan session by the search of stimulus class n, are as defined in Eqns. 4.1 and 4.2, respectively. Distances of 2.0 and greater in bold.

session was quite small across all searches. Δvar generally falls between -0.02 and 0.02, while the variance of locations explored in each half of a session fall between 0.02 and 0.07. Visited points are just as likely to be more dispersed (positive Δvar values) as they are to be more concentrated (negative Δvar values) as the search progresses. The lack of convergence over time as indicated by the Δvar measure in part may reflect the reinitialization of the simplex at the start of each new run within the scan session, as described in Sec. 3.1.5. Existence of multiple locations in search space evoking high cortical responses also may account for lack of convergence over time. In contrast to Δvar, the time-independent convergence measure defined in Eqn. 3.6 can reach high Z values while converging on multiple locations in space, provided the number of locations is small.

Changes in the center of the distribution of locations explored from the first half to the second half of each session is notable for several searches, with $dist \geq 2$ for 9 out of 80 searches and $dist \geq 1.5$ for 24 out of 80 searches. The 9 high shifts in distribution focus occurred with roughly equal frequency for searches of stimulus classes 2, 3, and 4. Most high shifts in focus (7 of 9) occur in the second session. In the second session, the starting locations were selected to be distant from the center of focus from the first session, as discussed in Sec. 3.5.2; in the first session, the starting locations were set to be the origin, around which stimuli are distributed in a roughly Gaussian manner. While this observation indicates a step towards cross-session consistency for several searches, the corresponding Z scores for the consistency metric defined in Eqn. 3.7 are predominantly negative.

Fribble objects search

Convergence of realtime searches, as defined in Sec. 3.5, is shown for Fribble objects in Table 4.8 for each subject, object class, and session. Convergence is assessed based on its Z score, with a threshold of $Z \geq 1.8$ set to be considered a score above chance.

Above-threshold convergence occured for 20 of 80 searches performed across all sessions

$\text{Subject}_{\text{session}}$	z1	z2	z3	z4
$S11_1$	-0.08	0.40	-0.13	**3.90**
$S11_2$	**3.40**	0.18	0.63	-0.38
$S12_1$	1.20	0.56	1.70	0.25
$S12_2$	0.42	1.10	0.51	1.90
$S13_1$	0.91	1.10	-0.43	0.79
$S13_2$	**2.42**	-1.2	1.42	**2.67**
$S14_1$	0.39	1.43	0.43	0.95
$S14_2$	1.45	0.60	0.52	1.40
$S15_1$	**2.76**	1.66	**2.20**	0.18
$S15_2$	1.45	1.69	-0.83	**1.87**
$S16_1$	1.72	**2.10**	**1.80**	0.98
$S16_2$	**2.87**	-1.10	-0.11	-0.22
$S17_1$	-0.47	-0.27	0.89	-0.59
$S17_2$	**2.42**	**2.97**	1.76	0.47
$S18_1$	0.54	1.57	**1.82**	1.72
$S18_2$	1.43	0.93	1.17	**2.30**
$S19_1$	**2.00**	0.93	**3.00**	**4.20**
$S19_2$	0.90	0.86	1.40	**2.10**
$S20_1$	0.77	1.07	**2.86**	**2.84**
$S20_2$	1.24	1.66	0.39	1.41

Table 4.8: Convergence for searches in Fribbles spaces as measured by Z score metric discussed in Sec. 3.5. Z scores of 1.8 and above in bold.

and object classes. Converged searches were performed for all stimulus classes, though more frequently for classes 1, 3, and 4 than for class 2. Higher frequency of delayed displays for search 1 compared to the frequency of delays for other searches did not appear to adversely affect performance of search 1 as it had for subjects viewing real-world objects. In part, this may be attributable to the smaller number of display errors for Fribble object searches overall, especially compared to the number of incorrect real-world stimuli displayed for search 1 and search 3 reported in Table 4.1. Several realtime sessions contained multiple searches with above-threshold convergence; three of four searches converged in session 1 for S19. However, session-specific characteristics, i.e., subject motion, were not apparent underlying factors in successful search convergence within Fribble-viewing subjects.

Above-threshold convergence Z scores generally were higher for Fribble object searches than they were for real-world object searches; 50% of above-threshold Fribble object searches showed $Z \geq 2.5$, compared to 33% of above-threshold real-world object searches. The greater frequency and magnitude of successful search convergence for Fribble objects may reflect the lesser motion of the subjects in these sessions or, potentially related, the seemingly more reliable results of fMRI signal processing during these sessions, reported in Sec. 4.3.2. The structure of the Fribble search spaces also may pose advantages over the SIFT-based image space, as discussed in Sec. 5.2.

Below-threshold convergence Z values still were assigned to 75% of searches, and ranged somewhat widely. However, unlike in real-world objects searches, negative Z values were much more infrequent and were relatively small in magnitude, i.e., $Z > -1.3$. Furthermore, many sub-threshold searches exhibited degrees of convergence, e.g., 22 searches have $1.0 \leq Z \leq 1.8$, compared to 6 searches fitting this criterion for real-world objects sessions.

Consistency of realtime searches, as defined in Sec. 3.5, is shown for Fribble objects in Table 4.9 for each subject and object class. Consistency is assessed based on its Z score, with a threshold of $Z \geq 1.8$ set to be considered a score above chance.

Above-threshold consistency occurred for 7 of 40 searches performed across all subjects and

$\text{Subject}_{\text{session}}$	z1	z2	z3	z4
S11	**2.10**	0.57	0.43	**2.20**
S12	-0.53	1.40	-0.03	1.40
S13	0.46	0.62	-1.20	-1.40
S14	-0.59	-0.19	-1.20	1.22
S15	-1.10	-1.10	1.43	**2.96**
S16	-0.29	0.85	0.39	0.54
S17	**2.28**	**3.14**	**3.28**	-0.99
S18	-1.70	-0.03	0.28	-1.80
S19	1.40	0.30	0.97	**3.80**
S20	0.63	0.15	0.46	0.05

Table 4.9: Consistency between searches in Fribbles spaces as measured by Z score metric discussed in Sec. 3.5. Z scores of 1.8 and above in bold.

object classes. The searches were performed for all stimulus classes, though somewhat more frequently for classes 1 and 4. Several realtime sessions contained multiple searches with above-threshold consistency; three of four searches were consistent for S17. However, session-specific characteristics, i.e., subject motion, were not apparent underlying factors in successful search convergence within Fribble-viewing subjects. Almost all searches showing above-threshold constistency also showed convergence in one scan session, and in both sessions for S19 search 4. Lack of convergence in the both scan sessions for most searches may reflect the search in one session starting close to the location(s) producing highest activity from the selected ROI and converging on the desired region(s) in visual space, while the search in the other session begins probing farther-away locations and searches around more widely, chancing upon the correct areas occassionally but lacking sufficient time to converge.

Below-threshold consistency Z values ranged widely. Several searches show values $Z < -1.3$.

The change in the distribution of locations visited by realtime searches, as reflected by change in the distribution's mean (dist) and variance (Δvar), is shown for Fribble object searches in Table 4.10 for each subject, object class, and session.

Change in the variance of locations explored from the first half to the second half of each

Subject$_{\text{session}}$	Δ var1	Δ var2	Δ var3	Δ var4	dist1	dist2	dist3	dist4
S11$_1$	0.02	0.01	0.05	0.03	1.38	1.15	0.58	1.13
S11$_2$	0.05	0.03	-0.04	0.03	1.42	**2.30**	1.55	1.49
S12$_1$	0.02	0.02	-0.11	0.00	1.38	1.29	**2.18**	**2.25**
S12$_2$	-0.04	0.01	-0.01	0.07	1.03	**2.49**	0.87	**2.28**
S13$_1$	-0.02	0.12	0.03	0.01	**2.39**	0.89	0.99	0.50
S13$_2$	0.10	0.01	-0.03	0.03	1.44	0.60	.84	1.46
S14$_1$	-0.08	-0.01	-0.07	0.03	**3.05**	**2.33**	0.77	1.30
S14$_2$	-0.06	-0.05	-0.03	-0.06	1.68	1.19	1.35	1.95
S15$_1$	0.02	0.01	-0.08	0.08	1.39	1.47	1.89	1.31
S15$_2$	0.06	-0.02	0.01	-0.08	1.83	0.98	1.07	1.57
S16$_1$	-0.00	0.09	-0.02	-0.08	0.52	0.54	1.73	0.95
S16$_2$	0.05	0.05	-0.08	0.05	0.60	1.34	1.17	1.32
S17$_1$	0.01	0.06	0.03	0.07	1.31	1.03	1.16	1.20
S17$_2$	0.03	0.02	-0.08	-0.05	**2.41**	0.81	0.76	1.09
S18$_1$	0.10	-0.03	0.05	0.00	0.25	0.22	0.81	1.07
S18$_2$	0.01	-0.06	0.07	0.00	1.31	1.06	1.09	0.39
S19$_1$	-0.01	-0.03	0.01	-0.02	1.14	1.72	1.42	1.71
S19$_2$	-0.08	0.00	0.01	0.03	**2.12**	0.78	1.97	**2.30**
S20$_1$	-0.03	-0.07	-0.05	0.03	1.93	1.80	**2.59**	1.09
S20$_2$	-0.05	0.00	0.06	0.09	1.34	1.09	0.84	0.85

Table 4.10: Temporal evolution of Fribble searches. Δvar$n$ and dist$n$, corresponding to the change in variance and mean of locations visited in the first and second half of each scan session by the search of stimulus class n, are as defined in Eqns. 4.1 and 4.2, respectively. Distances of 2.0 and greater in bold.

session was small across all searches, equivalent to observations made for search behavior using real-world objects. Δvar generally falls between -0.1 and 0.1. Visited points are just as likely to be more dispersed (positive Δvar values) as they are to be more concentrated (negative Δvar values) as the search progresses. Potential contributions to the lack of convergence over time as indicated by the Δvar measure are discussed above in the context of real-world objects search performance, for which there is a similar lack of observed decrease in variance of stimuli explored over time.

Also similar to real-world objects searches, changes in the center of the distribution of locations explored from the first half to the second half of each session is notable for several searches, with $dist \geq 2$ for 12 out of 80 searches and $dist \geq 1.5$ for 23 out of 80 searches. The 12 high shifts in distribution focus occurred with roughly equal frequency for searches of all stimulus classes. Unlike in real-world objects searches, high shifts in focus occur with equal frequency across the first and second sessions. Starting from the origin in the first session, each search initially probes stimuli whose component shapes are morphed intermediate appearances between two better-established shapes at the extreme -1 and 1 coordinates on each axis. Intuitively, a region involved in object perception — particularly more anterior ROIs shown in Fig. 4.4 — may be specifically selective for a clear circle or a clear star rather than a vague shape newly generated for my stimulus set (Fig. 3.6). Therefore, large shifts from the origin in session 1 may be expected. In contrast, the definition of real-world object feature space through SIFT and multi-dimensional scaling will place groups of salient visual features throughout space, not just at extremes, making large shifts from the origin less likely in the first session. As in real-world objects search, in the second Fribble sessions, the starting locations were selected to be distant from the center of focus from the first session, as discussed in Sec. 3.5.2 — thus, a significant shift in search focus would be required to identify the same stimuli producing high activity in the pre-selected cortical region. While these second session observations indicate a step towards cross-session consistency for several searches, the corresponding Z scores for the consistency metric defined in Eqn. 3.7 are predominantly below threshold, though for all but S11, $Z \geq 1.4$.

All measures of Fribble object search behavior indicate more stability and more consistency in identified visual selectivities when compared with search of real-world objects. However, there remains significant room for improvement to enable convergence, both over space and while operating across time, in many more than 25% of visual selectivity searches. Nonetheless, the current success rate of a relatively simple search method — simplex simulated annealing — to investigate a rather complex probelm in visual encoding constitutes a strong start for realtime fMRI methods in the field.

4.4 Complex visual selectivities

I developed the set of realtime programs for dynamic selection of stimuli to display to a subject in an fMRI machine in order to quickly identify visual properties producing the strongest activity from cortical regions in the ventral object perception pathway. In Sec. 4.2, I report the successful selection of ROIs for study in mid- and high-level visual areas — generally around the fusiform cortex for subjects viewing real-world objects and more broadly spread in fusiform, lateral occipital, and anterior inferotemporal cortex for subjects viewing Fribble objects. In Sec. 4.3, I discuss the generally proper functioning of the search programs used for exploration of visual properties, probed by real-world and Fribble objects. Below, I discuss the visual selectivities revealed through the use of these regions and programs.

I expected the search in each ROI to converge onto one, or a few, location(s) in the associated visual space producing greatest cortical response, corresponding to the regional selectivity. Convergence occurred for only 10% of searches of real-world objects and 25% of searches of Fribble objects. I examine cortical responses observed for these searches, as well as for searches that showed consistency between scanning sessions — i.e., revisiting the same locations in visual space when initialized from two different points in that space. In particular, I use visual inspection of the frequently-visited stimuli, ranked by ROI response size, to intuit important visual properties for each cortical region and use the scatter plot introduced in Sec. 4.1 to visualize cortical activity across visual space (and, to some extent, to observe search behavior). Regardless of the specific visualization used to examine them, the visual feature spaces I have developed provide a powerful new tool for characterizing and understanding cortical responses to complex visual properties. *Examination of points frequently visited by each search and the responses of the corresponding brain regions revealed multiple distinct selectivities within search of single ROIs, marked change in cortical response resulting from slight deviations in visual properties/slight changes in location in visual space, and several intuitive classes of visual properties used by the ventral pathway.*

When a search fails to show convergence or consistency, there is less confidence that the stimuli used by the search sufficiently captured cortical response properties across the space of visual features. Nonetheless, cortical response data was gathered from every ROI examined by a search, and this data provides a partial view into complex visual properties used in the ventral pathway. I compare patterns of activities from all searches to identify potentially smooth evolution of selectivities across the brain. This expected evolution was not apparent in my results, likely because the regions explored within each subject tended to be anatomically distant from one another and regions anatomically close across subjects reflect variability in cortical visual encoding across subjects.

4.4.1 Metric for cross-region comparisons

Beyond exploring the visual property selectivities for single ROIs in individual subjects, I study the generalization of selectivities across subjects and selectivity evolution as one moves across the cortex. To this end, I compared the distribution of high- and low-response regions in SIFT space for anatomically proximal ROIs within and across subjects. I expected nearby ROIs to have similar response profiles and distant ROIs to have distinct profiles. Instead, nearby ROIs generally were quite different in selectivities, both within and across subjects.

Given the sparse sampling of SIFT space by the search method — made more sparse by the removal of unreliable responses estimated over “too few,” i.e., fewer than three, stimulus repetitions — comparison of response profiles across ROIs poses its own challenge. I interpolate over the sparse samples using inverse distance weighting [58] to construct a four-dimensional grid of responses for each ROI and each search session. The grid spans between -1 and 1 on each axis, divided into 0.05 length intervals. Similarity between a pair of response profiles is computed using the Pearson correlation of the grid entries for the two ROIs.

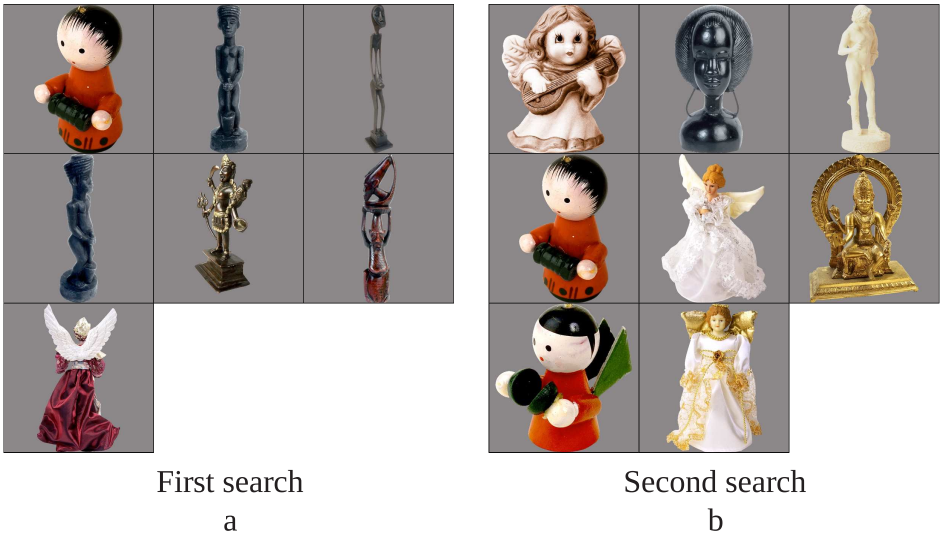
Cross-region comparisons failed to show clear patterns of similarity in regional selectivities related to anatomical proximity of the compared regions.

4.4.2 Real-world objects search

Among subjects viewing real-world objects, 9 selectivity searches converged and 2 searches showed consistency across searches, as measured in Sec. 4.3.3. Examination of points frequently visited by each search and the responses of the corresponding brain regions revealed multiple distinct selectivities within search of single ROIs, marked change in cortical response resulting from slight deviations in visual properties/slight changes in location in visual space, and several intuitive classes of visual properties used by the ventral pathway — including surface texture as well as two- and three-dimensional shape.

The two searches with highest Z score convergence values for object class 2 (human-forms) were performed in session 2 for S3 and in session 1 for S6; the two searches with highest Z score convergence values for obect class 4 (containers) were performed in the two sessions for S6, as reported in Table 4.5. Object class 3 had no searches showing above-threshold Z score convergence values, and the one above-threshold Z score for object class 1 was below those of the searches in class 2 and class 4 mentioned above. I examine the results of the four “most-converged” searches in detail, and summarize results for all other searches with above-threshold convergence.

The **class 2/human-forms search** in the second session for **S3** showed one of the greatest



First search
a

Second search
b

Figure 4.8: Stimuli visited three or more times in searches for S3, class 2 (human-forms). Images sorted in order of decreasing ROI response, averaged across all trials for each image.

convergence measures ($Z = 2.60$). The scatter plot view of this search was presented in Sec. 4.1 in Fig. 4.1. I return to analysis of this figure in the present discussion, in addition to analysis of the relative ROI response sizes for the stimuli frequently searched, shown in Fig. 4.8. Projecting the visited stimuli along the first two dimensions in SIFT-based space in Fig. 4.1a, and focusing on frequently-visited stimuli (for reasons discussed in Sec. 4.1), we see two clusters on the top left and the middle left for the second session (red and blue diamonds). The images visually are split into two groups[2]: one group containing light/generally-narrow-shape (the first, third, fifth and eighth highest response stimuli) and the second group containing less-light/wide-shape (the remaining stimuli), as shown in Fig. 4.8b. Notably, stimuli evoking high and low responses appear in both clusters, and similar-looking images can elicit opposite ROI activities — e.g., the two red characters.

