Modeling the Representation of Medial Axis Structure in Human Ventral Pathway Cortex

by

Haluk N. Tokgozoglu

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

April, 2016

Abstract

Computational modeling of the human brain has long been an important goal of scientific research. The visual system is of particular interest because it is one of the primary modalities by which we understand the world. One integral aspect of vision is object representation, which plays an important role in machine perception as well. In the human brain, object recognition is a part of the functionality of the ventral pathway. In this work, we have developed a computational and statistical techniques to characterize object representation among this pathway. The understanding of how the brain represents objects is essential to developing models of computer vision that are truer to how humans perceive the world.

In the ventral pathway, the lateral occipital complex (LOC) is known to respond to images of objects [31]. Neural recording studies in monkeys have shown that the homologue for LOC represents objects as configurations of medial axis and surface components [25, 49, 65]. In this work, we designed and implemented novel experiment paradigms and developed algorithms to test whether the human LOC represents medial axis structure as in the monkey models. We developed a data-driven iterative

ABSTRACT

sparse regression model guided by neuroscience principles in order to estimate the response pattern of LOC voxels. For each voxel, we modeled the response pattern as a linear combination of partial medial axis configurations that appeared as fragments across multiple stimuli. We used this model to demonstrate evidence of structural object coding in the LOC. Finally, we developed an algorithm to reconstruct images of stimuli being viewed by subjects based on their brain images. As a whole, we apply computational techniques to present the first significant evidence that the LOC carries information about the medial axis structure of objects, and further characterize its response properties.

Primary Reader: Prof. Gregory D. Hager

Secondary Reader: Prof. Charles E. Connor

Secondary Reader: Prof. Russell H. Taylor

Acknowledgments

First I would like to thank my advisors. Professor Gregory G. Hager, whose guiding influence has constantly pushed me to think about my work more critically and structurally, whose flexibility and understanding has given me the freedom to pursue the work that truly interested me. He has taught me to think critically and his encouragement for me to join the PhD program has provided me with the best opportunity of my life. Professor Charles E. Connor, whose relentless dedication to helping me navigate fields of research I am unfamiliar with, whose kind and friendly guidance has helped me direct my work and find motivation when it seemed like my data had abandoned me. I would like to thank you both for always supporting me and believing in me. I would also like to thank Professor Russell H. Taylor, who has taught me many things over many years in our many different interactions. He gave me the opportunity to work as his teaching assistant when I was trying to figure out what to do with my research, and what I thought would be a secondary assignment turned into a fulfilling learning experience.

I would also like to thank the late Professor Steven G. Yantis. I am thankful for

his taking a chance on me and supporting me. Without his guidance and support this work would not have been possible. I am saddened that he was not able to see the results our work together. I would not have been here today without the support of Dr. Yantis or my advisors.

I would like to thank Professors Ralph Etienne-Cummings, Han Liu and Randal Burns for their participation on my GBO committee. They all helped me, along with my advisors, strengthen myself as a researcher and ask myself the questions that I needed to ask in order to get my research to where it needed to be. I would additionally like to thank Professor Michael Kazhdan, Dr. Daniel Mirota and especially Dr. Eric Meisner for their help and friendship in my early years as a PhD student.

I would like to thank Professor Scott Smith, Cathy Thornton and Debbie DeFord for their help, guidance and welcoming attitude over the years. They have helped me through my years at Johns Hopkins in many ways, making my academic experience as enriching as possible. I would also like to thank Zach Burwell and Tracy Marshall for their help in facilitating the defense of this very dissertation.

I would like to thank Drs. Brian A. Anderson and Anthony W. Sali. Their help, friendship and endless patience, along with their wisdom and insight has made this work possible. They have been with me every step of the way. I would especially like to thank Dr. Sali, as he introduced me to Professor Yantis and Professor Connor in the first place. I would also like to thank Professor Michael Yassa and Zachariah Reagh for their help in developing our imaging technique. I would also like to thank

ACKNOWLEDGMENTS

Dr. Marcin Szwed for his help in implementing the localizer.

I would like to thank members of the Computational Interaction and Robotics Laboratory. For many years, their help and friendship has been invaluable. Thanks to Dr. Kel Guerin for all the fun conversation, the work we've done together in robotics, virtual reality and the display wall, and for all other forms of generosity and friendship. Thanks to Jon Bohren for his help on robotics, and all forms of stimulating conversation. Thanks to Colin Lea for his friendship and theoretical expertise that he lent to me whenever I had questions. Thanks to Berk Gonenc and Dr. Tutkun Sen for their companionship. Further thanks to all present and past members of CIRL that have been part of my academic career: Dr. Nicolas Padoy, Dr. Swaroop Vedula, Dr. Sharmishtaa Seshamani, Anand Malpani, Amod Jog, Chris Paxton, Dan Abretski, Amanda Edwards, Carol Reiley, Chi Li, Xiang Xiang, Narges Ahmidi, Dr. Ioana Fleming, Dr. Daniel Mirota, Dr. Zach Pezzementi and Dr. Raphael Sznitman.