The class 2 search in the first session for S3 shows a quite weak convergence measure ($Z = 0.77$). Projecting the visited stimuli along the first two dimensions in SIFT-based space in Fig. 4.1a appears to reveal a concentration of frequently-visited stimuli (red and blue circles)

[2] For the interpretation of real-world objects results, grouping was done by visual inspection of a single linkage dendrogram constructed in the four-dimensional SIFT-based space.

at the bottom of the first two dimension of the space, but the stimuli are broadly spread out horizontally and along the additional two dimensions shown in Fig. 4.1b. Unlike results for the second session, there is no concentration of focus around one (or two) spatial locations. Despite a very low consistency Z score ($Z = -0.82$), there is evidence for a degree of consistency between session results. The stimuli evoking the strongest and weakest responses in the first session appear in the lower cluster of visited points in the second session. The red wingless character, again, elicits high response while the purple winged character in the first session and the red-green winged character in the second session, nearby in visual SIFT-based space, elicit low responses. The winged character in the first session is projected as a very small blue circle at $(-0.05, 0.02, 0.15, 0.10)$ in the SIFT space in Figs. 4.1a and b. By starting from a separate location, the second search finds the highest-response spatial neighborhood for the ROI found by the first search, but also finds a second local ROI-response maximum in SIFT space. This partial consistency between searches is too weak by my metric to earn a notably non-random Z score in Table 4.6. Unfortunately, these nuances are not fully captured in the Z score metric used in my study, as defined in Sec. 3.5.

The class 2 search in the first session for **S6** showed the greatest convergence measure across all searches ($z = 2.77$). Projecting the visited stimuli along the SIFT dimensions in Fig. 4.9, we see one cluster (of red and blue circles) around the coordinates $(-0.1, -0.15, 0, -0.15)$ and several outliers for the first session. The members of the central cluster are the images producing the second, third, fourth, and seventh highest responses from the ROI, as ordered in Fig. 4.10a. The three stimuli in the cluster producing the highest responses may be linked by their wide circular head/halo, while the smallest-response stimulus is notably thin — potentially indicating response intensity as a wide/thin indicator. Stimuli evoking high and low responses, coming from the two ends of the wide/thin spectrum, are nearby one another in the part of the SIFT space under study by the search.

The class 2 search in the second session for S6 shows a quite weak convergence measure ($Z = -1.43$). Projecting the visited stimuli along the first two and second two dimensions in

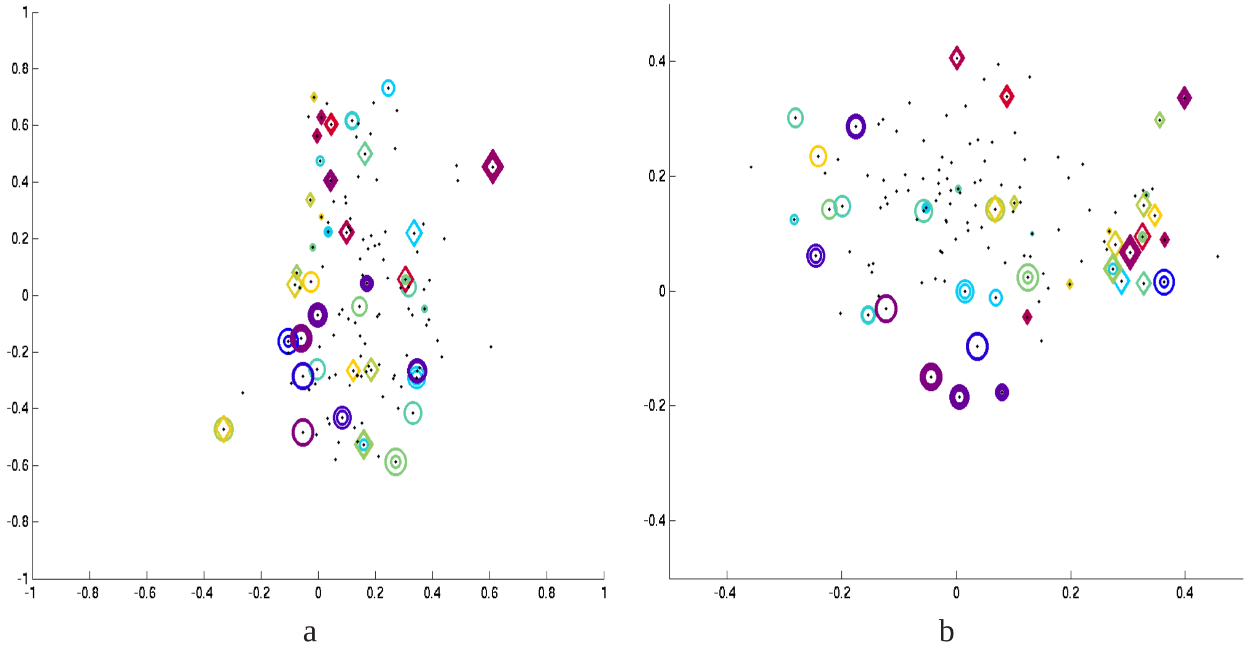
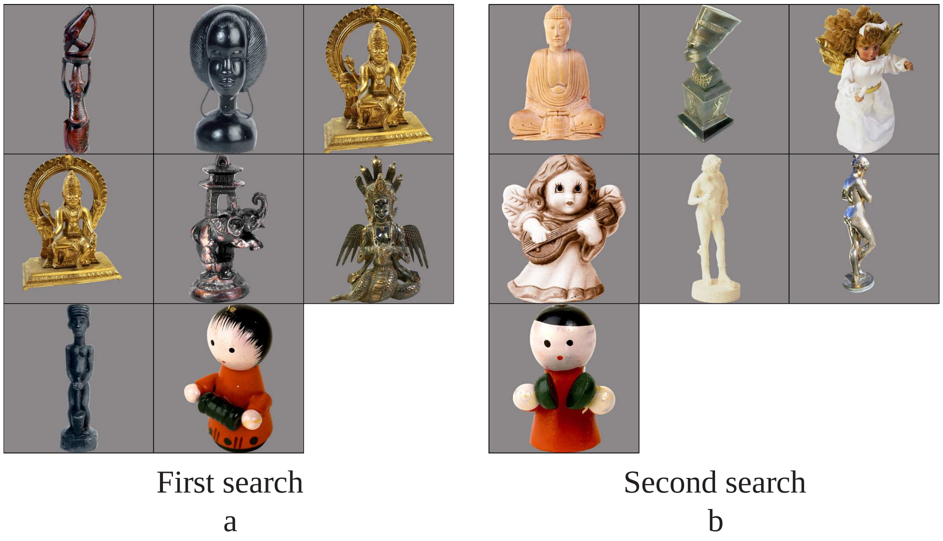


Figure 4.9: Search results for S6, class 2 (human-forms), shown in (a) first and second SIFT space dimensions and (b) third and fourth dimensions. Colors and shapes used as in Fig. 4.1


First search
a

Second search
b

Figure 4.10: Stimuli visited three or more times in searches for S6, class 2 (human-forms). Images sorted in order of decreasing ROI response, averaged across all trials for each image.

Fig. 4.9a and b appear to show a similar structure of one cluster with additional outliers in each view, but careful inspection reveals the points clustered together in the first two dimensions are split apart along the second two dimensions, and outliers in each view are farther flung across the space, supporting the low convergence measure. Similarly, as the consistency Z score is a low $Z = -0.67$, the stimuli frequently visited in the second session fail to overlap with similar feature space locations and "similar-looking"[3] stimuli frequently visited in the first session. Although a red character produces the minimum responses in each of the two searches (Fig. 4.10), the two characters are located in distinct corners of the SIFT space (dark red diamond and blue circle in Fig. 4.9).

Comparison of searches for S3 and S6, in Figs. 4.1 and 4.9, respectively, shows a similar pattern of visited stimuli in the feature space. For both subjects, there is a focus close to the first dimensional axis, i.e, a vertical line of red and blue circles and diamonds along the first two dimensions; visited stimuli follow a V pattern in the second two dimensions. Furthermore, some of the highest ROI response stimuli appear (in red) at high locations along the second and third dimensions. Similarly, frequently-visited stimuli for S6 session 1 (dark blue circles) appear close to the the observed lower cluster for S3 session 2, though the cortical response sizes for the two subjects appear to differ. Comparing Figs. 4.8 and 4.10, we also can confirm a degree of overlap between the images frequently shown for each subject. In both subjects, frequently visited stimuli seemed to show regional selectivity, and potentially differentiation, for narrow-versus-wide shapes. While searches for the two subjects show great similarities, it is worth noting the ROIs studied, labeled as c2 and f2 in Fig. 4.3, are anatomically distant from one another. Thus, different subjects achieve similar coding strategies in different areas of their brains.

The **class 4/containers search** in sessions 1 and 2 for **S6** showed the third and fourth greatest convergence measures ($Z = 2.37$ and $Z = 2.58$, respectively) across all searches. Projecting the

[3] Similarity in appearance is not well-defined, as explored by my related work in Chap. 2. Generally, I limit my similarity judgements to identification of identical pictures, e.g., in Fig. 4.8. Here, I occasionally use more rough intuition.

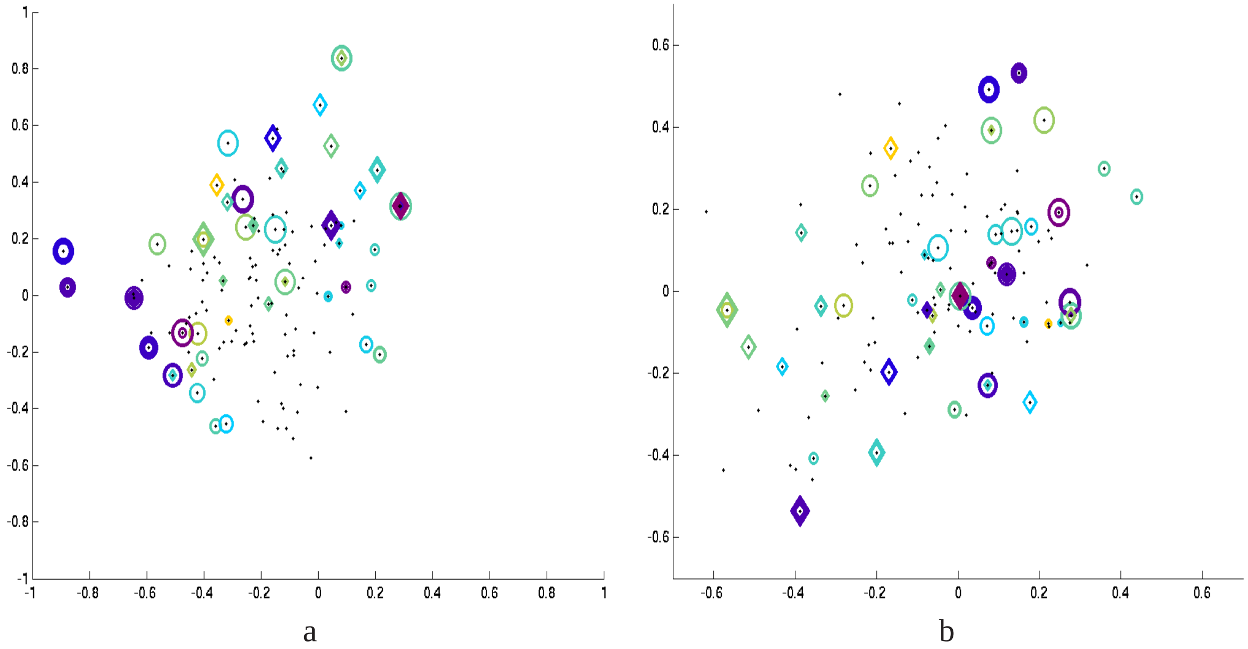
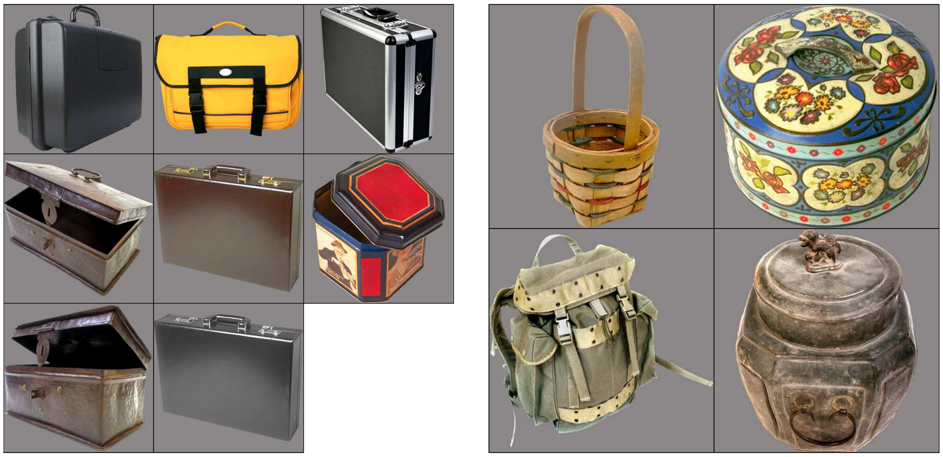


Figure 4.11: Search results for S6, class 4 (containers), shown in (a) first and second SIFT space dimensions and (b) third and fourth dimensions. Colors and shapes used as in Fig. 4.1.



First search
a

Second search
b

Figure 4.12: Stimuli visited three or more times in searches for S6, class 4 (containers). Images sorted in order of decreasing ROI response, averaged across all trials for each image.

visited stimuli along the SIFT dimensions in Fig. 4.11, we see two clusters for the first session (red and blue circles), centered around $(-0.9, 0.1, 0.17, 0.55)$ and $(-0.55, -0.15, 0.15, -0.15)$, respectively, in addition to a few outliers. The members of the first cluster are the images producing the fifth and eighth highest responses from the ROI, and the members of the second cluster are the images producing the first, fourth, sixth, and seventh highest responses, as ordered in Fig. 4.12a. While the members of the first group have clear visual similarities, differing only by their color, the members of the second group are much more assorted. Perhaps they are linked by their multiple long sides and by the illumination focused on corners rather than edges. Stimuli within the same group draw both high and low activities from the ROI. The two partially-open hinged boxes produce mid-sized and low responses, indicating notably different responses to slight changes in complex visual properties.

Projecting the stimuli in the second session (red and blue diamonds) appears to show less strong grouping, though the convergence metric Z score is larger. The stimuli evoking the first and third strongest responses are grouped very closely together, overlaping in the first two dimensions and surrounding the origin in the second two dimensions, $(0.35, 0.35, 0, 0)$, in Fig. 4.11. The remaining two stimuli are outliers that appear relatively close in the first two dimensions, but lie farther to the bottom right in the second two dimensions. It is unclear by visual inspection why two nearby stimuli are grouped together, though their top handles and predominance of horizontal and vertical lines may underlie their grouping. Similar to the other search results observed above, the two nearby stimuli evoke a particularly high and low response, respectively, showing regional sensitivity to a slight change in position in visual feature space — though the visual differences between the stimuli is rather apparent. As the consistency Z score is a low $Z = -0.83$, the stimuli visited in the two sessions fail to overlap in location and in intuitive visual appearances, though almost all stimuli visited in both searches have small handles on top — a common but not essential characteristic of the class as shown in Fig. 3.2.

Visual comparison of searches for class 2 and class 4 in, e.g., Figs. 4.9 and 4.11, respectively, show distinct patterns of visited stimuli in the feature space. Stimuli frequently visited by class

4 searches are spread to greater extremes along the first and third dimensions of the space while stimuli frequently visited by class 2 searches are spread to greater extremes along the second and fourth dimensions. These search behavior in part reflect the differing spreads of stimuli in each class across the space, as shown by the black dots in each scatter plot.

Study of frequently visited stimuli in the search sessions showing lower, but above-threshold, convergence reveals a mix of results.

- The **class 2/human-forms search** in sessions 1 and 2 for **S7** showed high convergence measures ($Z = 1.95$ and $z = 1.91$, respectively). The first session frequently visited a cluster of stimuli evoking the third, fourth, fifth, sixth, and seventh (out of seven) highest ROI responses. The stimuli in this cluster appear to be grouped by high spatial frequency details, and particularly frequent shiny spots, across the surface. Stimuli close together in the visual space evoked particularly high and low ROI responses. The second session frequently visited points forming one cluster, but there is no clear visual grouping among these stimuli. Stimuli close together in the visual space evoked high (first highest) and low (eighth and tenth out of eleven) ROI response. As the consistency Z score is a low $Z = 0.74$, the stimuli visited in the two session fail to overlap in location and in intuitive visual appearances.
- The **class 4/containers search** in the first session for **S1** showed high convergence ($Z = 2.14$). The search frequently visited a cluster of stimuli evoking the second, third, fourth, and fifth (out of seven) highest ROI responses. There are no clear visual properties linking these images. Notably, the stimuli evoking the highest and lowest responses fall outside the converged cluster.
- The class 4 search in the first session for **S4** showed high convergence ($Z = 2.27$). The search frequently visited a cluster of stimuli evoking the first, second, third, fourth, and eigth (out of eight) highest ROI responses. The two stimuli closest in space (evoking the second and fourth highest responses) appear to be linked by surface texture, earth-tone

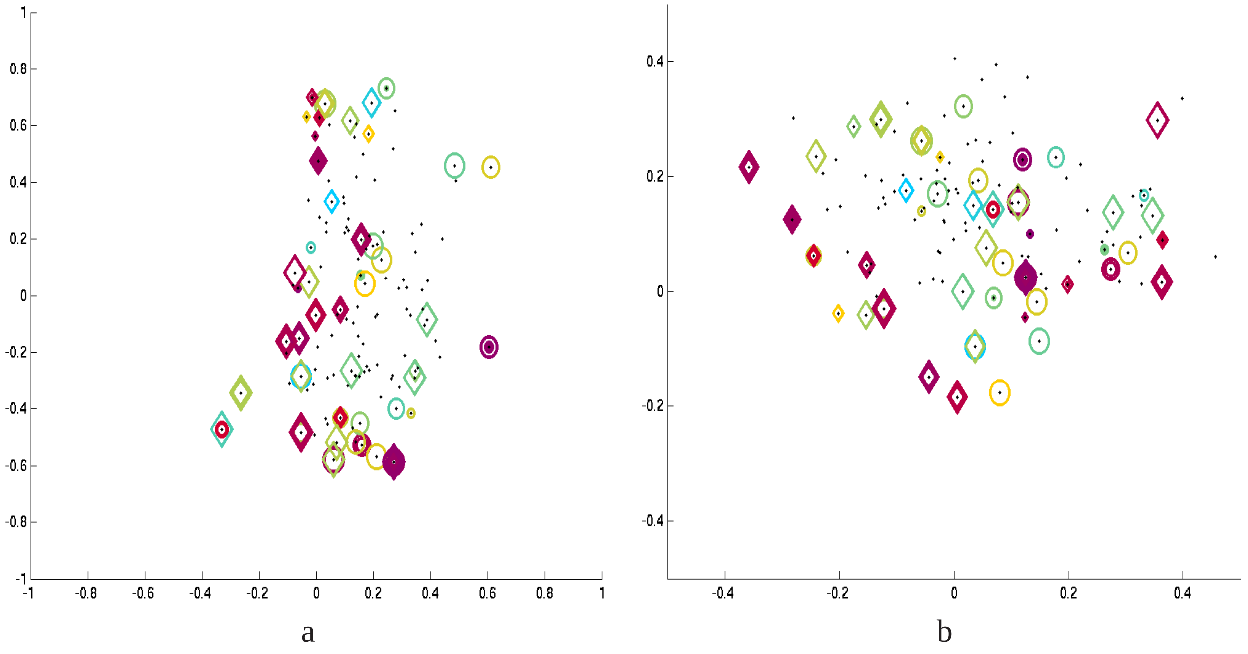


Figure 4.13: Search results for S1, class 2 (human-forms), shown in (a) first and second SIFT space dimensions and (b) third and fourth dimensions. Colors and shapes used as in Fig. 4.1.

colors (although the SIFT representation space operates on black-and-white versions of the images), and similar handle shapes; visual patterns across all five clustered stimuli are not apparent. Stimuli clustered together in the visual space evoked the highest and lowest ROI responses.

The two searches with cross-session search consistency were performed for object class 2 (human-forms) for S1 and S5. Object classes 1, 3, and 4 had no searches showing above-threshold Z score consistency values values. I examine the results of the two consistent search pairs below.

The **class 2/human-forms searches** for **S1** showed a high consistency measure across the two scan sessions ($Z = 1.80$). Projecting the visited stimuli along the SIFT dimensions in Fig. 4.13, we see stimuli frequently visited in the second session (red and blue diamonds) are spread widely across the SIFT space. Half of the stimuli frequently visited in the first session (red and blue circles) are focused around $(0.15, -0.55, 0.15, 0.1)$, in the same area as the stimulus producing the lowest ROI activity in the first session. The three stimuli in the first session exploring the

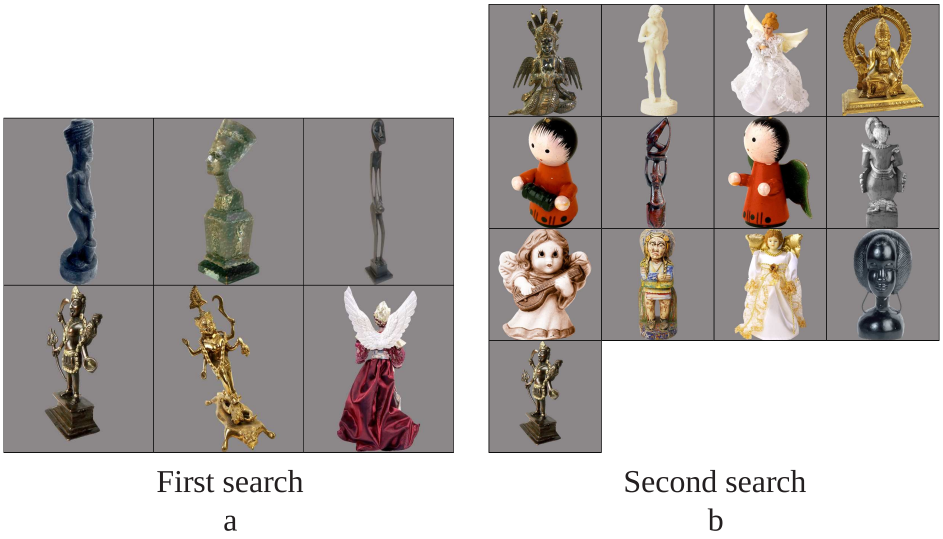


Figure 4.14: Stimuli visited three or more times in searches for S1, class 2 (human-forms). Images sorted in order of decreasing ROI response, averaged across all trials for each image.

same location as the second session were those producing the second, third, and fourth highest responses, as shown in Fig. 4.14a. While, based on their ordering, these first-session stimuli appear to be producing mid-level responses, distinct from the "lowest" response produced by the metal samurai stimulus of interest in the second search, comparison of absolute computed values within each scan session shows all four responses to be quite low, further indicating similar activity discover in both searches. It is difficult to intuit the visual properties grouping these stimuli together, though they may be linked by their metalic surfaces, and by their rectangular bases and heads. Reviewing the cross-session grouping in Fig. 4.13, In summary, we see my consistency measure will award one search focusing its efforts on a single location frequently visited by a second search — potentially indicating the need for further modifications to my consistency metric.

Comparison of class 2 searches for S1, S3, and S6, in Figs. 4.13, 4.1, and 4.9, respectively, shows a similar pattern of frequently visited stimuli in feature space. There is a vertical line of stimuli along the first two dimensions and a V pattern in the second two dimensions. Some of the highest ROI response stimuli appear at high locations along the second and third dimensions for

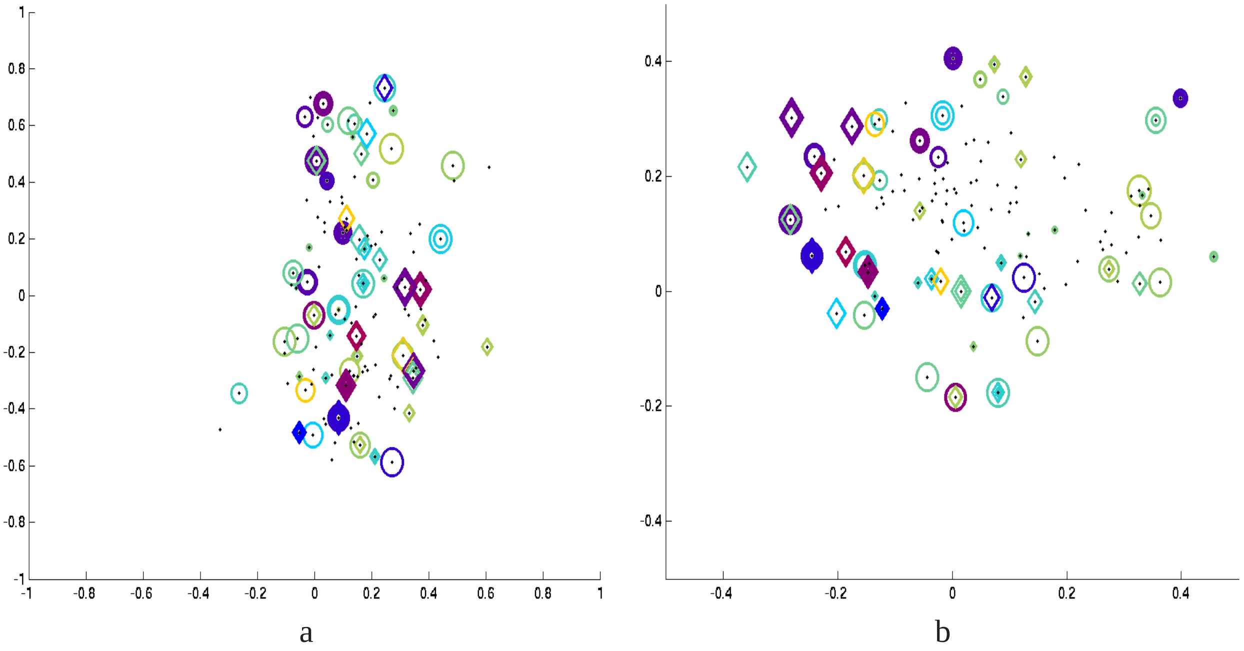
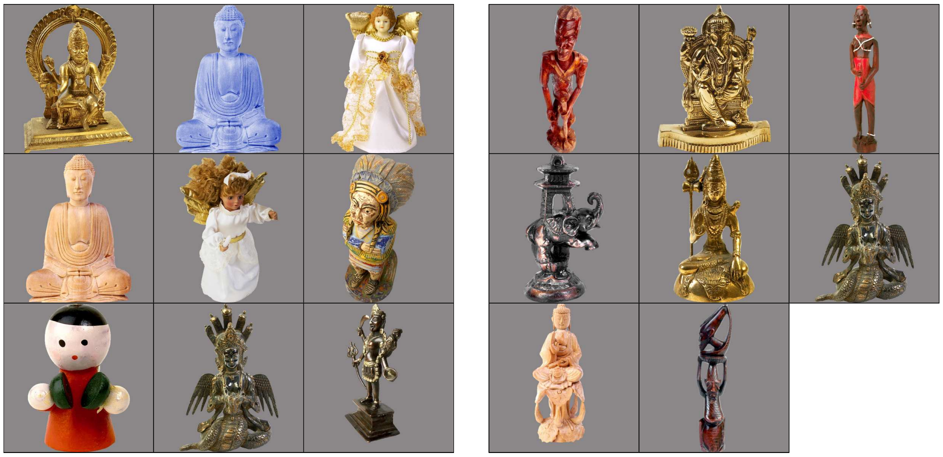


Figure 4.15: Search results for S5, class 2 (human-forms), shown in (a) first and second SIFT space dimensions and (b) third and fourth dimensions. Colors and shapes used as in Fig. 4.1.

S1 session 2, S3 session 2, and S6 session 2. Notably, the 4 mostly-white figures from this region frequently displayed to S3 in session 2 also are frequently displayed to S1 in session 2; 3 of the 4 figures are sorted in the same order based on ROI response size, as shown in Figs. 4.8 and 4.14. Examining the relative anatomical location of the ROIs studied, labeled as a2, c2, and f2 in Fig. 4.3 for S1, S3, and S6, respectively, we see the ROIs for S1 and S6 are very close to each other when projected on the Talairach brain, though S3's ROI is more distant.

The class 2 searches for **S5** showed a high consistency measure across the two scan sessions ($Z = 2.19$). Projecting the visited stimuli along the SIFT dimensions in Fig. 4.15, we see stimuli frequently visited in the first session (red and blue circles) are spread across the space. Quite similar to the S1 searches for object class 2, half of the stimuli frequently visited in the second session (red and blue diamonds) are focused around $(0.05, -0.45, -.025, 0.05)$, in the same area as the stimulus producing the second lowest ROI activity in the first session. The four stimuli in the second session exploring the same location as the first session were those producing the first, fourth, sixth, and eighth largest responses, spanning high and low response values. The



First search
a

Second search
b

Figure 4.16: Stimuli visited three or more times in searches for S5, class 2 (human-forms). Images sorted in order of decreasing ROI response, averaged across all trials for each image.

visual properties grouping these stimuli together appear to be the presence of sharp local angles defining internal holes or feathers in the shape. Further inspection of the SIFT space in Fig. 4.15 shows the two search sessions explore largely different regions along the first two dimensions, though there appear to be closer overlaps along the second two dimensions.

Comparison of class 2 searches for S5 with those of the subjects reported above, S1, S3, and S6, shows a great degree of difference in the pattern of frequently visited stimuli in feature space and in the pattern of cortical responses across space. This finding reflects the expected diversity of selectivities employed in perception of a given object class, e.g., human-forms.

The high consistency Z score despite relatively limited overlap in results between two sessions may point to the need for modifying the metric defined in Eqn. 3.7. The metric only considers clusters of points in SIFT space containing stimuli visited by both searches. If most visited stimuli are dispersed throughout the space, each one will be defined as its own cluster and its presence will not effectively decrease the metric value through the $\ell1$ term, which will offset the $\ell2$ term for small clusters. If there is a suffficiently large cluster containing only one point from one search and several points from the other search, this will be considered strong consis-

$(\mathrm{Sn_m},\mathrm{Sp_q})$	corr	Tdist	$(\mathrm{Sn_m},\mathrm{Sp_q})$	corr	Tdist	$(\mathrm{Sn_m},\mathrm{Sp_q})$	corr	Tdist
$(\mathrm{S2_1},\mathrm{S4_2})$	-0.37	1.5	$(\mathrm{S5_4},\mathrm{S9_4})$	-0.55	1.5	$(\mathrm{S1_2},\mathrm{S6_2})$	-0.49	1.6
$(\mathrm{S5_2},\mathrm{S8_1})$	0.42	2.0	$(\mathrm{S3_3},\mathrm{S5_1})$	0.60	1.3	$(\mathrm{S2_4},\mathrm{S6_1})$	0.23	2.3
$(\mathrm{S2_3},\mathrm{S4_4})$	**0.86**	2.6	$(\mathrm{S1_3},\mathrm{S3_4})$	0.45	2.8	$(\mathrm{S2_1},\mathrm{S8_2})$	0.42	1.7
$(\mathrm{S7_2},\mathrm{S7_4})$	-0.48	3.1	$(\mathrm{S3_3},\mathrm{S3_4})$	**0.82**	3.2	$(\mathrm{S2_3},\mathrm{S2_4})$	-0.31	3.6
$(\mathrm{S8_2},\mathrm{S8_3})$	0.72	3.6	$(\mathrm{S1_1},\mathrm{S1_3})$	-0.25	4.5	$(\mathrm{S4_1},\mathrm{S4_4})$	0.54	4.5

Table 4.11: Correlation of activation profiles in SIFT-based space for anatomically proximal ROIs in real-world objects searches. ROIs selected to be the closest pairs in Fig. 4.3 in first three rows, closest pairs within same subject in last two rows. The corr value for $(\mathrm{Sn_m}, \mathrm{Sp_q})$ corresponds to the maximum-magnitude correlation for the class m ROI for subject n and the class q ROI for subject m across search session pairs — e.g., (sess1,sess1), (sess2,sess1). corr ≥ 0.8 in bold. Tdists is distance (in voxels) between the two ROI centers in the Talairach brain.

tency, even if that was not intended when I designed the metric. Indeed, these are the results I observe above.

Comparison of neighboring regions

Profile correlations are distributed roughly as a Gaussian distribution around 0, with a standard deviation of 0.37. ROIs with high correlations, e.g., $r \geq 0.8$, show marked similarities on visual inspection, such as the pair of interpolated response profiles shown in Fig. 4.17 for S2, object class 3, session 2 and S4, object class 4, session 1. However, anatomically proximal ROIs tend to have less similarity, seen across multiple pairwise comparisons in Table 4.11 and for an individual pairwise comparison in Fig. 4.18. We can see that $r = 0.42$ is associated with a pair of profiles retaining partial similarities in peak positive and negative activity locations, with further differences in broader surface details. The pair $(\mathrm{S5_2},\mathrm{S8_1})$ is considered here as it is the pair of brain regions across closest to each other across all subjects with a non-negative profile correlation.

Looking within individual subjects, the ROIs are slightly more separated in the brain, as indicated in Fig. 4.3, but generally have larger correlation magnitudes than the closer region pairs across subjects. Most correlations remain below $r = 0.8$, but may indicate a weak trend. The presence of some moderately high negative correlations within and across subjects, particularly

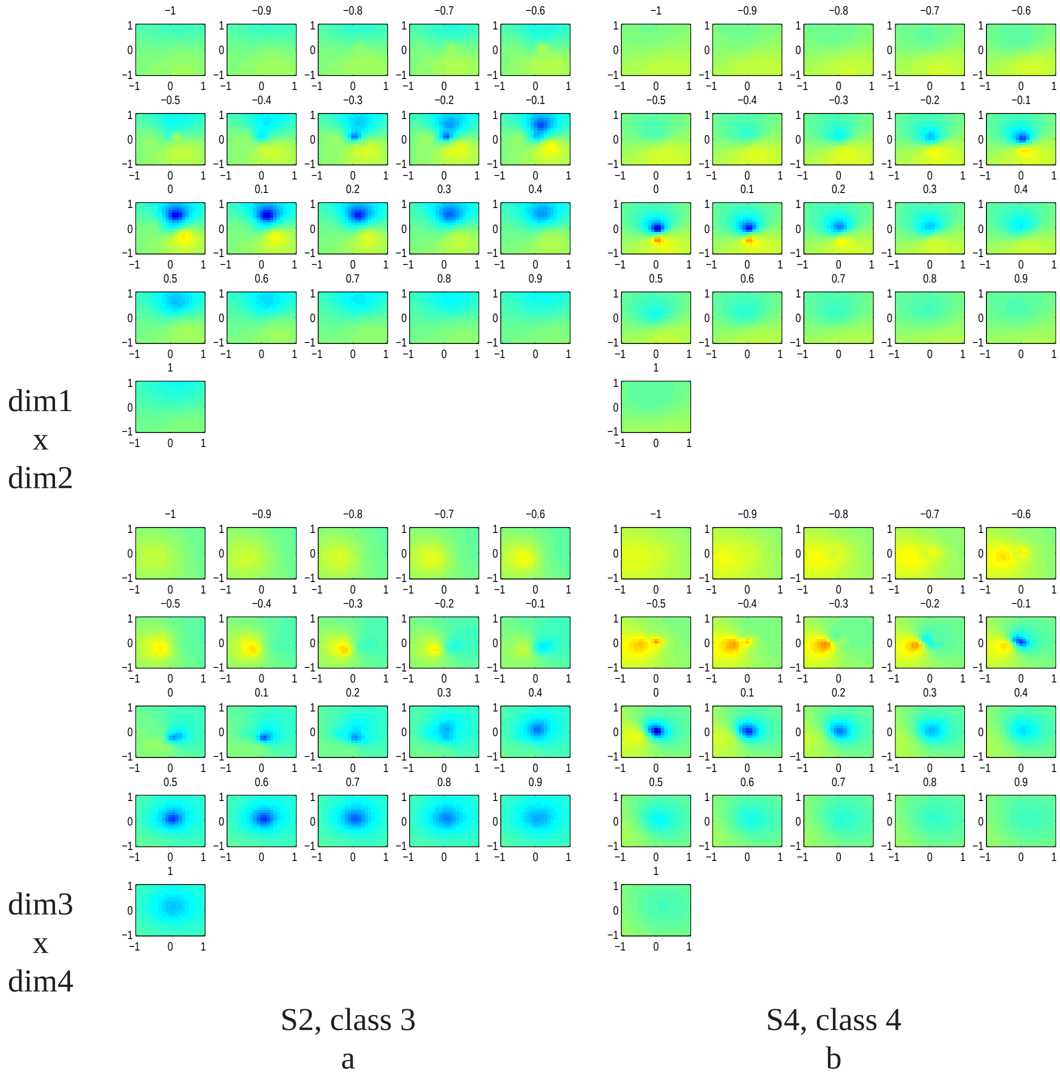


Figure 4.17: Interpolated activation profiles for two ROIs, with correlation $r = 0.86$. Responses normalized for each profile with maximum value red and minimum value blue. dim1 x dim2 slices taken along first and second dimensions, varied along third dimension and fixed at fourth dimension to 0. dim3 x dim4 slices taken along third and fourth dimensions, varied along first dimension and fixed at second dimension to 0.