I would also like to thank my friends and fellow colleagues throughout my years in the Graduate Representative Organization. Their friendship and mentorship throughout the many ordeals we endured together will be remembered. Specifically, I would like to thank Dr. John Matsui for his leadership, Dr. Dominik Gothe for his stimulating and entertaining friendship, Dr. Manu Madhav for his guidance and friendship, Ian Miers and Christina Garman for their friendship and enduring dedication to the GRO, Linda Braun for her friendship and hard work for the GRO. I would also like to thank Dr. Kitty Z Xu for her friendship and help with my research.

ACKNOWLEDGMENTS

I am also glad to have been a small part of the large envelope of the Laboratoty for Computational Sensing and Robotics. I would like to thank Anton Deguet and Xingchi He for their help over the years.

I would like to thank three professors from Bilkent University who set me on this long path that brought me here today. David Davenport, William Sawyer and Mehmet Bayar. Without their friendly encouragement, help and guidance, I would not have pursued graduate school. They were integral in my becoming a computer scientist. I like to thank all my friends and family who have helped me along this journey. I would specifically like to thank my parents, Professors Mazhar and Lale Tokgozoglu, and my grandparents Perihan and Akin Altinok. It is not possible to state the extent of their contribution to my life and work without overwhelming this document. Finally, I would like to thank my fiancée Merve Kaya, who has been an inseparable part of my life and supported me both personally and academically, making this work possible.

Haluk Noyan Tokgozoglu

Baltimore,

21 April 2015

Dedication

This thesis is dedicated to my family. Without their encouragement and support I would not be here today.

Contents

Abstract **ii**

Acknowledgments **iv**

List of Tables xiv

List of Figures xv

1 Introduction **1**

1.1 Motivation 3

1.2 Proposed Approach 6

1.3 Thesis Statement 9

1.4 Overview of Contributions 9

1.4.1 Contribution 1 10

1.4.2 Contribution 2 10

1.4.3 Contribution 3 10

1.4.4 Contribution 4 11

1.5 Attribution 11

1.6 Notation and Definitions 11

CONTENTS

2 Experiment and Stimulus Design **13**

2.1 Overview 15

2.2 Related Work 18

2.2.1 Experiment design 18

2.2.2 Neurophysiology Studies Motivating Stimulus Design 21

2.3 Experiment Setup And Equipment 23

2.4 Stimulus Shape Generation 28

2.5 Cognitive Task Design 34

2.6 Localizer Experiment Designs 42

2.6.1 LOC Localizer Design 43

2.6.2 VWFA Localizer Design 45

2.7 Experiment Software 48

2.8 Complications And Configurations Of Experiments 55

2.9 Conclusions 58

3 Preprocessing of fMRI Data **62**

3.1 Overview 63

3.2 Related Work 65

3.3 fMRI Data Format 67

3.4 Registration 70

3.5 Localizers And Noise 75

3.5.1 LOC Localizer Results 77

3.5.2 VWFA Localizer Results 79
3.6 Extracting Stimuli 81
3.7 Conclusions 81

4 Statistical Modeling Of Structural Representation In LOC 83

4.1 Overview 84
4.2 Related Work 87
4.2.1 General Guidelines For fMRI Analyses 87
4.2.2 Neurophysiology Studies On Object Structure 89
4.2.3 fMRI Studies On The Visual Cortex 91
4.2.3.1 Early Visual Areas 91
4.2.3.2 Lateral Occipital Complex 93
4.2.3.3 Visual Word-Form Area 93
4.2.4 Statistical Modeling Techniques 94
4.3 Ascertaining The Distinguishability Of Stimuli 95
4.3.1 Initial Study 95
4.3.2 Final Studies 102
4.4 Developing a Computational Model 103
4.4.1 Developing A Novel Modeling Approach 112
4.4.2 Advantages of Two-Group Fitting 115
4.5 Validating The Model 117
4.5.1 Cross-Validation Analyses 118

4.5.1.1 Leave-One-Out Cross-Validation 118
4.5.1.2 Leave-Two-Out Cross-Validation 119
4.5.1.3 Leave-Six-Out Cross-Validation 120
4.5.2 Applying Our Model To The Visual Word-Form Area 121
4.5.3 Applying V1 Models To The LOC 122
4.6 Conclusions 132

5 Reconstruction Of Stimuli Using fMRI Images 135

5.1 Overview 136
5.2 Related Work 137
5.2.1 Neurophysiology Studies On Object Structure 138
5.2.2 fMRI Studies On The Visual Cortex 140
5.2.2.1 Early Visual Areas 140
5.3 The Reconstruction Process 142
5.3.1 Computing Candidate Fragments 142
5.3.2 Rank-Based Fragment Selection 143
5.4 Results 147
5.5 Practical And Future Applications 159
5.6 Conclusions 160

6 Conclusions 163

6.1 Discussion of Contributions 166

6.2 Limitations 169

6.3 Future Work 170

Vita **183**

List of Tables

2.1 The performance of each subject on the one-back task. True Positive and False Positive presses from each subject are denoted. 40