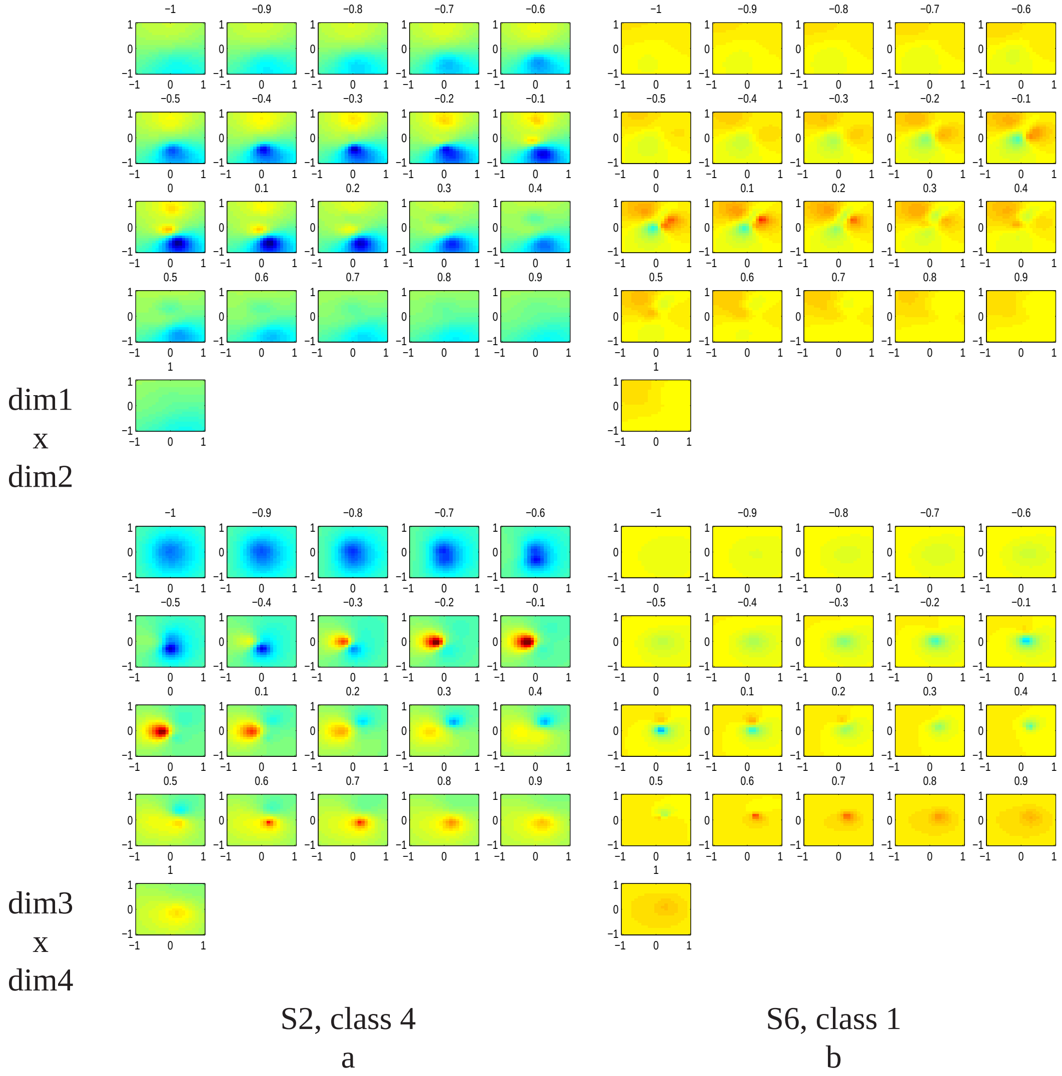


Figure 4.18: Interpolated activation profiles for two ROIs, with correlation $r = 0.42$. Same cross-section display method employed as in Fig. 4.17.

$r = -0.55$ for the $(\mathrm{S5}_4, \mathrm{S9}_4)$ pair, may be similar in principle to the close proximity of maximum and minimum activity stimuli in the SIFT visual feature space. Interpolated profiles such as in Fig. 4.17 and scatter plots such as in Fig. 4.1 show nearby regions in feature space can evoke extremely polar opposite cortical responses.

4.4.3 Fribble objects search

Among subjects viewing Fribble objects, 20 selectivity searches converged and 7 searches showed consistency across searches, as measured in Sec. 4.3.3. As in real-world object searches, examination of stimuli frequently visited by each search and the responses of the corresponding brain regions revealed multiple distinct selectivities within search of single ROIs, marked change in cortical response resulting from slight deviations in visual properties/slight changes in location in visual space, and several perception approaches used by the ventral pathway — including focus on the form of one or multiple component "appendages" for a given Fribble object.

The search with the highest Z score convergence value for object class 1 (curved tube object, see Fig. 3.6) was performed in session 2 for S11; the search with the highest value for object class 4 (wheelbarrow object) was performed in session 1 for S19, as reported in Table 4.8. The highest Z score convergence values for searches of object classes 2 (blue-bodied, yellow-winged object) and 3 (bipedal, metal-tipped tail object) were below those of the searches in class 1 and 4. The two searches with highest cross-session search consistency were performed for object class 3 for S17 and for object class 4 for S19. I examine in detail the results of the two searches listed above as most-convergent for their respective object classes as well as the results of the two most consistent search pairs (noting S19 is both most consistent and most convergent). I also summarize results for all other searches with above-threshold convergence and consistency.

The **class 1/curved tube object search** in the second session for **S11** showed high convergence ($Z = 3.40$). Projecting the visited stimuli along the three Fribble-specific morph dimensions in Fig. 4.19, noting the third dimension is indicated by diagonal displacement, we see one

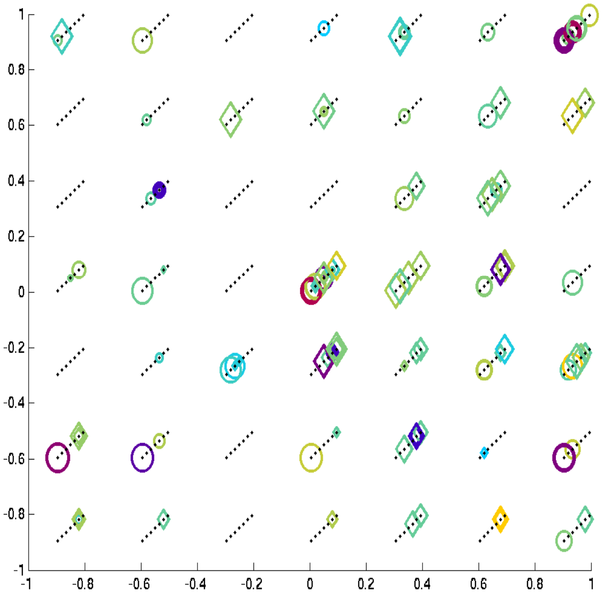


Figure 4.19: Search results for S11, class 1, shown in three-dimensional Fribble space, with third dimension represented as diagonal offset. Positive third dimenion results in displacement up and to the right. Location of all potential stimuli in space shown as black dots. Results from realtime scan session 1 are circles, results from realtime scan session 2 are diamonds. For stimuli visited three or more times, colors span blue–dark blue–dark red–red for low through high responses; for stimuli visited one or two times, colors span cyan–yellow–green for low through high responses. Size of shape corresponds to time each point was visited in search, with larger shapes corresponding to later points in search.

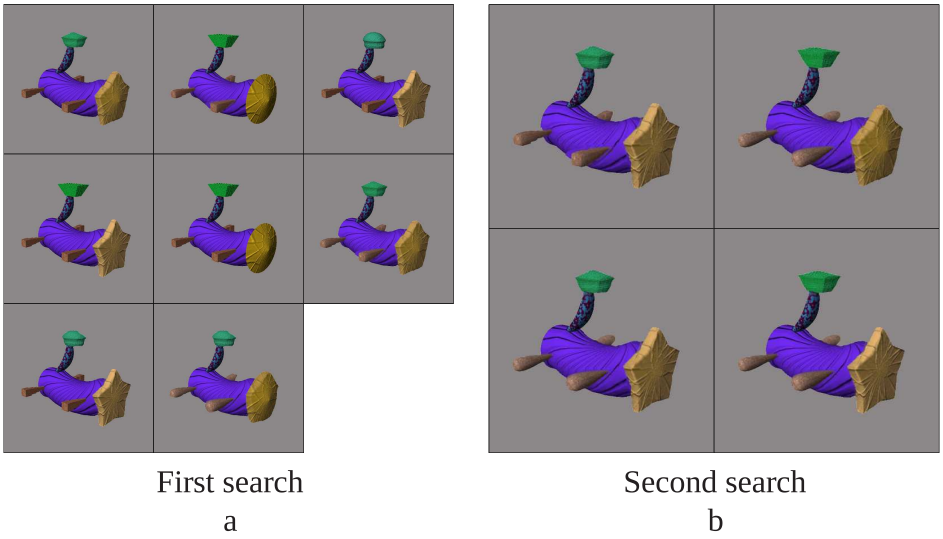


Figure 4.20: Stimuli visited three or more times in searches for S11, class 1. Images sorted in order of decreasing ROI response, averaged across all trials for each image.

cluster[4] (of red and blue diamonds) centered around $(0, -0.33, -0.66)$. The cluster contains three of the four stimuli visited three or more times in the second session — all but the stimulus evoking the second highest response from the ROI in Fig. 4.20b. These stimuli show mid-extremes green tail tip and mid-extremes tan head (see Fig. 3.6 for range of Fribble appearances); their legs generally are mostly round (morphed away from the rectangular shape at the other extreme). The outlying stimulus, while deviating in its more circular head and more flat-topped tail tip, retains the round leg shape. I observe Fribble ROIs often are most selective for the shape of a subset of the component appendages, although clustering appears to indicate the head and tail-tip shape remain important for S11's ROI as well, as does cross-session comparison of results below.

The class 1 search in the first session for S11 shows a quite weak convergence measure ($Z = -0.08$). Projecting the visited stimuli along the three Fribble-space dimensions (red and blue circles) shows the search spreading to all corners of the first two dimensions of the space, while focusing on $dim3 > 0$. In several locations, pairs of near-adjacent stimuli were visited, as in the lower left, upper right, and center of Fig. 4.19. In each location, the stimuli evoked opposite strength responses from the ROI — the second and seventh highest responses are coupled, as are the first and sixth, and the third and seventh. Sensitivity to slight changes in visual features had been observed previously for several brain regions of subjects viewing real-world objects.

The stimuli with positive values for the third dimension have more-rectangular legs, as seen by visual inspection of Fig. 4.20a, seemingly contradictory to the round leg selectivity posited for the second session. As part of my procedure, described in Sec. 3.5, the search in the second session is started at a position distant from the locations frequently visited in the first session to observe whether the search will return to the same location, showing consistency. For S11 class 1, the second search found and focused on one location, close to the frequently visited stimuli producing the first and sixth highest responses in the first search session, but shifted along the leg-shape dimension. This focus of one search session around a point of interest from the other

[4] For the interpretation of Fribble results, grouping was done by visual inspection of the three-dimensional scatter plots, e.g., Fig. 4.19.

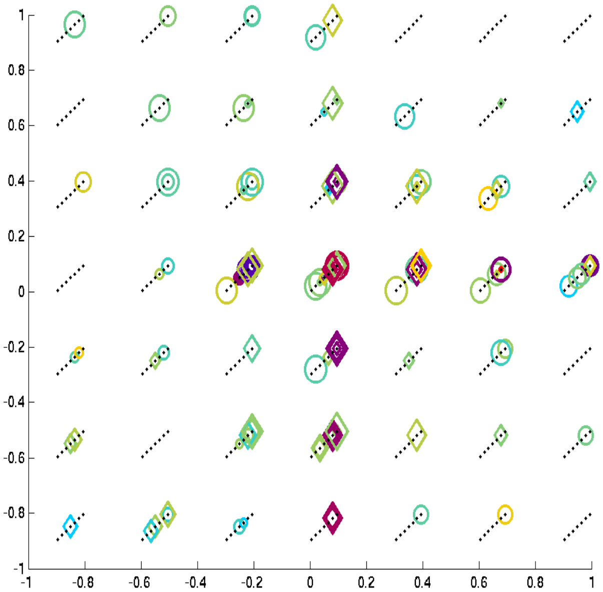
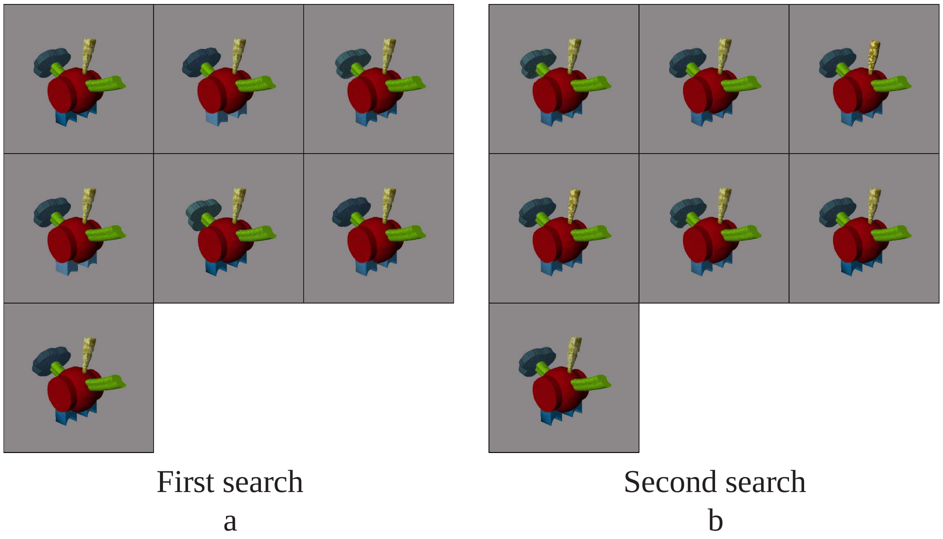


Figure 4.21: Search results for S17, class 3, shown in three-dimensional Fribble space. Colors and shapes used as in Fig. 4.19.

search session produces a consistency value, $Z = 2.10$. Comparing across searches, it appears all three attributes are important to producing high regional responses.

The **class 3/bipedal, metal-tipped tail object searches** for **S17** showed high cross-session consistency ($Z = 3.28$). Projecting the visited stimuli along the three Fribble-specific morph dimensions in Fig. 4.21, we see the first session focuses on the axis of dimension 1, the second session focuses on the axis of dimension 2, and both emphasize stimuli with $dim3 \approx 0.66$. As the convergence Z scores are low ($Z < 1.8$), the visited points for each session spread widely, albeit roughly confined to a single axis. The sessions' shared focus around $(0, 0, 0.66)$ results in the high consistency Z score. These points correspond to the stimuli evoking the first, sixth, and seventh highest responses for the first session and those evoking the second, fifth, sixth, and seventh highest responses for the second session, shown in Figs. 4.22a and b, respectively. Visually, these stimuli are grouped for their spiked feet ($dim3 = 0.66$), as well as for their tails appearing half-way between a circle and a cog shape (see Fig. 3.6) and their yellow "plumes" half-way between a round, patterned and angled, uniformly-shaded. The importance of spike-


First search
a

Second search
b

Figure 4.22: Stimuli visited three or more times in searches for S17, class 3. Images sorted in order of decreasing ROI response, averaged across all trials for each image.

shaped feet indicated in both searches, even beyond the $(0, 0, 0.66)$ cluster focus, may relate to prominance of edge detection in biological vision, expanding to the detection of sharp angles. As noted for other Fribble and real-world objects searches above, stimuli evoking the lowest and highest responses are notably clustered in the search space.

Visual comparison of searches and of regional responses for different subjects cannot be made across classes, as each Fribble space is defined by a different set of morph operations. Within class comparisons do not reveal strong consistent patterns across ROIs, as discussed below.

The **class 4/wheelbarrow object search** for **S19** showed the highest convergence measure ($Z = 4.20$) in session 1 across all subjects and object classes and a high convergence measure in session 2 ($Z = 2.01$). Furthermore, the two searches together showed the highest cross-session consistency ($Z = 3.80$) across all subjects and object classes. Projecting the visited stimuli along the three Fribble-specific morph dimensions in Fig. 4.23, we see clustering along $dim1 = 0$ and $dim3 = -0.33$ for the first session (red and blue circles); dimension 2 location of the stimuli is more broadly-distributed, but limited to $dim2 \leq 0$. The stimuli at the center of the first session cluster — those generating the first, fourth, fifth, sixth, seventh, and eighth highest responses as

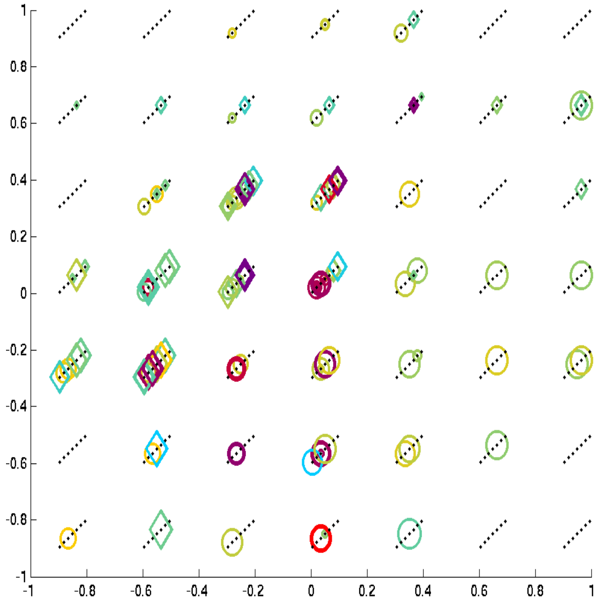


Figure 4.23: Search results for S19, class 4, shown in three-dimensional Fribble space. Colors and shapes used as in Fig. 4.19.

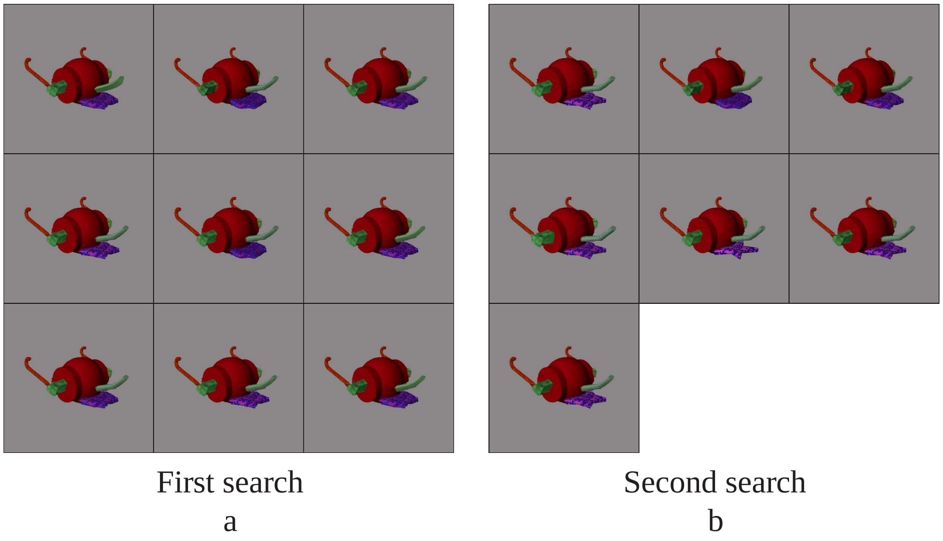


Figure 4.24: Stimuli visited three or more times in searches for S19, class 4. Images sorted in order of decreasing ROI response, averaged across all trials for each image.

shown in Fig. 4.24a — are linked by their purple tongue and green ear shapes, both intermediate compared with the extremes observable in Fig. 3.6. The ROI appears to be selective for the shape of a subset of component appendages, without regard for other elements of the object (i.e., the green nose). As observed throughout my search results, stimuli evoking high and low responses appear in the same cluster, sometimes adjacent to one another in space and appearing rather similar by visual inspection, indicating ROI sensitivity to slight changes in appearance.

Projecting the visited stimuli for the second session along the three Fribble dimensions (as red and blue diamonds) shows two clusters, one focused around $(0, 0.33, 0.66)$ and the other (consisting of two stimuli) focused around $(-0.66, -0.33, -0.66)$. The presence of multiple selectivity centers is consistent with observed ROI response properties for subjects viewing real-world objects, as well as Fribble subject S11 discussed above. The stimuli at the center of the larger second session cluster — those generating the first, fifth, and sixth highest responses in Fig. 4.24b — show a similar green ear and similar mid-extremes nose but a more star-shaped purple tongue. The two stimuli with the most-circular tongues, ordered second and third in ROI response, form the second cluster. This second cluster has the highest consistency with two of the cluster outliers from the first session, i.e., the second and third most active stimuli for the first session. The strong consistency among a small number of stimuli from the two sessions together produces the high Z score value for the consistency metric, $z = 3.80$. Stimuli evoking high and low responses appear in the same cluster, sometimes adjacent to one another in space and appearing rather similar by visual inspection.

Study of frequently visited stimuli in the remaining search sessions showing above-threshold convergence and consistency reveals a mix of results.

- The **class 1 search** in the second session for **S13** showed high convergence ($Z = 2.42$). The search frequently visited a cluster of stimuli evoking all but the lowest ROI response. The stimuli are focused closely around $(0.33, 0.66, 0.33)$, corresponding to a star-headed, square-legged object, indicating selectivity for the form of all three component appendages.

Stimuli close to one another in the visual space evoked evoked opposite high and low cortical responses.

- The class 1 search in the first session for **S15** showed high convergence ($z = 2.76$). The search frequently visited stimuli along the $dim1 = 0$ (head between star and circle shape) plane, indicating selectivity for the form of only one of the three component appendages.
- The class 1 search in the second session for **S16** showed high convergence ($Z = 2.42$). The search frequently visited a cluster of stimuli evoking the first, fourth, sixth, and seventh (out of seven) highest ROI responses. The cluster stimuli are focused around $(0.33, 0.33, 1)$, corresponding to an object with round legs ($dim3 = 1$), and mid-extreme tail tip and head. Stimuli close to one another in visual space evoked the highest and lowest cortical responses.
- The class 1 search in the second session for **S17** showed high convergence ($Z = 2.42$). The search frequently visited two clusters of stimuli evoking the third, fifth, and seventh (out of seven) highest ROI responses and the second and sixth highest responses, respectively. The presence of multiple selectivity centers is consistent with observed ROI response properties for several subjects viewing real-world objects and Fribble objects discussed above. The first cluster stimuli are focused around $(0.33, 0.33, 0.33)$, corresponding to an object with mid-extreme head, tail tip, and legs. The second cluster stimuli are focused around $(-0.66, -1, -0.66)$ corresponding to an object with a star head, muffin tail-tip, and rectangular legs. The class 1 searches performed across the two scan sessions for S17 showed high cross-session consistency ($Z = 2.28$). Frequently visited stimuli in the first session are limited to four points more-broadly distributed across the space, resulting in a low convergence value ($Z = -0.47$). The point evoking the highest ROI response in the first session is close to the $(0.33, 0.33, 0.33)$ stimulus cluster seen in the second search session, producing the high consistency measure.
- The class 1 search in the first session for **S19** showed high convergence ($Z = 2.00$). The

search frequently visited three clusters of stimuli evoking the second, third, and seventh (out of seven) highest ROI responses, the first and fifth highest responses, and the fourth and sixth highest, respectively. These observations again show the possibility of multiple selectivities in a selected ROI. The first cluster is focused around the origin, the second is focused around $(-0.33, -0.33, -0.33)$, and the third is focused around $(0, -1, 0)$.

- The **class 2/blue-bodied, yellow-winged object** (see Fig. 3.6) in the first session for **S16** showed high convergence ($Z = 2.10$). The search frequently visited only three stimuli, which cluster together around $(-0.33, 0, 0)$. The small number of frequently-visited stimuli limits broader conclusions about ROI responses across visual space and span of ROI activity.

- The class 2 search in the second session for **S17** showed high convergence ($Z = 2.97$). The search frequently visited a cluster of stimuli evoking all but the ninth highest ROI response. The stimuli are focused around $(-0.66, 0, 0.33)$, corresponding to a square-tipped eared object with mid-extreme wings. The class 2 searches performed across the two scan sessions for S17 showed high cross-session consistency ($Z = 3.14$). Frequently visited stimuli in the first session are more dispersed across space, resulting in a low convergence value ($Z = 0.89$). The stimuli evoking the first and third (of seven) highest responses in the first session are close to the cluster of stimuli frequently visited in the second search session, producing the high consistency measure.

- The **class 3 search** in the first session for **S16** showed high convergence ($Z = 1.80$). The search frequently visited two clusters of stimuli. The first cluster contains stimuli evoking the third, fourth, and fifth (out of seven) highest ROI responses; it focuses around $(-0.33, 0, -0.66)$, corresponding to a mostly flat-footed object with mid-extreme tail tip and mid-extreme yellow plume. The second cluster contains stimuli evoking the second, sixth, and seventh highest response; it focused around $(dim1, dim2) = (0, 0.66)$, corresponding to an object with a cog tail tip and a mid-extreme plume, while varying in dimension 3, affecting foot shape. The ROI selects for three appendage properties in one case and only two properties in another case. Stimuli close to one another in the visual space in the second cluster evoked opposite high and low cortical responses.

- The class 3 search in the first session for **S18** showed high convergence ($Z = 1.82$). The search frequently visited a cluster of stimuli evoking all but the first and sixth (of seven) highest ROI responses. The stimuli are focused around $(dim1, dim2) = (0, 0.66)$ while varying in dimension 3. Stimuli close to one another in space evoked high and low cortical response.
- The class 3 search in the first session for **S19** showed high convergence ($Z = 3.00$). The search frequently visited a cluster of stimuli evoking all but the highest ROI response. The stimuli are focused around $(0, 0.33, 0)$, corresponding to an object with mid-extreme tail tip, foot, and yellow plume.
- The class 3 search in the first session for **S20** showed high convergence ($Z = 2.86$). The search frequently visited a cluster of stimuli evoking all but the first three highest ROI responses. The stimuli are focused around $(-0.66, -0.66, 0)$, corresponding to an object with a narrow plume and almost-circular tail tip.
- The **class 4 search** in the first session for **S11** showed high convergenc ($Z = 3.90$). The search frequently visited a cluster of stimuli evoking all but the lowest ROI response. The stimuli are focused closely around $(0.33, 0, -0.33)$, corresponding to a fork-tongued, block-eared object. Stimuli close to one another in space evoked high and low cortical responses. The class 4 searches performed across the two scan sessions for S11 showed high cross-session consistency ($Z = 2.20$). Frequently visited stimuli in the second session are dispersed across space, resulting in a low convergence value ($Z = -0.38$). The stimuli evoking the third and fourth (of seven) highest responses in the second session are close to the cluster of stimuli frequently visited in the first search session, producing the high consistency measure. As usual, the measure reflects consistency of only a few stimuli

from one session with a high-activity stimulus or cluster of stimuli in the other session.

- The class 4 search in the second session for **S13** showed high convergence ($Z = 2.67$). The search frequently visited two clusters of stimuli. The first cluster contains stimul evoking the second, third, fifth, and seventh (out of seven) highest ROI responses; the second cluster contains stimuli evoking the first, sixth, and fourth highest responses. Both clusters are focused around $(dim1, dim2) = (0.33, 0.66)$. They differ by the extent to which the green ears are bright and curved. This ROI is another example of selectivity for a subset of Fribble component appendages, while requiring the "less-selected" appendage — the ear — to fall into one of two appearance categories.
- The class 4 search in the second session for **S18** showed high convergence ($Z = 2.30$). The search frequently visited a cluster of stimuli evoking all but the second (of six) highest ROI responses. The stimuli are focused around $(0.66, 0.66, 0.33)$, corresponding to an object with a curved ear, almost-circular tongue, and mid-extreme nose. Stimuli close to one another in visual space evoked high and low cortical responses.
- The class 4 search in the first session for **S20** showed high convergence ($Z = 2.86$). The search frequently visited a cluster of stimuli evoking the second, third, fourth, fifth, eighth, and ninth (of nine) highest ROI responses. The stimuli are focused around $(0.33, 0.33, -0.33)$, corresponding to an object with mid-extreme tongue, ear, and nose.

In sum, searches in most ROIs studied above cluster around a single location, indicating a single selectivity in visual space specific for all three component appendages in a given Fribble, though several for searches find multiple clusters and some results show Fribble location along certain dimensions does not affect ROI response. The invariance of ROIs to variation along a certain dimension, but selectivity along other dimensions is difficult to be detected when thresholding by convergence and consistency, which favors tight clustering along all dimensions. Locations of clusters, and of high ROI responses, are roughly equally likely to be in the middle of the space (morphing between clear end-point shapes) or close to the extreme ends (showing clear

$(\mathrm{Sn_m},\mathrm{Sp_q})$	corr	Tdist	$(\mathrm{Sn_m},\mathrm{Sp_q})$	corr	Tdist	$(\mathrm{Sn_m},\mathrm{Sp_q})$	corr	Tdist
$(\mathrm{S12_4},\mathrm{S15_4})$	0.53	0	$(\mathrm{S13_1},\mathrm{S15_1})$	0.70	1.3	$(\mathrm{S11_2},\mathrm{S16_2})$	-0.42	1.7
$(\mathrm{S12_2},\mathrm{S15_2})$	0.74	2.8	$(\mathrm{S15_1},\mathrm{S16_1})$	0.57	2.8	$(\mathrm{S13_1},\mathrm{S16_1})$	-0.25	2.8
$(\mathrm{S14_4},\mathrm{S17_4})$	-0.63	3.8	$(\mathrm{S19_2},\mathrm{S20_2})$	0.66	3.8	$(\mathrm{S11_1},\mathrm{S13_1})$	-0.05	3.8
$(\mathrm{S11_1},\mathrm{S15_1})$	**0.81**	3.8	$(\mathrm{S11_1},\mathrm{S16_1})$	0.68	3.8	$(\mathrm{S11_1},\mathrm{S14_1})$	-0.47	3.8
$(\mathrm{S13_1},\mathrm{S14_1})$	-0.67	3.8	$(\mathrm{S15_1},\mathrm{S14_1})$	-0.78	3.8	$(\mathrm{S16_1},\mathrm{S14_1})$	-0.67	3.8

Table 4.12: Correlation of activation profiles in Fribble space for anatomically proximal ROIs. ROIs selected to be the closest pairs within same subject in Fig. 4.3. The corr value for $(\mathrm{Sn_m}, \mathrm{Sp_q})$ and Tdist between the pair of ROIs are as defined in Table 4.11. corr≥ 0.8 in bold.

end-point shapes like star heads or sharp-toed feet). For several (but not all) ROIs, stimuli close to one another in visual space evoked high and low cortical responses — indicating sensitivity to slight changes in visual properties.

Comparison of neighboring regions

As for real-world objects searches, Fribble ROI response profile correlations are distributed roughly as a Gaussian distribution around 0, with a standard deviation of 0.36. Because the space for each Fribble class is defined by component-specific morphing operations, meaningful profile comparisons in the initial spaces only can be made across regions selected for the same class of stimuli. Anatomically proximal ROIs have widely varying correlations, as seen in Table 4.12. The positive values are above 0.5, but generally below $r = 0.8$, indicating similarities are weak but present. High negative correlations, as also seen in real-world objects results, may indicate a regions may suppress its response to stimuli excitatory to neighboring regions, constituting lateral inhibition on the level of ˜$10mm^3$ cortical regions. This potential property has parallels to the cortical inhibition for stimuli neighboring high-activity stimuli in Fribble-morph visual space.

Beyond metrics, cross-region consistency can be assessed visually, particularly in Fribble spaces in which the three dimensions of stimuli can easily be visualized in the two dimensional scatter plot. Looking across subjects at frequently visited stimuli (red and blue circles and diamonds) for class 2, searches for S3, S4 and S9 appear to pursue a focus around $(-0.5, 0, 0.33)$,

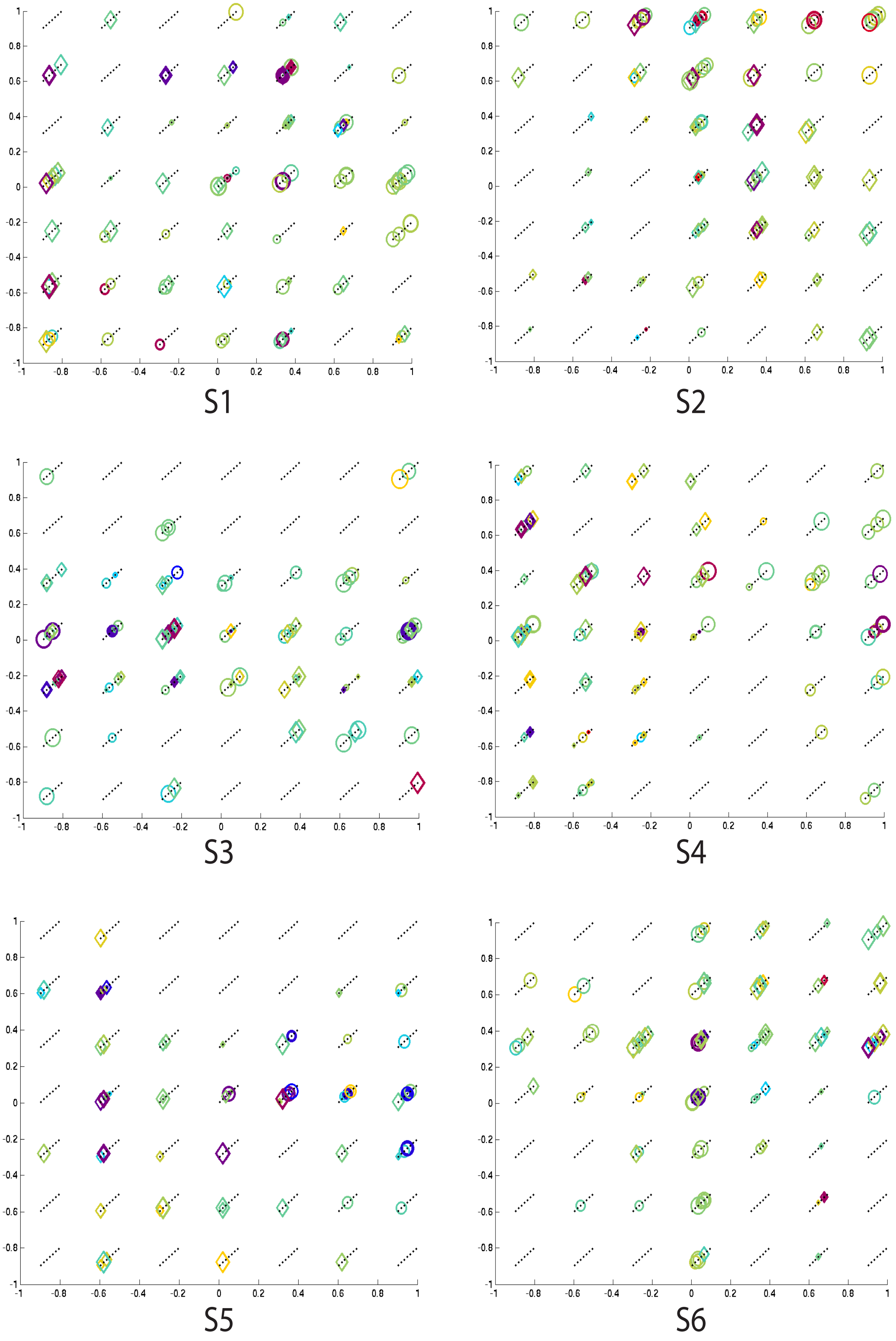


Figure 4.25: Comparison of search results and cortical responses across visual space for S11–S16, class 2. Colors and shapes used as in Fig. 4.19.