2.2 The set of nouns used for the VWFA localizer. 6-letter neutral English nouns with an occurence rate of over a million in the database of contemporary American English [62] 46

4.1 The accuracy of choosing between the predicted values of two left-out stimuli. 119

List of Figures

2.1 Comparison of a regular fMRI image with voxels of size $3mm \times 3mm \times 3mm$ acquired with a 3 Tesla scanner to our paradigm that uses voxels of size $1.5mm \times 1.5mm \times 1.5mm$ acquired with the same scanner. The field of view is aligned along the visual cortex. Note the oblique angle of the high resolution image. 27

2.2 Samples from the original stimulus set. Fixation point is displayed as it was presented to the subjects, and it highlights the fact that many stimuli had quadrants with no fragments in them. The original stimulus set had stimuli that went at most two fragments deep from the fixation point. 33

2.3 The entire stimulus set used for the final experiment. 35

2.4 Demonstration of the one-back task for main experiment and the viewing task of the localizer runs. Stimuli are presented at a time interval, and the subject is required to attend to them. 39

LIST OF FIGURES

2.5 The entire stimulus set used for the VWFA localizer. Note that inside the scanner these stimuli were legible, which does not come across as well in this format. Subjects were asked post-experiment if they were able to read all the words, and they responded positively. 47

2.6 The Cedrus RB-830 response box used in our experiments. This device sends all fMRI-related signals over the serial port. 49

2.7 The state machine used by the software. Initially, the program runs a loop as it waits for the scanner pulse. Then, when the pulse is received, it schedules all stimulus-related activities, and gives the scheduler control. The scheduler ensures a regular update loop, in addition to calling stimulus-related event handles when needed. When the run is over, the subject is given a brief period to rest, after which the software starts waiting for a pulse again. 52

2.8 Flow of the experiment. Black lines indicate the signal's start pulses, which trigger each part of the run. Red lines are the stimuli being displayed through the projector. Blue lines are the subject's responses, which are a button press depending on a query stimulus presented. . 54

LIST OF FIGURES

2.9 Image of the acquisition FOV. The yellow box indicates the limits of the FOV. In this instance the angle of the visual cortex wasn't very oblique, but we had other scans where the field was up to -45 off degrees from the X-axis of the left image. The blue areas were used as virtual fixtures to constrain the FOV placement and weren't used for purposes of the analyses. 57

3.1 The registration process. 74

3.2 Depiction of the GLM model for an imaginary voxel with time-series Z predicted by a design matrix X including regressors of interest e.g., 3 task regressors and seven nuisance regressors e.g., six motion parameters and one linear drift of unknown amplitude, and an error term. . 77

3.3 LOC localizer GLM results for Subject 1. The color map from green to red indicates the f-statistic of the GLM. 78

3.4 VWFA localizer GLM results for Subject 2. Note that the images for Subject 2 were acquired at a very oblique angle (as per Section 2.8) and thus are being viewed from an unconventional perspective. The color map from green to red indicates the f-statistic of the GLM. . . . 80

4.1 Plot of the area under the ROC curve for each fragment from the initial experiment. The fragments are sorted by the area under ROC value. 97

LIST OF FIGURES

4.2 Leave-One-Out results. Each color represents a different LOO trial, and their residual error percentage according to NRMSE is displayed. Voxels are not sorted. The error percentage is quite high for all trials. 100

4.3 The set of all fragments used in our final model, except for the "blank" fragments. The figure is arranged such that every row is one quadrant (demonstrated by the graphic on the first column) and the fragments in each column are consecutive 90 deg rotations of each other. The first five fragments in each row are the individual fragments, the rest are compound fragments. 103

4.4 Plot of the area under the ROC curve for each fragment from the final experiment, compared to the initial experiment. The fragments are sorted by the area under ROC value. Note that the initial study had 200 fragments, and the final study has 44, so the initial study's curve is resampled down to 44 points for the purposes of this comparison using linear interpolation. The original curve can be seen in Figure 4.1 . . . 104

4.5 The residual error percentage according to NRMSE of applying a linear regression solution to Equation 4.8. Voxels are not sorted. 106

4.6 The residual error percentage according to NRMSE of applying a ridge regression solution to Equation 4.8. Voxels are not sorted. 108

LIST OF FIGURES

4.7 The residual error percentage according to NRMSE of applying lasso. Only Subject 1's results are displayed as they were the best. The mean NRMSE across all voxels is calculated, with the error bars showing cross-validation performance. 109

4.8 The residual error percentage according to NRMSE of applying the BSR approach in Algorithm 3 111

4.9 The residual error percentage according to NRMSE of applying the TGF approach in Algorithm 4 115

4.10 The residual error percentage according to NRMSE of applying each approach. The NRMSE curves are averaged across each subject via resampling, then displayed against each other. Note that our TGF algorithm performs comparably to Backwards Stepwise Regression. . 116