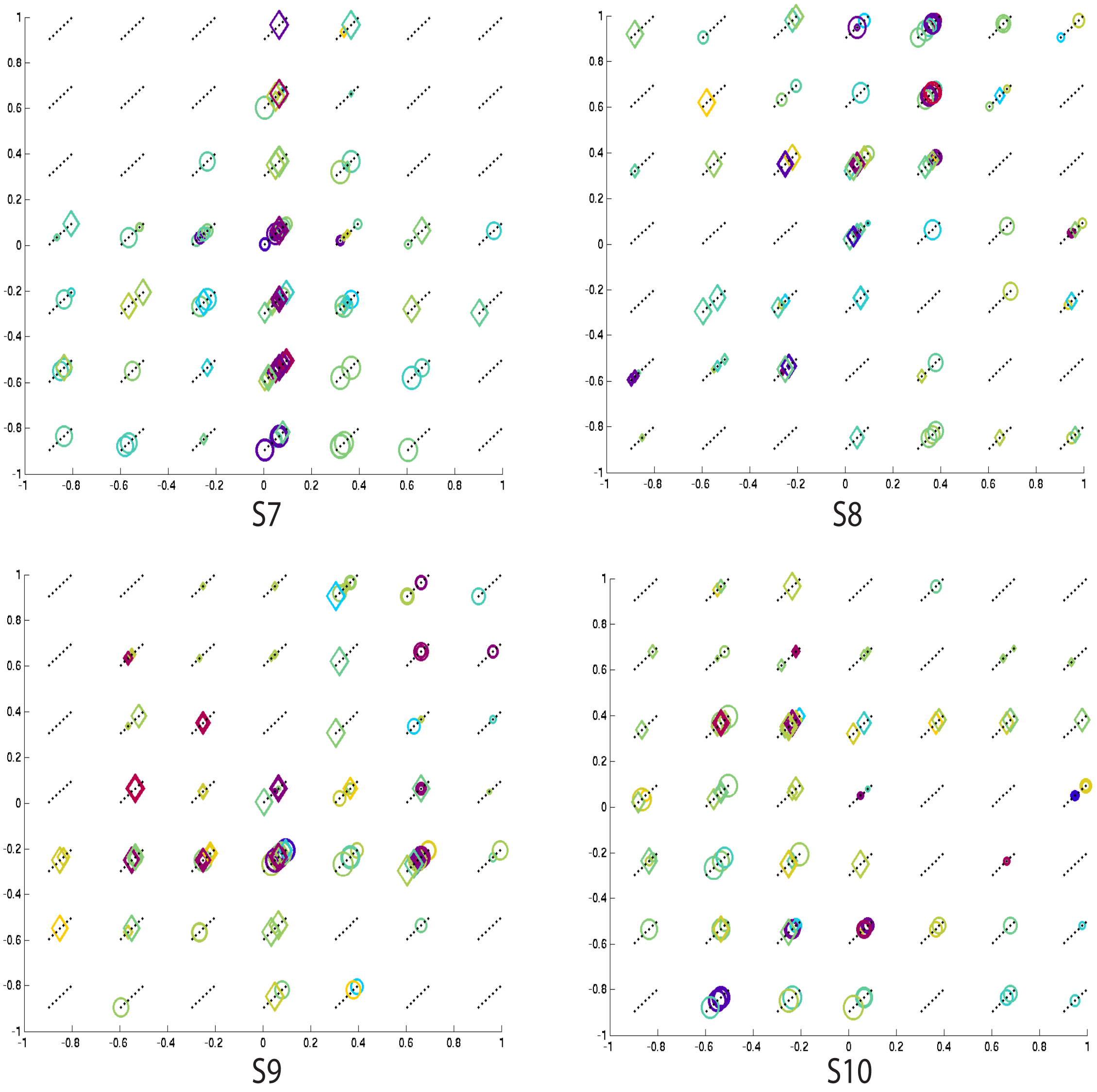


Figure 4.26: Comparison of search results and cortical responses across visual space for S17–S20, class 2. Colors and shapes used as in Fig. 4.19.

evoking strong, spatially adjacent positive and negative responses. Searches for S2, S8, and S9 appear to pursue a focus around $(0.5, 1, 0.33)$, also evoking strong, spatially adjacent positive and negative responses. Regardless of the specific visualization used to examine them, the visual feature spaces I have developed provide a powerful new tool for characterizing and understanding cortical responses to complex visual properties.

Chapter 5

Discussion

My goal in this study was to better elucidate the complex visual properties used by the brain for visual object perception. In contrast to our understanding of early visual processing (e.g., edge detection in primary visual cortex) and the high-level organization of visual cortex (e.g., broad category identification in LOC, FFA, and PPA), intermediate representation along the ventral pathway is poorly understood. In my recent work (Chap. 2), I identified computer vision methods that successfully modelled object representation at different stages of the ventral pathway — methods predicted what object stimulus pairs would produce similar or distinct cortical responses from selected brain regions. My present work procedes to use computational models of perception to establish low-dimensional visual feature spaces as a context in which to characterize brain region activity across the world of visual objects. Where Hubel and Wiesel explored varying orientations and locations of edges to excite neurons in V1 [22], I define spaces of complex object-related visual features to explore which properties will excite a ˜$10mm^3$ brain region at higher levels of the ventral pathway. I develop and employ novel techniques for realtime fMRI study to quickly identify brain region selectivities — those visual properties evoking maximal cortical response from a pre-select region — given limited scanning time.

My work begins with the scientific question "What are complex visual properties used in cortical object perception?" and the related technical question "How can we select stimuli to best identify cortically-preferred visual properties in limited fMRI scanning time?” Both questions are addressed through the development and application of a set of programs that dynamically choose visual object stimuli to show to a subject in the scanner based on the subject’s cortical responses to previously-shown stimuli. This “realtime” stimulus selection is performed in the context of a search of visual feature space, using the simplex simulated annealing method [7], to quickly find the stimuli (corresponding to feature space locations) producing highest cortical responses from a pre-selected brain region. Many modeling and technical choices underlie the operation of my realtime search methods, pertaining to: definition of visual feature space, selection of brain regions to study (fixed prior to the search of visual space), rapid and accurate computation of regional responses, and effective communication among programs running in parallel to perform all elements of the search. In the definition of feature spaces, I test two approaches — first organizing real-world object stimuli based on their similarities as measured by the SIFT computer vision method [36] and then organizing synthetic Fribble objects [76] based on morph operations to component parts required to transition from one object appearance to another. I observe most aspects of each search generally behaved as expected, supporting my choices. However, there remains much room for further development. Most searches fail to converge on a clear location in visual feature space as the regional selectivity, indicating shortcomings in the feature spaces — particularly in the space defined based on SIFT — and in assumptions about the nature of cortical responses across each space — the simplex search method expects a unique maximum, which is not seen in the presently-gathered data.

Those realtime searches that successfully converge, however, provide new insights into the complex visual properties utilized by mid- and high-level brain regions in the ventral pathway. Observing cortical activities over the defined visual feature spaces, I find multiple brain regions producing high responses for several sets of visual properties, i.e., for two or three locations in space. I also find many regions suppress their responses for stimuli adjacent in space — and slightly varied in visual appearance — from those stimuli evoking markedly high cortical activity, indicating a high-level “surround suppression” computation as discussed below. Visual inspection of stimuli corresponding to the spatial selectivity centers provides visual intuition about high-level visual properties of interest, such as holistic object shape, shapes of component parts, and surface textures.

5.1 Regional selectivities

My novel methods in realtime fMRI search seek to identify visual object stimuli producing maximal activity for a given brain region — revealing the complex visual property selectivities of the brain region. I study the cortical responses recorded in the course of realtime searches in the context of visual feature spaces to understand brain region activity across a world of visual objects, finding flaws in the initial search assumptions that there is a unique maximum to a region's cortical activity across visual space. I seek intuition about visual properties of interest to brain regions in the ventral stream through inspection of stimuli evoking extreme cortical activity, finding holistic object shape, shapes of component parts, and surface textures are included among cortical selectivities.

5.1.1 Selectivities in feature space

Examination of the distribution of cortical responses across the defined visual feature spaces indicates repeating patterns across subjects and ROIs not anticipated in my search design. The simplex method assumes the activity of a given brain region reaches a maximum for a stimulus corresponding to one location in feature space and activity monotonically decreases for stimuli corresponding to points increasingly distant from the maximum. In contrast, in both SIFT and Fribble spaces, several searches show extreme high and low response stimuli cluster together, with mid-level response stimuli spread further from the cluster center. This pattern of slightly differing stimuli causing extremely different neural responses is familiar from visual coding properties in earlier stages of the ventral pathway. In both SIFT and Fribble spaces, several searches also show cortical response maxima distributed broadly across space, rather than concentrated in one location as a unique regional selectivity.

The proximity of stimuli evoking ROI responses of opposite extremes can be seen through scatter plots (e.g., Figs. 4.19 and 4.23[1]), sorted stimulus figures (e.g., the red figures in Figs. 4.8b and the hinged-ajar boxes in 4.12a[2]), and activation profile cross-sections (e.g., Figs. 4.17a and 4.18a), as well as through discussion of further high-convergence search examples in Chap. 4. These findings are consistent with the principle of surround suppression observed at early stages of the visual system. Hubel and Wiesel observed spatially adjacent "on" and "off" edge regions in visual stimuli exciting or inhibiting, respectively, the spiking of neurons in V1 [22]. In modern hierarchical models of the human visual system, the first stage reflects these early findings by using a series of Gabor filters [27, 57]. Prior to cortical coding, retinal ganglia cells similarly are known to have receptive fields characterized by concentric "on" and "off" rings in the image plane of any given stimulus [49]. Wang et al. found evidence for surround suppression, again based on location in the image plane, for perception of the second order texture statistics of noise [73]. These multiple stages of alternating patterns of excitation and suppression are consistent with principles of successful neural coding models, in which lateral inhibition of representational units "located" adjacent to or nearby one another are found to be advantageous to computational perception tasks [25, 50]. While past perceptual studies have focused on suppression of percepts neighboring one another in the plane of the image falling on the retina, local competition in alternative feature spaces are conceptually plausible from neural coding models. My work indicates the use of surround suppression in more complex representational spaces employed at more advanced stages of cortical visual object perception. From a methodological perspective, my results illustrate both the descriptive power of the feature spaces I have defined — based on SIFT and Fribble-morph representations — and the ability of the realtime search

[1] Scatter plot examples only are given in Fribble spaces as they are more easily evaluated visually in one two-dimensional plot.

[2] Stimulus examples only are given for SIFT searches as similarities of the real-world object stimulus set are easier to see than they are for Fribbles that all look predominantly similar within a given class to the uninitiated reader.

method to capture meaningful properties of cortical behavior.

The presence of multiple local ROI response maxima in each feature space also can be seen through scatter plots (e.g., Figs. 4.1 and 4.19) and activation profile cross-sections (e.g. Fig. 4.18b), as well as through discussion of further results. Indeed, the high frequency of low convergence scores in Tables 4.5 and 4.8 may correspond to large numbers of local maxima in each search session, which may explain the results for subject S11, class 1 in Fig. 4.19. Overall, these findings suggest the one cubic centimeter 125-voxel cubes studied may contain multiple cortical sub-regions selective for distinct visual properties. Given the millions of neurons present in each region and the specificity of properties explored, functional variability is a potential risk of the analysis design. Regions purposely were limited in expanse to lessen the likelihood of variability, but multiple voxels were retained to build upon the visual representation-searchlight analysis findings of Chap. 2. A weighted average of voxel responses was used to compute a single number for regional response for each trial, as discussed in Sec. 3.3.6. The weighting was intended to further suppress less-prominent multi-voxel activity patterns, though the method for selection of these weights may be modified to focus on voxels with a single selectivity, as discussed further in Chap. 6.

5.1.2 Selectivity visual intuitions

Analysis of cortical activities over visual space provides valuable understanding of the presence of one or several selectivities for a brain region and the presence of surround suppression within the defined visual space. However, intuition about the nature of preferred stimuli, and their underlying visual properties, is better obtained through visual inspection of those stimuli frequently visited by each search and evoking extreme cortical responses. For many real-world objects searches, it was not possible to identify unifying visual patterns of preferred stimuli. However, for a few searches I did observe potential selected shape and surface properties. For Fribble object searches, executed in carefully-constrained visual spaces, unifying visual patterns for stimuli producing high cortical activity largely were holistic Fribble shapes. There were no clear patterns across subjects regarding the preferred types of holistic shapes, dependent upon the shapes of the three components of each Fribble class.

Frequently visited stimuli clustered together in SIFT space — evoking both extreme high and low responses, consistent with the observations in Sec. 5.1.1 — can be united by broad shape (e.g., width in Figs. 4.8b and 4.10a or relative three dimensional proportions in Fig. 4.12a), surface properties (e.g., brightness in Fig. 4.8b or texture in Fig. 4.14), and fine internal contours (e.g., sharp-edged holes in Fig. 4.16). Observed selectivity for shapes is consistent with the findings of Yamane et al. and Hung et al., who successfully identified two- and three-dimensional contour preferences for neurons in V4 and IT using uniform-gray blob stimuli [24, 78]. Unlike these prior studies, my work employs real-world stimuli and thus identifies classes of preferred shapes likely to be encountered in normal life experience. Observed selectivity for surface properties is a more novel finding, though Tanaka observed such selectivities in primate IT neurons in the context of perception of object drawings [63]. Many searches performed for real-world objects revealed no clear patterns among stimuli evoking extreme cortical region responses, clustered together in SIFT-based space. This lack of clear patterns likely reflects the difficulty of capturing the diversity of real-world visual properties in a four dimensional space, as discussed in Sec. 5.2.3.

Fribble objects, and corresponding “Fribble spaces,” were used to study ten subjects with stimuli more controlled in their span of visual properties. Frequently visited stimuli in each Fribble space can cluster around a three-dimensional coordinate. Each dimension corresponds to variations of a component shape morphed between two options, such as a star/circle head or flat/curved feet, as in Fig. 3.6. Thus, clustering around a point indicates slight variations on three component shapes, with focus around a fixed holistic shape. Across subjects, there is no clear pattern of the nature of favored holistic Fribble shapes, nor of favored shapes for the three varying component “appendages.” For some searches, frequently visited Fribble stimuli evoking strong cortical responses can vary along one axis or two axes while staying contant on the remaining one(s). Depending on the brain region, one to three component shapes can account for selectivities. Regional selectivity for parts of an object, rather than the whole, may be associated with cortical areas particularly early in the ventral pathway; this finding would be consistent with the focus of early and intermediate stages of vision on spatially-distinct parts of a viewed image, pooled together over increasingly broad parts of the image at higher stages of vision [48].

For both real-world and Fribble objects searches, visual inspection of the ordering of stimuli by ROI response, e.g., Fig. 4.8 and 4.20, fails to yield any further insights. A priori, we would expect shape properties to smoothly transition as measured responses decreases. The lack of this transition may stem from the mix of multiple coding units, noise in fMRI data (despite averaging), and particularly from the presence of surround suppression, placing similar-looking stimuli at opposite ends of the line of sorted stimuli.

5.1.3 Comparison of neighboring regions

Similar to the notion of retinotopy [72], in which neighboring brain regions encode neighboring parts of visual space, we expected selectivities in nearby brain regions to exhibit selectivities for stimuli drawn from nearby parts of the more-complex visual feature spaces. Comparison of regional selectivities was performed by smoothing the responses for the scattered visited stimulus points over space to form “activation profiles” as in Fig. 4.17. As discussed in Sec. 4.4, selectivities for nearby ROIs rarely showed strong similarity, both within subject and across subjects[3]. There are several moderate positive and negative matches between nearby ROIs in Fribble spaces, but the results are not as strong as desired ($0.5 \leq |r| \leq 0.8$). While these weak results may indicate a lack of continuous transitions in selectivities across the brain, they also may reflect limitations in the method of comparison. For example, an activation profile may contain multiple maxima reflecting behavior of neurons in different locations across a region. A region neighbor-

[3] Cross-subject ROI distances were computed based on projection to a Talairach brain, as discussed in Sec. 4.2.1.

ing from the left may only contain maximum responses for stimuli activating neurons in the left portion of the initial region, and additional maxima for neurons towards the right end of the new region. This partial overlap may not be sufficiently reflected in the correlation metric describe in Sec. 4.4.1. Similarly, a neighboring ROI slightly shifting in feature space a pattern of strong negative response surrounding strong positive response, corresponding to surround suppression discussed above, may cause the two respective responses to have strong negative correlations rather than strong positive correlations, potentially explaining moderately negative correlation Fribble results in Table 4.12.

5.2 Influences on search behavior

In my work, the study of complex visual feature selectivities of regions in the ventral pathway was driven by a set of programs that dynamically selected stimuli to display during each scan session based on cortical responses to stimuli displayed seconds earlier in the session. Dynamic, or "realtime," stimulus selection was pursued to most effectively search the defined space of visual properties in limited scan time and to most quickly identify objects that produce the highest responses from each brain region under study. The performance of realtime searches for maximally preferred stimuli in areas of the ventral pathway has not been pursued previously in neuroimaging to my knowledge, and is quite new to neuroscience studies of vision in general [24, 78]. I implemented and applied a set of programs for this study, and assesses the performance of these programs. The three programs — responsible for cortical response measurement (called "preprocessing"), stimulus selection ("search"), and stimulus display ("display"), respectively — generally acted as expected and successfully worked together. However, the selection of stimuli by the search program frequently failed to converge on a visual selectivity for a given brain region. A variety of inaccurate visual selectivity and technical assumptions underlying my methods likely challenged the effectiveness of these searches.