4.11 A demonstration of the Two-Group Fitting algorithm for a sample voxel. Every row represents the weights assigned to each fragment at a particular iteration, with a gradient from red to black denoting normalized values between 1 and 0. At each step, one fragment is removed from the model by setting its weight to zero. 123

4.12 The correlation coefficients between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for each subject. 124

4.13 The correlation coefficient between the predicted and actual value of the stimulus set according to Leave-One-Out analyses for each voxel. 125

4.14 The correlation coefficients between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for each subject. 128

4.15 The correlation coefficient between the predicted and actual value of the stimulus set averaged across all L6O analyses for each voxel for each subject. 129

4.16 Performance of the structural parametrization on VWFA voxels compared to the LOC voxels. 130

4.17 Performance of the V1-parametrization compared to the structural parametrization on LOC voxels. 131

5.1 Projecting the novel stimulus onto the line between the means of the present and absent clusters. 143

5.2 The rank sum of each stimulus for each subject. Perfect reconstruction would have a rank of 44. 150

5.3 The number of correctly predicted fragments for each stimulus. 151

5.4 The correlation coefficients between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for each subject. 151

5.5 Comparison of a reconstruction set (below) to the stimulus set (above). For most fragments, even when they are incorrectly predicted, their appearance is close to the actual fragment. 152

5.6 The confusion matrix displaying how many times a given fragment has been mis-predicted as another fragment within the 60 stimuli. 153

5.7 Demonstrating the reconstruction of stimulus 5. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. This reconstruction has an error in the second and fourth quadrants, which correspond to the quadrants facing 90 and 270 degrees, respectively. In both cases, instead of predicting an empty fragment, the algorithm predicted a straight line. In general, the algorithm seems to not favor predicting empty fragments. 154

5.8 Demonstrating the reconstruction of stimulus 6. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. This reconstruction has an error in the first quadrant, facing 0 degrees. Again, the algorithm prefers a straight line over an empty fragment. However, in this case it is able to correctly predict and empty fragment in the third quadrant, facing 180 degrees. 155

5.9 Demonstrating the reconstruction of stimulus 33. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. This stimulus is reconstructed accurately. The algorithm does better with three-fragment stimuli than it does with two-fragment stimuli, as can be seen in Figure 5.4. 156

5.10 Demonstrating the reconstruction of stimulus 45. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. Similar to Figure 5.7a, the algorithm predicts a straight line in place of an empty fragment. 157

5.11 Demonstrating the reconstruction of stimulus 57. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. This stimulus is reconstructed accurately. The algorithm does better with three-fragment stimuli than it does with two-fragment stimuli, as can be seen in Figure 5.4. 158

Chapter 1

Introduction

Understanding the workings of the human brain is an ongoing pursuit. The visual system in particular plays a huge role in perception, and thus characterizing its behavior is key to this understanding. While the behavior of earlier visual areas is rather well understood, higher-level areas in the visual system remain hard to define, especially in humans. Even in areas where the general capability of the region is known, the way it specifically translates a given stimulus into brain signals is harder to define. Since earlier visual areas (e.g. V1) code for visual stimuli in more direct ways, there is a large body of work identifying those areas. Higher level areas have more complex schema for representing visual information, and thus the exact way they represent visual information isn't know, even though their general task is. In general, how the human brain represents structural characteristics of objects is not well-defined. Neurophysiological work in monkeys has led to significant understand-

ing of these representations, however due to the invasive nature of this process it is not applicable to humans.

Technological advancements in the past few decades have made such analyses possible, enabling non-invasive measurement of brain activity. The key tool for mapping out the functionality of the human brain in this manner is functional magnetic resonance imaging (fMRI). Measuring blood flow to parts of the brain using magnetic fields caused by blood oxygenation differences, fMRI has made it possible to image the human brain as it operates. Use of fMRI has enabled us to delineate brain areas based on their functional characteristics. The problem with fMRI is that it measures a complex signal that is affected by many factors, and it requires precise experiment design and statistical analysis to truly leverage to its full extent. Higher visual areas that are purely functionally defined are especially difficult to image and model. These statistical analyses and experiment protocols must be driven by algorithms and tools created specifically for the purposes of each experiment and visual area. In this work we will discuss a set of computational models, experiment paradigms and software developed for the purpose of measuring and modeling activity in the human lateral occipital complex.

1.1 Motivation

This dissertation is concerned with the problem of developing a computational model of how the visual cortex represents object structure. If the visual stimulus an eye receives can be thought of as analogous to a 2-dimensional image, the functionality of early visual areas could be described as a direct product of the image's pixel content, based on linear filters. Further along the ventral pathway, in higher visual areas, the exact nature of how the visual stimulus is represented is unknown. Specifically, it is known that the lateral occipital complex (LOC) represents structural information for visual stimuli in some fashion, but a concrete parametrization has not been established.