For each subject and brain region, the realtime search method I implement — simplex simulated annealing [7] — is expected initially to probe the brain with stimuli broadly distributed in the defined feature space, but quickly to narrow its focus to stimuli drawn around the area in feature space evoking the highest cortical responses. Furthermore, searches are expected to produce the same results for a fixed ROI regardless of the starting point in feature space. These expectations frequently were not met, as shown in Tables 4.5, 4.6, 4.7, 4.8, 4.9, and 4.10. The simplex method frequently revisited points spread across SIFT and Fribble spaces, with little clear change in focus over time. Consistency of search results for the same brain region across scan sessions was similarly poor. Nonetheless, the successful convergence and consistency of searches for several ROIs, and the insights resulting from these searches, indicate great robustness and ongoing promise for my methods.

5.2.1 Simplified search assumptions

As discussed above, the simplex method expects a given ROI's stimulus response function to have a unique maximum in feature space. In contrast, my data often show multiple local maxima. If there are three or more maxima in a region — particularly if the number of maxima is larger — it is unlikely the search will repeatedly probe a sufficient number of stimuli to associate each maximum location with a large enough cluster of stimuli to produce a high convergence value, defined in Eqn. 3.6. Similarly, the presence of a large number of maxima increases the likelihood that starting searches from different points in feature space will produce different sets of results, each focusing on points closest to their respective starting location, producing poor consistency measures as defined in Eqn. 3.7.

5.2.2 Technical challenges in realtime programs

Myriad technical choices were required in the implementation of the three programs executing the realtime search for regional selectivities. While the majority of these choices enabled the smooth operation of my experiments, some challenges did arise for each program.

The persistent variability of visited stimulus locations across each search session, indicated by low $|\Delta\text{var}|$ values in Tables 4.7 and 4.10, reflects the simplex re-initialization strategy described in Sec. 3.1.5. Each search session was divided into multiple runs. At the beginning of each run, a new simplex was defined centered around the location eliciting the highest ROI response in the previous run — except in the first run, in which the starting point was selected as discussed in Sec. 3.1.5. The four initial new simplex points in the four-dimensional space were selected by taking random-sized steps away from the initial point along each dimension. The random distribution used to generate these steps remained the same for each run, causing an equal spread of points to investigate at the first run of the session as the last run of the session. Further developments to the search program allowed for partial continuation of searches across runs, rather than requiring new simplex initializations — however, these improvements have not yet been thoroughly tested to confirm proper execution and, therefore, were not used in the present work.

Further technical choices in the realtime computational system posed additional challenges to effective search. Insufficiently fast computation and network-communication times prevented the display program from showing subjects the correct stimuli at the proper time on as many as 40% of trials, as seen in Tables 4.1 and 4.2. While the frequency of display errors was significantly reduced by switching methods of inter-program communication — from sharing a file over a mounted drive to passing information directly through a socket — errors still occured, sometimes on as many as 10% of trials. The preprocessing and search programs, described in Sec. 3.1, assume the correct stimulus is shown for each trial and select new stimuli to show based on computed ROI response regardless of the validity of the visual stimulus actually reaching the subject. Incorrect displays misinform the simplex search about stimulus responses and can lead to sub-optimal exploration and acceptance of future simplex points. However, the search still often recovers sufficiently to identify ROI selectivities — one of the nine examples of significant real-world objects search convergence, $S8_1$, comes from a session in which over 25% of stimulus displays were incorrect, and most convergent searches include a more limited number of erroneous stimulus displays. The assumption of noisy stimulus response measurements embeded in the simplex simulated annealing approach may contribute to the robustness of realtime search to display errors.

Shortcomings in motion correction during preprocessing also may mislead realtime stimulus selection. As discussed in Sec. 4.3.3, the location of ROIs for each search is determined at the beginning of each scan session. Optimally, all functional volumes collected through a session should be realigned to the brain position at the beginning of the session, to ensure the proper voxels are used to compute stimulus responses. Instead, in my study, volumes were aligned to the brain position at the start of their respective runs. These positions potentially could be shifted from the beginning of the session. Comparing offline computation of ROI responses based on start-of-session alignment with responses computed using realtime start-of-run alignment, as reported in Tables 4.3 and 4.4, regional activity estimates could differ significantly. Similar to the risks of undetected display errors, incorrect response calculations could lead to sub-optimal exploration and acceptance of future simplex points. However, counter to this theoretical concern, we can observe that 66% of significant convergence results for real-world objects searches and 50% of significant convergence results for Fribble searches correspond to sessions whose realtime–offline calculations have correlations $r < 0.3$. Limiting convergence and consistency measures to stimuli visited three or more times may permit averaging activity over multiple trials to overcome errors in individual measurements. Alternatively, for sessions with highly negative correlations, particularly noticable in Fribble spaces, searches may effectively be searching for stimuli evoking particularly low responses; this strategy may successfully identify maxima in stimulus space as well because of the observed phenomenon of surround suppression.

5.2.3 Limitations of SIFT multi-dimensional scaling space

The use of a SIFT-based Euclidean space yielded particularly poor search performance across subjects and ROIs, despite the abilities of SIFT to capture representations of groups of visual objects in cortical regions associated with “intermediate-level” visual processing, discussed in Chap. 2. Significant convergence and consistency statistics were observed more rarely than expected — certainly compared to those statistics in Fribble spaces — and visual inspection of frequently-visited stimuli frequently failed to provide intuition about visual properties of importance to the brain region under study.

Confining the SIFT representation to four dimensions, found through multi-dimensional scaling as discussed in Sec. 3.3.2, limited SIFT space’s descriptive power over the broad span of visual properties encompassed by real-world objects. Use of a small number of dimensions was required to enable effective search over a limited number of scan trials. However, Fig. 3.3 shows that at least 50 dimensions would be required to explain 50% of the variance in a SIFT-based pairwise distance matrix for 1000 images. Even among the $\tilde{}100$ stimuli employed for each object class, the four dimensions used account for less than 50% of variance. The missing dimensions acount for grouping pairwise distance patterns across large sets of images — therefore, more-careful selection of stimuli included in a given object class still renders four-dimensional SIFT space insufficiently-descriptive.

Intuitively, it is not surprising that there are more than four axes required to describe the visual world, even in the non-linear pooling space of SIFT. Indeed, the method used successfully in Chap. 2, and repeated for the realtime study, employs 128 descriptors and 128 visual words [35]. Further study shows that tailoring SIFT space for each of the four object classes used in my sessions still requires over 10 dimensions each to account for 50% of variance. The exploration of selectivities for real-world objects using Euclidean space may well require more dimensions, and thus more trials or a more efficient realtime analysis approach. The number of dimensions may be kept small by identification of a superior feature set, or by limitations on the stimuli. I pursue the latter through Fribble spaces, with notable improvement.

My definition of SIFT space also may obscure visual intuitions for properties unifying stimuli producing high cortical activations. Multi-dimensional scaling identifies dimensions maximizing the preservation of pairwise distances between images. This method allows groups of objects deemed similar by SIFT to be clustered together, but may not capture patterns of visual variability smoothly transitioning between two extremes across a diverse set of objects. Fribble space, again, is defined to capture such variability, and reveals ROIs invariant to changes in some dimensions but selective to changes in others.

5.2.4 ROI selection

Weak matches in response properties of nearby ROIs in part may be attributed to the lack of sufficiently close pairs of regions under study. My study explores a new method — realtime fMRI search of a complex visual space — to gain insight into visual object encoding. Given the multiple sources of uncertainty, from recording and physiological noise to questions of optimal realtime analysis techniques, brain regions are selected in each subject to maximize search performance rather than to maximize opportunities for cross-region comparisons. ROI selection was performed based on the strength of my models to explain activity during reference scans, as discussed in Secs. 3.3.6 and 3.4.5. Furthermore, a diversity of anatomical locations were selected in each subject, as reference scan data allowed, to gain perspective on cortical selectivities in a breadth of regions across the ventral stream — a goal opposing study of transitions across neighboring cortical areas.

Regions were selected manually by assessing localizer results, anatomically restricted to the ventral stream. Often there were multiple candidate areas that could have been selected. Regions centered in a larger group of voxels with high matches to SIFT and class-specific activitations, or to Fribble-class specific encodings, were favored. However, I decided the balance of SIFT and class-specific matchings along with proximity to anatomical locations of interest in the ventral stream, e.g. lateral occipital, fusiform, or anterior temporal, on an ad hoc basis. Further analyses are needed to determine the optimal balance. In Chap. 6, I propose a potential fixed method for future ROI selection.

5.3 Promise of realtime stimulus selection

My work employs a collection of novel methods in realtime analysis of cortical data to explore complex visual properties used in perception. This exploration faces myriad technical and biological challenges — from scanner and physiological noise in fMRI recordings to uncertainty about the nature of higher level visual representations — compounded by the small number of stimuli able to be shown in the limited scanning time. Realtime selection of stimuli based on cortical responses to recently displayed visual objects optimizes the use of this limited scanning time, building on similar approaches in primate neurophysiology [24, 63, 78]. My present application of simplex simulated annealing [7] for stimulus selection faces considerable additional challenges — from occasional faults in stimulus display to frequent simplex resets — resulting in lack of convergence and consistency for searches across a number of brain regions across subjects. However, numerous brain regions studied, particularly using Fribble stimuli, produced successful search performance revealing novel insights into visual object perception. In the novel search spaces I defined, I observe surround suppression, likely reflective of local competition between neural units, and multiple sets of featural selectivies, likely reflecting the large size of the studied brain regions. I identify local and global shapes and surface properties such as texture and brightness as biologically relevant complex features. I observe similarities in activation patterns across visual feature space across subjects, confirming complex selectivities are shared across subjects. These results, found for the convergent and consistent searches in my study, point the way to future models of higher level vision and more refined methods for realtime stimulus selection. These improvements can further overcome the many challenges facing my present work. Indeed, the success of my initial approaches in searching for stimuli for a number of brain regions provides encouragement for the potential of future work in realtime fMRI. This encouragement builds on the promise of realtime stimulus selection already seen theoretically as an efficient use of limited scanning time and empirically from past successes in neurophysiology.

Chapter 6

Future work and conclusions

I have developed a realtime search method to determine the selectivity of ventral stream regions for complex visual properties using fMRI. My work has provided new understandings about visual representation of objects in the brain. I identified brain regions selective for holistic and component object shapes and for varying surface properties. The visual feature spaces I defined in my work provide a powerful new tool for characterizing cortical responses to complex visual properties. My findings also serve as a compass to further development of realtime fMRI methodology to study the visual system more effectively. It informs important choices in processing of signals from across the brain and from within selected voxel regions, modification of the simplex search method, and assessing the evolution of selectivities between regions.

6.1 Voxel selection

Selection of “optimal” regions of interest, based on prior reference scan data, is at the center of realtime analysis. These regions specify the voxels that will control the choice of stimuli displayed, in turn revealing regional selectivity. The ROIs used in my study often showed multiple selectivities across a given feature space, potentially reflecting the presence of multiple neural groups effectively competing for control over the search. To focus on single group selectivities, it may be advisable to decrease the number of voxels within a region — potential to $3^3 = 27$. Alternatively, the method for consolidation of voxel responses to a single region-wide number may be revised to emphasize activities from only one neural group. The method used in the present study performs a weighted sum based on the first component learned from principal component analysis on voxel responses in the reference scan, as described in Sec. 3.3.6. This method identifies and emphasizes the most common activation pattern across the region, which may be a mix of multi-voxel patterns resulting from multiple commonly-activated neural groups. In contrast, use of independent component analysis, or similar sparse methods, potentially including spatial constraints, is more likely to define separate components for voxel patterns corresponding to each neural group. Summing with the weights of an independent component may better emphasize a single selectivity. Analyses on already collected data can indicate the effects of these approaches for future experiments.

Beyond size and weighting, the optimal locations on which to center voxel regions merits further exploration. As discussed in Sec. 5.2.4, the manual selection of ROIs in my work balanced desires for broad coverage of the ventral stream with maximizing the chances for strong signal and for strong search performance. These balances may be set through a deterministic algorithm. Furthermore, selection in future studies may confine ROIs to a more restricted functional or anatomical area, potentially favoring use of neighboring ROIs for same-session searches to compare selectivity transitions across a cortical area. Algorithmic selection can assign each voxel values drawing on:

- Z scores from class and representation-space localizers, such as those discussed in Sec. 3.3.4
- Z scores of nearby voxels and voxel-searchlights
- distances from desired brain regions, defined through functional localizers, anatomical landmarks, and location of other selected ROIs for realtime analysis

These values can be weighted and summed to produce a score for each voxel that represents its desirability as the center for a region. Weights can be determined based on analyses of already collected data.

6.2 Search method evolution

The simulated annealing simplex search, spread across multiple runs for each realtime scan session, incorporates a variety of algorithmic choices intended to most effectively identify selectivity of a region under investigation. Results from my study suggest directions for improved search performance. Convergence over scan time — rather than the clustering measured by my "convergence" measure employed in this work — is limited by the re-initialization of the search simplex at the start of each run with widely spread simplex points around a carefully chosen center. Two methods may be used to approach more desirable search behavior.

- At the start of each new run, the re-initialized simplex points can be defined to be offset from the simplex center with a uniform random distribution, as discussed in Sec. 3.1.5, scaled with a decreasing maximum with increasingly later runs in the session. The scaling can decrease using the temperature reduction equation defined in Eqn. 3.4. Furthermore, the simulated annealing temperature for each search — affecting acceptance/rejection of candidate points into the simplex — can be updated by the same equation for each new run. This approach is relatively easy to introduce into the current software, but artificially speeds search convergence.

- The simplex state at the end of each run can be provided as the starting condition to the search program at the beginning of the next run. Using this method, searches may run a sufficient number of steps across runs to meet convergence criteria given in Cardoso et al. for termination and for interim temperature updates, truly allowing the search to incorporate simulated annealing [7]. Transfer of full search information across runs and simultaneously management of searches at different stages of temperature update and termination presents further implementation challenges, which are unfortunate but merit the effort to address.

Analyses on already collected data may be used to indicate the likely impact of these modifications and to fine-tune optimal parameters for simplex and temperature updates.

The simplex method is intended to find a unique maximum in feature space, assuming response to a given point decreases with distance from the maximum location. The presence of multiple maxima with surrounding regions of suppression in feature space, violates this assumption. From my data, it appears the simplex can successfully identify a small number of local selectivities. The development of realtime fMRI search would benefit from simulation studies of my method's behavior incorporating the reality of multiple maxima, and exploring the utility of modifications to the method and of entirely different methods for probing the feature space.

6.3 Techniques in preprocessing

Realtime preprocessing of fMRI signal is needed for meaningful analyses. However, the constellation of registration, detrending, and normalization steps employed require computation time that can lead to delays in selecting new stimuli, in turn leading to incorrect stimulus displays that misinform the ongoing searches. At the same time, I observe that post hoc performance of a more complete, but slower, set of preprocessing steps in Tables 4.3 and 4.4 can result in different estimates for cortical region responses, indicating the steps currently used in realtime may not always be sufficient for proper stimulus response computations. It would be valuable to perform further post hoc analyses on the data already collected to identify processing the steps best added, changed, and removed to strengthen the reliability of regional response computations while decreasing the risk of delays associated with these computations. Incorporating the findings from Sec. 4.3.2, future studies likely should perform volume registration with the brain position at the start of each session, rather than with the position at the start of each run.

6.4 Lessons learned

My work uses realtime selection of stimuli in conjunction with fMRI to explore visual properties used by pre-selected voxel regions in the visual object perception pathway. These methods have broad potential applications to study, for example, diverse levels of vision, alternative senses such as hearing, and more abstract semantic representations in the brain.

For continued pursuit of realtime fMRI exploration of complex perceptual spaces, there are six central design factors to consider:

- Selection of cortical region(s)/voxel(s) for study
- Selection of pool of potential stimuli (e.g., images, words, or sounds) from which to draw during the experiment
- Organization of potential stimuli for effective realtime selection
- Stimulus selection method
- Stimulus presentation design — particularly, subject task and stimulus onset asynchrony (i.e., the time between onsets of stimuli)
- Realtime fMRI signal processing

Selection of voxels for study As realtime stimulus selection still is a young field, it is advisable to continue along the relatively simple path already pursued with some success in primate neurophysiology, identifying a neural "unit" likely selective for only one set of properties and determining those properties for which it is selective [63]. In fMRI, minimization of the neural units studied may be achieved by analysis of activity from a small voxel region, e.g., containing $2 \times 2 \times 2$ voxels rather than $5 \times 5 \times 5$ voxels, if not fewer. Alternatively, independent component analysis can be used to identify a set of voxels that tend to reliably vary together across stimulus presentations — either across a $5 \times 5 \times 5$ voxel region of interest or across the ventral temporal cortex in general. These covarying voxels are likely responding to similar visual properties, permitting study of a single set of visual properties relevant to cortical perception. The initial selection of voxels for analysis should utilize class and representation-space localizers as discussed in Secs. 3.3.6 and 3.4.5, to increase measured cortical responses over noise and to ensure any visual representations in the stimulus organization and selection methods are reasonably accurate for the voxels under study.

Selection of potential stimulus pool For early realtime explorations of complex perceptual properties, the pool of potential stimuli should have limited “dimensions” of variation — though these dimensions may not need to be explicitly defined a priori. The highest realtime search performances in my work, as measured by convergence and consistency, were found for Fribble stimuli with three fixed dimensions of variation (Sec. 4.3.3); similar successful work in primate neurophysiology also has utilized strong visual constraints, synthesizing controlled blob stimuli [24, 78]. Use of stimuli found in the real world, rather than synthesized by a computer model, can provide valuable additional intuitions about cortical perception performed in more natural settings (Sec. 5.1.2); I recommend the use of natural stimuli if they can be controlled for perceptual variability.

Organization of stimuli Each stimulus is best represented through a set of parameters indicated by prior work to be important to cortical perception. In my realtime study, the use of the SIFT computer vision representation [36, 40] was motivated by my previous work indicating SIFT constitute a reasonable model for representation of visual objects in intermediate stages of the visual object processing pathway in the brain (Chap. 2)., The use of Fribbles was intended as a controlled exploration of the effects of shape and texture, observed to be relevant to cortical object perception by past studies [24, 63, 78] and by my realtime exploration of real-world object stimuli (Sec. 5.1.2). For most efficient implementation and operation of the chosen realtime stimulus selection method, the stimulus representations can be arranged into a Euclidean space, a graph/tree, or another structure. As mentioned above, the brain regions studied also optimally should be selected to best support the stimulus organization assumptions.