This is partly due to challenges associated with fMRI. Unlike neurophysiological recordings, fMRI does not directly measure the activity of neurons. Instead, it measures the blood oxygenation level changes in a unit of 3-dimensional space known as a volumetric pixel (voxel). This signal is a function of the neural response, but not in a direct fashion. The activity of neurons causes blood flow to blood vessels contained within the voxel, which increases the oxygenation within the voxel. This signal is called the blood-oxygen-level-dependent (BOLD) signal. Studies measuring the electrophysiological signal from a certain region and the fMRI BOLD signal from that region have shown a direct correlation between those signals, which means fMRI can be used as an extremely coarse approximation for neural recordings.

While fMRI is an essential tool for analyzing functionality of brain areas, it is

also one with many limitations. The spatial resolution of a voxel is limited to a few milimeters, which means that a voxel usually contains hundreds of thousands of neurons. As a result, trying to ascertain fine behavior of the brain is challenging. The BOLD response to a given stimulus is not a value that is trivially extractable either. Scanners are prone to magnetic interference and drift in signal, subjects often get uncomfortable and move within the scanner, and BOLD signal is inherently noisy. To counteract this, clever experiment design and computational models are necessary for obtaining meaningful information from fMRI.

The task of defining object representation in the LOC is difficult in itself. While work in monkeys shows that the homolog for LOC, inferotemporal cortex (IT) codes for objects in terms of shape, orientation and position in a continuous manner, the limitations of fMRI mean that such a parametrization is too fine to verify in humans. To obtain a robust characterization of the BOLD signal for any given stimulus, many repetitions of the image need to be shown to participants, and to define a "shape space" where shapes are parametrized sufficiently requires many different stimuli to be shown. Coupled with limitations on the attention span of participants, this requires solving experimental design challenges that explore both the depth and breadth of any given parametrization within the mimimum amount of time. Such an experiment design would also require the development of special-purpose methods to accommodate and implement it.

Given a robust stimulus set and an efficient experiment, there still remains the

task of preparing, then analyzing this data to show evidence of the parametrization being a valid way of representing how the LOC responds to object images. Once this parametrization is established with a computational model, it would need to be tested to ensure its validity and specificity. The viability of the computational model would also need to be verified separately.

Our goal is to design an fMRI experiment that can provide robust enough signal for a set of objects that are varied enough to be parametrized such that we can show evidence for coding of object structure in LOC using statistical analyses. Such analyses would establish the LOC as a brain area that codes for object shapes in a parametric fashion. The establishment of such a coding would enable the reconstructing of visual stimuli based on fMRI images. This would set the path for future research in exploring a more concrete representation of the function of LOC in humans, and be the next step in modeling the visual system after the early areas. These goals, however, come with significant computational challenges. The experiment design would require algorithms to generate visual stimuli specific to the task of modeling LOC activity, and experiment suite to interface with fMRI scanners and run the procedure facilitating the experiment. The fMRI data, once acquired, would require processing with imaging and registration methods to be ready for statistical analysis. The analyses themselves would require the development of a statistical modeling algorithm that takes into account the nature and domain of the data. Finally, the reconstruction process would also require the design of an algorithm for this specific purpose.

1.2 Proposed Approach

Concretely, the problems we are attempting to solve are:

> How can we design an experiment paradigm and computationally implement this paradigm in practice? How can we algorithmically create a visual stimulus set for use in fMRI to incite the LOC? How can we prepare LOC images using pre-processing and registration methods for statistical analysis? How can we design an algorithm to model the activity of LOC voxels based on object parameters, and how can we validate these algorithms? Can we reconstruct stimuli subjects are viewing based on their brain activity using these methods?

The approach we take in this work to solve this problem is based on work in monkeys, showing that neurons in the IT code for object structure in terms of curvature, orientation and polar position. Since the LOC is the human homolog for IT, we expect there to be similar representation of object structure in LOC. To do this, we need to develop a stimulus set consisting of images that contain variations in structure similar to the coding of IT. Due to limitations of fMRI, this stimulus set should be one that contains many presentations of each stimulus, and should represent a large enough variety of parameters within "structure-space" to be able to cut through noise and other fMRI artifacts. Such a stimulus set would be difficult to design by hand, so we would need to design an algorithm to generate these stimuli based on parameters, then use this algorithm to help us choose an optimal stimulus set.

In order to make the statement that this structural coding parametrization is specific to the function of LOC, it is necessary to show that this parametrization is inapplicable in brain areas that code for similar shapes in different ways, and that

other non-structural parametrizations aren't applicable in the LOC. The visual word-form area (VWFA), which codes for letter-like shapes similar to the medial axis stimuli that this work relies on, is a good candidate for the former analysis. Conversely, existing work shows that V1 codes for objects with a spectral coding based on Gabor Wavelet Pyramids (GWP), and showing that this coding does not work on LOC voxels would be a candidate for the latter. To further solidify this parametrization as a model for the behavior of LOC, it is also necessary to demonstrate that this parametrization can be used to reconstruct stimuli being viewed by subjects based on activity in their LOC. We will need to design an algorithm that is capable of fitting statistical models to all of these brain areas and object parametrizations, given the degenerate data conditions of fMRI.

To achieve all of this, it is necessaryto design an experiment where an abundance of images can be delivered to subjects within the constraints of a scanning run. It is also important to leverage as much data as possible from each scan, both in terms of time and spatial resolution. This will require very precise experiment design. Such experiment design requires the development of a software tool that is highly configurable and temporally precise. The design of this tool also needs to be intertwined with the stimulus generation process. Following that, the data needs to be pre-processed in a way that takes into account the limitations of the experiment and goals of the analysis. These pre-processing steps also require the design of a pre-processing software tool developed with the constraints imposed by the other aspects of the process in

mind.

Our goals translate to the following questions. These are the questions we need to answer in order to provide evidence of the lateral occipital complex coding for object structure:

1. *How can we create computational methods based on neuroscience principles that are capable of preparing stimuli and executing an fMRI experiment to obtain data about very fine features from a complex brain area?*

2. *How can we use conventional registration methods to pre-process narrow-field-of-view fMRI images and identify certain brain areas?*

3. *Can we develop a computational model that can describe the activity of LOC voxels informed by a structural parametrization of medial axis stimuli?*

4. *Can we develop an algorithm that uses the parametrization of medial axis stimuli to reconstruct images being viewed by a subject from their LOC voxel activity?*

The goal of this thesis is to answer these four questions. This thesis will be laid out in the following fashion. Chapter 2 will address the first challenge. We will discuss our experiment design and imaging paradigm as motivated by neuroscience, and the implementation details of the methods we designed to facilitate them. Chapter 3 will address the second challenge. We will detail the process we use to take the raw output of the fMRI scanner and turn it into data that can be interpreted in a spatial and temporal fashion. Chapter 4 will begin addressing the main point of this

work, which is the third challenge. We will develop a computational model for how shape parameters can be used to explain the behavior of LOC voxels, and validate our model. Finally, Chapter 5 will complete the argument from the previous chapter by addressing the last challenge, validating the parametrization by developing an algorithm that can predict stimuli being viewed by the subjects.

1.3 Thesis Statement

Informed by neuroscience principles and knowledge, it is possible to computationally model the human lateral occipital complex's activity with a model based on parametrizing the medial axis structure of objects.

1.4 Overview of Contributions

This thesis is composed of four key contributions. These contributions provide new approaches to addressing the challenges outlined above, starting with the ultimate goal of building a computational model of human LOC activity, followed by the algorithms and principles developed to facilitate this goal.

1.4.1 Contribution 1

A Computational Model For Explaining The Activity of LOC Voxels Using Medial Axis Parametrization of Stimuli Our first contribution is showing that the activity of LOC voxels can be modeled using a data-driven sparse linear regression algorithm that reduces overfitting computational model based on structural parametrization of medial axis object stimuli.

1.4.2 Contribution 2

Reconstructing Stimulus Images from LOC Activity Our second contribution is the development of an algorithm for reconstructing stimulus images from LOC voxel activity.

1.4.3 Contribution 3

Experiment Paradigms For Identifying Object Representation in LOC Our third contribution is a set of experiment paradigms, algorithms and tools for creating visual stimuli and facilitating an fMRI experiment with them, with the purpose of facilitating Contributions 1 and 2.

1.4.4 Contribution 4

High Resolution Visual Cortex Imaging Our fourth contribution is a set of principles and an imaging paradigm conducting fMRI experiments designed specifically to acquire and process higher-than-normal resolution images of the visual cortex using conventional scanners, and the process for registering these images.

1.5 Attribution

The entirety of this work was done by the author under the guidance of Charles E. Connor and Steven Yantis. Anthony W. Sali and Brian A. Anderson assisted with the experiment design in chapter 2 and pre-processing in 3. The narrow-view fMRI technique was developed with the guidance of Zach Reagh and Michael Yassa.

1.6 Notation and Definitions

In this section we will define terms and acronyms that are commonly used throughout this work.

- **BOLD** - Blood-oxygen-level dependent signal. The method with which fMRI images are acquired.
- **fMRI** - Functional Magnetic Resonance Imaging.
- **GLM** - Generalized Linear Model.

- **HRF** - Hemodynamic Response Function. The curve that characterizes the change in BOLD signal as a result of neuron stimulation.
- **IT** - Inferotemporal Cortex. The homolog of LOC in monkeys.
- **LOC** - Lateral Occipital Complex. A human visual area that is thought to respond to images of objects.
- **Medial Axis** - The medial axis of an object is the set of all points having more than one closest point on the object's boundary. In the context of this work, we use it to refer to our stimuli which consist of straight and curved lines.
- **Tesla** - Unit of magnetic field. Generally used to refer to the strength of an fMRI scanner.
- **Voxel** - Volumetric Pixel. Three-dimensional analogue of a pixel.
- **VWFA** - Visual Word-Form Area. A human visual area that is thought to respond to images of words.