Realtime selection of stimuli A variety of stimulus selection methods may be used to explore the visual properties used by a selected set of voxels in the brain. Identification of visual properties most activitating a neural unit through a simplex search of feature space is appealing for its simplicity and for its success in my current work. However, the results of realtime study may be improved by modifications to the simplex simulated annealing method [7] to account for observed surround suppression — maximum response locations in visual feature space may be surrounded by visual space regions producing markedly suppressed cortical responses. Perhaps new simplex points can be accepted when they either are prominently lower *or* prominently higher than present points in a simplex. Alternative explorations of visual properties may be pursued by evolutionary algorithms, modified from Tanaka, Yamane et al. and Hung et al. [24, 63, 78], exploring stimuli slightly deviating from those producing the highest and lowest responses earlier in the scan session. Further methods for characterization of a given visual feature space may be explored, using active learning with noisy measurements for learning classification boundaries and for regressing parameterized response surfaces in the space. It is advisable to test new stimulus selection methods in simulation, accounting for expected noise, potential models of surround suppression, and potentially multiple maxima in the feature space.

Stimulus presentation design should be developed carefully to maximize signal quality and to minimize unintended perceptual biases. Dimness detection, used in Sec. 3.4.3 is preferable because it limits notions about perceptual features of importance while forcing subject attention onto the object stimuli. Fixation onset detection with passive viewing of stimuli, used in Chap. 2, is similarly appealing for its lack of potential perceptual biasing. However, it may cause less attention to the stimuli which may result in weaker measured responses. In contrast, the one-back location task in Sec. 3.3.4 should be avoided because it caused some subjects to use shape-based strategies to determine stimulus location, a strategy that may have biased their perception of objects. Other tasks also may be worth pursuing.

The use of an 8 s stimulus onset asynchrony (SOA) in my work (Sec. 3.1.3) prevents significant overlap between cortical responses to consecutive stimuli, aiding in reliable estimates of cortical responses to each stimulus. While more rapid display, i.e., lower SOA, the increased overlap between stimulus responses would hinder the interpretability of single trial measurements and would require pooling of response data over many more trials to perform optimal selection of further stimuli to study. Faster stimulus display also may require development of approaches to decrease fMRI signal processing times, to process new response data at the speed at which it becomes available.

Performance of multiple visual property searches in parallel, alternating displays between distinct classes of stimuli for each search (Sec. 3.1.1), also is an appealing approach to continue from my work. Interspersal of visually distinct stimuli lessens the risk of potential adaptation to visually similar properties, reducing what otherwise could be a strong response from the selected voxels.

Realtime fMRI processing as described in Sec. 3.1.4 is necessary to properly compute cortical responses to displayed stimuli. However, it remains an open question which stages of preprocessing are necessary and how best to compress cortical responses across time and space. Detrending, motion correction, and normalization preprocessing steps all appear to be valuable, though visual property search performance appears to be robust to reductions in motion correction (Sec. 5.2.2). Compression of cortical response across time, through fitting of an HRF, makes response computations more robust against scanner and biological noise. Further compression across voxels may add further robustness, or may obscure significant information about multi-voxel cortical encodings. I believe my weighted summation across voxels provides a compromise between the two concerns while embracing more tractible realtime search methods, requiring fewer stimulus response evaluations. However further work performing optimization, regression, or classification given a multi-dimensional output from the cortical region of interest also merits future pursuit.

6.5 Conclusions

My work develops a novel method for probing complex visual selectivities in the ventral visual pathway. Despite a variety of biological and technical challenges, I identified brain regions selective for holistic and component object shapes and for varying surface properties, further developing our understanding of the visual properties used for cortical object perception. I also found examples of “surround suppression,” in which cortical activity is inhibited upon viewing stimuli slightly deviating from the visual properties preferred by a brain region. Here we see the mechanism of surround suppression extends up the visual hierarchy, past its established use in V1 [22], supporting the notion of a perceptual advantage for local competition between “neighboring” percepts in higher levels of vision [25]. The multiple spaces I used to parameterize complex visual properties — based on SIFT [36], multidimensional scaling [55], and Fribble objects [76] — provide promising representational frameworks on which to build future studies of object perception. Chap. 2 suggests additional computer vision representations for pursuit in brain research. My method for realtime selection of stimuli, to rapidly identify stimuli most activating a given brain region, presents a further way forward in neuroimaging investigations of object vision making maximal use of limited scanning time.

Bibliography

[1] Freesurfer, 2012. URL `http://surfer.nmr.mgh.harvard.edu/`. [Online; accessed 20-August-2012]. 2.2.7

[2] A.C. Berg, T.L. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *IEEE Computer Science Conference on Computer Vision and Pattern Recognition*, volume 1, 2005. 2, 2.4.1

[3] D.H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:443–446, 1997. 2.2.3, 3.1.3

[4] C. Cadieu, M. Kouh, A. Pasupathy, C.E. Connor, M. Riesenhuber, and T. Poggio. A model of v4 shape selectivity and invariance. *Journal of Neurophysiology*, 98(3):1733–1750, 2007. 1.1, 2.1, 2.2.1, 2.4.1, 3.3, 3.3.1, 3.4

[5] A.J. Calder and A.W. Young. Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6:641–651, 2005. 1.1, 2.1

[6] J. Canny. A computational approach to edge detection. pages 679–698, 1986. 2

[7] M.F. Cardoso, R.L. Salcedo, and S.F de Azevedo. The simplex-simulated annealing approach to continuous non-linear optimization. *Computers and Chemical Engineering*, 20 (9):1065–1080, 1996. (document), 1.3, 3.1.5, 3.1.5, 4.3.3, 5, 5.2, 5.3, 6.2, 6.4

[8] V. Chandrasekhar, M. Makar, G. Takacs, D.M. Chen, S.S. Tsai, R. Grzeszczuk, and B. Girod. Survey of sift compression schemes. In *International Conference on Pattern Recognition*, 2010. 2.4.1

[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Science Conference on Computer Vision and Pattern Recognition*, volume 1, 2005. 2.5

[10] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America [A]*, 2(7):1160–1169, 1985. 1, 2.4.1

[11] Brecht Donckels. Global optimization algorithms for matlab, 2012. URL `http://biomath.ugent.be/~brecht/downloads.html`. [Online; accessed 31-May-2012]. 3.1.5, 3.1.5

[12] R. Epstein, A. Harris, D. Stanley, and N. Kanwisher. The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, 23(1):115–125, 1999. 1.1, 2.1

[13] J. Freeman and E.P. Simoncelli. Metamers of the ventral stream. *Nature Neuroscience*, 14 (9):1195–1201, 2011. 1.1

[14] W.A. Freiwald, D.Y. Tsao, and M.S. Livingstone. A face feature space in the macaque temporal lobe. *Nature Neuroscience*, 12:1187–1196, 2009. 1.1, 2.1

[15] C.R. Genovese, N.A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15:870–878, 2002. 2.2.7

[16] K. Grill-Spector, Z. Kourtzi, and N. Kanwisher. The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10-11):1409–1422, 2001. 1.1, 2.1, 2.2.3, 2.3, 2.4.1, 3.3.4

[17] K. Grill-Spector, N. Knouf, and N. Kanwisher. The ffa subserves face perception not generic within category identification. *Nature Neuroscience*, 7(5):555–562, 2004. 1.1, 2.1, 2.4.1

[18] A. Harel, S. Ullman, B. Epstein, and S. Bentin. Mutual information of image fragments predicts categorization in humans: neurophysiological and behavioral evidence. *Vision Research*, 47(15):2010–2020, 2007. 1.1, 2.1

[19] J.V. Haxby, L.G. Ungerleider, V.P. Clark, J.L. Schouten, E.A. Hoffman, and A. Martin. The effect of face inversion on activity in human neural systems for face and object perception. *Neuron*, 22(1):189–199, 1999. 1.1, 2.1

[20] J.V. Haxby, E.A. Hoffman, and M.I. Gobbini. The distributed human neural system for face perception. *Trends in Cognitive Science*, 4(6):223–233, 2000. 1.1, 2.1, 2.4.1

[21] Inc. Hemera Technologies. Hemera photo objects volumes i, ii, and iii, 2000-2003. 3.3.1

[22] D.H. Hubel and T.N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195:215–243, 1968. (document), 1.1, 1.2, 2.1, 2.4.1, 4, 5, 5.1.1, 6.5

[23] J.E. Hummel and I. Biederman. Dynamic binding in a neural network for shape recognition. *Pschological Review*, 99(3):480–517, 1992. 1, 1.2, 1

[24] C. Hung, E.T. Carlson, and C.E. Connor. Medial axis shape coding in macaque inferotemporal cortex. *Neuron*, 74(6):1099–1113, 2012. (document), 1.1, 1.2, 1.3, 2.1, 2.4.1, 3.3, 3.4, 5.1.2, 5.2, 5.3, 6.4

[25] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition. In *IEEE International Convference on Computer Vision*, volume 1, 2009. 2.5, 5.1.1, 6.5

[26] M.A. Just, V.L. Cherkassky, S. Aryal, and T.M. Mitchell. A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoSone*, 5(1), 2010. 2.2.1, 2.2.5, 3.1.4, 3.3.5

[27] K.N. Kay, T. Naselaris, R.J. Prenger, and J.L. Gallant. Identifying natural images from human brain activity. *Nature*, 452:352–355, 2008. 1.1, 1.2, 2.1, 1, 2.3, 2.4.1, 5.1.1

[28] R. Kiani, H. Esteky, K. Mirpour, and K. Tanaka. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, 97(6):4296–4309, 2007. 2.1.1

[29] B.B. Kimia, A.R. Tannenbaum, and S.W. Zucker. Shapes, shocks, and deformations i:

the components of two-dimensional shape and the reaction-diffusion space. *International Journal of Computer Vision*, 15(3):189–224, 1995. 4, 2.4.1

[30] Z. Kourtzi and N. Kanwisher. Cortical regions involved in perceiving object shape. *The Journal of Neuroscience*, 20(9):3310–3318, 2000. 2.2.3, 2.3, 3.3.4

[31] D.J. Kravitz, C.S. Peng, and C.I. Baker. Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *Journal of Neuroscience*, 31(20):7322–7333, 2011. 1.1, 2.1

[32] N. Kriegeskorte, R. Goebel, and P. Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the USA*, 103(10):3863–3868, 2006. 2, 2.2.5, 2.2.7, 3.3.5

[33] N. Kriegeskorte, M. Murr, D. Ruff, R. Kiani, J. Bodurka, H. Esteky, K. Tanaka, and P. Bandettini. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–1141, 2008. 2, 2.1, 2.1.1, 2.2.7, 2.4.1

[34] C. Kung, J.J. Peissig, and M.J. Tarr. Is region-of-interest overlap comparison a reliable measure of category specificity? *Journal of Cognitive Neuroscience*, 19:2019–2034, 2007. 2.4.1

[35] D.D. Leeds, D.A. Seibert, J.A. Pyles, and M.J. Tarr. Unraveling the compositional basis of cortical object representation. *Journal of Vision*, submitted. (document), 1.2, 1.4, 2, 3.3.2, 3.3.4, 3.3.6, 4.2.1, 5.2.3

[36] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. (document), 1.2, 1.3, 2, 2.1, 3, 2.5, 3, 3.3, 5, 6.4, 6.5

[37] Diego Macrini. Shapematcher 5, 2008. URL `http://www.cs.toronto.edu/~dmac/ShapeMatcher/`. [Online; accessed 20-September-2011]. 4

[38] MATLAB. *version 8.0.0.783 (R2012b)*. The MathWorks Inc., Natick, Massachusetts, 2012. 2.2.3, 3.1.2, 3.1.3

[39] J.A. Nelder and R. Mead. A simplex method for functional minimization. *The Computer Journal*, 7(4):308–313, 1965. 3.1.5, 4.3.1

[40] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *Computer Vision - ECCV 2006*, volume 3954 of *Lecture Notes in Computer Science*, pages 490–503. 2006. 3, 2.4.1, 6.4

[41] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001. 5, 2.4.1

[42] B.A. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set: a strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997. 2.4.1

[43] A.J. O'Toole, F. Jiang, H. Abdi, and J. Haxby. Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17(4): 580–590, 2005. 2.1.1

[44] D.G. Pelli. The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10:437–442, 1997. 2.2.3, 3.1.3

[45] D.I. Perrett, P.A. Smith, D.D. Potter, A.J. Mistlin, A.S. Head, A.D. Milner, and M.A. Jeeves. Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. *Human Neurobiology*, 3(4):197–208, 1984. 1.1, 2.1

[46] Brian Pittman. Afni main page — afni and nifti server for nimh/nih/phs/dhhs/usa/earth, 2011. URL `http://afni.nimh.nih.gov/afni`. [Online; accessed 20-September-2011]. 2.2.5, 3.1.1, 3.3.5

[47] J.A. Pyles and E.D. Grossman. Neural adaptation for novel objects during dynamic articulation. *Neuropsychologia*, 47(5):1261–1268, 2009. 2.2.3, 3.3.4

[48] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, 1999. 1, 2.1, 2.4.1, 5.1.2

[49] R.W. Rodiek and J. Stone. Analysis of receptive fields of cat retinal ganglion cells. *Journal of Neurophysiology*, 28:833–849, 1965. 5.1.1

[50] E.T. Rolls and T. Milward. A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation*, 12(11):2547–2572, 2000. 5.1.1

[51] Ziad Saad. Suma - afni surface mapper – afni and nifti server for nimh/nih/phs/dhhs/usa/earth, 2006. URL `http://afni.nimh.nih.gov/afni/suma/`. [Online; accessed 20-August-2012]. 2.2.7

[52] Alexander Schrijver. *Theory of Linear and Integer Programming*. John Wiley & sons, 1998. 3.3.2

[53] J. Schultz and K.S. Pilz. Natural facial motion enhances cortical responses to faces. *Experimental Brain Research*, 194(3):465–475, 2009. 2.2.3, 3.3.4

[54] P.G. Schyns, L. Bonnar, and F. Gosselin. Show me the features! understanding recognition from the use of visual information. *Psychological Science*, 13(5):402–409, 2002. 2.4.1

[55] G.A. Seber. *Multivariate Observations*. John Wiley and Sons, Inc., Hoboken, NJ, 1984. 3.3.2, 6.5

[56] D.A. Seibert, D.D. Leeds, J.A. Pyles, and M.J. Tarr. Exploring computational models of visual object perception, 2012. Poster presented at Vision Sciences Society Annual Meeting, Naples, Florida. 2.4.1

[57] T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 109(21):6424–6429, 2007. 1.1, 2.1, 2.4.1, 5.1.1

[58] D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. pages 517–524, 1968. 4.4.1

[59] K. Shibata, T. Watanabe, Y. Sasaki, and M. Kawato. Perceptual learning incepted by de-

coded fmri neurofeedback without stimulus presentation. *Science*, 334(6061):1413–1415, 2011. 1.1

[60] K. Siddiqi, A. Shokoufandeh, S.J. Dickinson, and S.W. Zucker. Shock graphs and shape matching. *International Journal of Computer Vision*, 35(1):13–32, 1999. 4

[61] J.G. Snodgrass and M. Vanderwart. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2):174–215, 1980. 2

[62] J.D. Swisher, J.C. Gatenby, J.C. Gore, B.A. Wolfe, C.-H. Moon, S.-G. Kim, and F. Tong. Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *Journal of Neuroscience*, 30(1):325–330, 2010. 2.2.5, 3.1.4, 3.3.5

[63] K. Tanaka. Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*, 13 (1):90–99, 2003. (document), 1.1, 1.2, 2.1, 2.4.1, 3.3.6, 5.1.2, 5.3, 6.4

[64] Michael Tarr. Novel object – the cnbc wiki, 2013. URL `http://wiki.cnbc.cmu.edu/Novel_Objects`. [Online; accessed 15-January-2013]. 3.4.1

[65] M.J. Tarr and I. Gauthier. Ffa: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3(8):764–769, 2000. 2.4.1

[66] B.S. Tjan. Adaptive object representation with hierarchically-distributed memory sites. In *Advances in Neural Information Processing Systems*, volume 13, pages 66–72, 2001. 2.5

[67] Antonio Torralba. Spatial envelope, 2006. URL `http://people.csail.mit.edu/torralba/code/spatialenvelope/`. [Online; accessed 20-September-2011]. 5

[68] D.Y. Tsao and M.S. Livingstone. Mechanisms of face perception. *Annual Review of Neuroscience*, 31:411–437, 2008. 1.1, 2.1

[69] S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5:682–687, 2002. 1, 1.1, 2.1

[70] Andrea Vedaldi. A. vedaldi - code - sift for matlab, 2011. URL `http://www.vlfeat.org/~vedaldi/code/sift.html`. [Online; accessed 20-September-2011]. 3

[71] R. Vogels. Effect of image scrambling on inferior temporal cortical responses. *Neuroreport*, 10(9):1811–6, 1999. (document), 1.2, 2.4.1

[72] B.A. Wandell, S.O. Dumoulin, and A.A. Brewer. Visual field maps in human cortex. *Neuron*, 56:366–383, 2007. 5.1.3

[73] H.X. Wang, D.J. Heeger, and M.S. Landy. Responses to second-order texture modulations undergo surround suppression. *Vision Research*, 60:192–200, 2012. (document), 1.2, 4, 5.1.1

[74] B.D. Ward, J. Janik, Y. Mazaheri, Y. Ma, and E.A. DeYoe. Adaptive kalman filtering for real-time mapping of the visual field. *Neuroimage*, 59(4):3533–3547, 2011. 1.1

[75] Martin Wennerberg. *version 2.9.5*. Norrkross Software, 2009. 3.4.1

[76] P. Williams and D.J. Simons. Detecting changes in novel, complex three-dimensional objects. *Visual Cognition*, 7:297–322, 2000. (document), 1.2, 1.3, 3, 3.4, 3.4.1, 4, 5, 6.5

[77] A.C.-N. Wong, T.J. Palmeri, B.P. Rogers, J.C. Gore, and I. Gauthier. Beyond shape: how you learn about objects affects how they are represented in visual cortex. *PLoS ONE*, 4 (12):e8405, 2009. 2.4.1

[78] Y. Yamane, E.T. Carlson, K.C. Bowman, Z. Wang, and C.E. Connor. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*, 11(11):1352–1360, 2008. 1.1, 1.2, 2.1, 2.2.1, 2.4.1, 3.3.1, 3.4, 5.1.2, 5.2, 5.3, 6.4

[79] L. Zhang and G.W. Cottrell. Holistic processing develops because it is good. In *Annual Cognitive Science*, 2005. 2.4.1

[80] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005. 3.5.1