Chapter 2

Experiment and Stimulus Design

Our goal in this chapter is to design a stimulus set and parametrization along with an fMRI experiment in order to be able to characterize the responses in LOC in terms of object structure. Both of these tasks require the development of specialized software tools, and these tools are intertwined in their functionality and operation. While there is a significant amount of work concerned with creating stimuli that can be parametrized and mapped to earlier brain areas, the complexity of the function of LOC means that a different approach is necessary. Characterizing such a complex behavior requires a more expansive stimulus set.However, with increased complexity in the stimulus design comes a need to make a large stimulus set to cover the space of objects to be parametrized. Creation of such a stimulus set by hand would be inefficient, thus requiring a program to generate and select stimuli.

When it comes to the experiment, each stimulus also needs to be presented several

times to eliminate fMRI noise and artifacts, and thus the number of stimulus presentations to the subject during the fMRI experiment grows quickly. As the number of presentations increases, the length of the experiment also increases, which means subjects need to spend long periods of time in the scanner. The longer subjects stay in the scanner, the easier it is for them to get distracted and lose focus, hence the experiment should be designed to account for all these factors. While numerous previous works have designed experiments to address similar problems, the fineness of the parametrization this work is looking to formulate requires a more focused experiment software created specifically for this problem.

The rest of the work in this thesis depends entirely on the strength of the experiment, the stimulus set and the parametrization, as without a set of data that is robust enough to overcome the challenges associated with fMRI, it is not possible to infer meaningful information from series of fMRI images. Without a good enough signal-to-noise ratio, the statistical modeling of the brain's activity will fail, and reconstruction will not work, thus the cognitive paradigms, neurologically-based stimuli and the programs developed to facilitate them in this chapter are key to the entirety of the work.

2.1 Overview

In this chapter we will discuss the steps we took to designing the principles of the stimulus set to be used in the rest of the work and the algorithm used to generate these stimuli. We will also describe the experiment paradigm used to display these stimuli to the subjects in ways to maximize the amount of information that can be obtained from a scan, and the development of the software that facilitates this paradigm. There are several design elements which will be explored:

1. **Design, number and complexity of stimuli to present**

   Our initial study showed that using many stimuli that are complex is not tractable. This chapter will expand upon the algorithmic generation process of the stimuli, the initial study and explain how we arrived at the final stimulus set.

2. **Timing of stimulus presentations**

   In tandem with the amount of stimuli, the timing of the presentations of stimuli is an additional consideration. Subjects can only spend a limited amount of time in the scanner, and in this chapter we will discuss how we balanced these factors.

3. **The cognitive task demanded from the subject during the experiment**

   Due to the length of the experiment, picking a cognitive task that is engaging

yet not too demanding is necessary. This chapter will describe the cognitive task asked of subjects during the experiment, and the design of a paradigm to ensure their attention.

4. **The structure of the experiment**

   Efficient procedures for dividing the experiment into parts and structuring those parts both help reduce overall experiment duration and improve subject comfort. This chapter will describe the process of designing the overall structure of the experiment with all of the mentioned elements being considered, and our technical implementation designed to address these concerns.

5. **The technology used to display the stimuli**

   The physical setup of the equipment used to perform the experiment is integral to designing the structure of the experiment and programming the software used to construct it. This chapter will enumerate the equipment used and detail how they were used to achieve the desired final result.

6. **The fMRI scanner image acquisition parameters**

   The fMRI scanner is often taken for granted in fMRI trials, with many studies using traditionally accepted settings for image acquisition. Our findings indicated that we needed more signal clarity from the scanner. This chapter will describe the steps taken to acquire special-purpose fMRI images for our experiment.

Another consideration parallel to the above is identifying the LOC and the VWFA in the brain. Since LOC and VWFA are functionally defined areas, anatomical markers in brain images are unreliable for determining their locations. Thus, a functional localizer [31, 57] must be used to find voxels associated with these areas. Functional localizers also require each of the above design elements to be realized separately.

More specifically, this chapter will discuss the design process for the stimuli in depth, including the basis it has on prior work, the algorithm developed for generating and presenting these stimuli to the subject during the experiment, the design of the experiment and software created to facilitate this design, an experiment that was performed as a preliminary analysis to iterate on the stimuli, the lessons learned from that experiment, three more experiments that were conducted based upon the findings from the initial experiment, and the changes made to those experiments and paradigms.

Section 2.3 will discuss the overall structure of the experiment and the equipment used during it. This section includes our contribution to image acquisition techniques for the image the visual cortex, specifically the lateral occipital complex in high resolution. Section 2.4 will discuss our stimulus design, motivated by the experimental goals and considerations for the rest of this work. This section includes the algorithms used to generate and select stimuli. Section 2.5 will describe the design of our experiment's overall structure and cognitive task, taking into account time considerations and feedback from our preliminary experiments. Section 2.6 will describe

the design and implementation of functional localizers we use in our experiments to identify what brain regions to target for analyses in the rest of this work. This section also includes our novel English VWFA localizer derived from the French-based work of Szwed et al. [57]. Section 2.7 will detail the design and implementation of our software that implements all of the experiment protocols. Section 2.8 will discuss the specific issues that came up during individual experiments and how we overcame them and what we learned from them. We will then discuss our contributions and findings in this chapter and how they can be used for future experiments and be applied to other use cases.

2.2 Related Work

The space of performing fMRI experiments is quite large and well-established. Here we will focus on studies we were informed by or have based our approaches on. Specifically, this section can be broken down into several categories. General-purpose works that helped inform our fMRI design, and special-purpose studies that have helped us target specific brain areas or derive stimuli.

2.2.1 Experiment design

Many of the key techniques of experiment design in functional magnetic resonance imaging have been outlined by the seminal work of Huettel et al. [24]. Our work

follows the guidelines set by them and Smith et al. [56] with respect to experiment design. Both works have established standards for how stimuli should be presented, in terms of timing, randomization, stimulus blocks and subject response expectations. Buxton et al. [5] have defined how the fMRI signal responds to stimuli over time, and our experiment design in regards to timing has been directly influenced by trying to account for these principles. While these works, and many others have presented the groundwork for experiment design in fMRI, each experiment is unique in terms of the constraints that have to be managed. For our studies, we needed to run exceptionally long experiments, which required both significant planning and some trial-and-error in addition to knowledge obtained from the previous works in the field.

One of the novel contributions of this chapter is the paradigm for imaging the visual cortex in high resolution. Yassa et al. [66] have established a technique for imaging the hippocampus at a higher resolution than normal. Typically, 3 Tesla scanners, which are widely used in fMRI studies, allow for voxels of size $3mm \times 3mm \times 3mm$. Yassa et al. use different parameters to narrow down the field of view of the scan, which lets them acquire for voxels of size $1.5mm \times 1.5mm \times 1.5mm$. However, the areas this study will image are larger than the hippocampus, to the point that they do not fit within the field of view of the hippocampal scan. Our approach involved careful positioning of the field of view and adjusting its parameters to fit all the regions we needed into the image.

There have been many cognitive tasks designed to ensure subjects pay attention

to the stimuli being presented for the prolonged duration of experiments. In some experiments, the task itself is the goal. However, in our experiment, the task is a distractor and the real goal is to have subjects attending to the visual stimuli. As such, we have sought a simple, non-intrusive task design. The one-back task, proposed originally by Kirchner [29] was used. The one-back task requires subjects to retain stimuli in short-term memory. While doing so for visual stimuli, they also engage higher visual areas to retain the images [39]. This task is very appropriate for our experiments as a result.

A key aspect of our imaging studies is narrowing down the window of voxels being looked at for the analysis. Kourtzi and Kanwisher [31], Kourtzi and Kanwisher [32], Kourtzi et al. [33], Grill-Spector et al. [19], Amedi et al. [1] have all demonstrated evidence of object shape representation in the LOC. These works have also provided a variety of approaches towards identifying voxels found within the LOC with functional localizers. In this work, we follow the design established by Kourtzi and Kanwisher [31].

Similarly to localizing the LOC, we also seek to identify voxels found within the visual word-form area. Szwed et al. [57] have described a robust method of localizing the VWFA, and we follow their approach. However, their work is with a French audience and using a French text corpus, and our work adapts their approach to an English audience.

All the programming in this section was done with Python, and the visualization

software was written with the OpenGL wrapper pyglet [23].

2.2.2 Neurophysiology Studies Motivating Stimulus Design

Our goal in this work is to find a parametrization for structural coding of object shape in the human LOC. This approach however, is motivated by work in monkeys. Since it is difficult if not impossible to perform neuron recording studies on humans, the motivation for the parametrization comes from work done in the field of monkey neurophysiology. Specifically, the inferotemporal cortex is considered the homolog of the LOC in monkeys [37]. Thus, understanding the IT is critical to deciphering the LOC. In this section we will discuss monkey work that has explored the IT and other relevant areas.

Evidence of the inferotemporal cortex as a visual area goes back to the 1950s, with work by Pribram and Barry [51] and Wilson [64] showing ablation of inferior temporal regions in monkey impairs visual function. Further work by Dubner and Zeki [17] has solidified the IT as an area that receives inputs from the primary visual area (V1) relayed by V2 and V4. More detailed ablation studies by Dean [15], Gross et al. [20] and Mishkin [40] have shown that ablating the IT specifically impairs visual discriminiation or recognition of objects. Tanaka et al. [58] have demonstrated that there is a specific coding for objects of differing structural parameters. This finding

is key to our work because it establishes the IT as an area that codes for structural properties of objects.

There has been more recent work in trying to exactly establish how the IT codes for object shape. Pasupathy and Connor [49] have demonstrated evidence for a parametric coding of object shapes in monkeys. They have also used this coding to reconstruct the stimuli being viewed using the neuronal spike information. This has given us motivation to try to describe a similar parametric framework for LOC voxels, treating them similarly to how Pasupathy et al. have treated neurons. In addition, their reconstruction approach has inspired the methods we will apply on Chapter 5. Brincat and Connor [4] have described a quantitative model for how the IT codes for straight and curved shape fragments that compose stimuli. Our approach builds on their work, as we will also use fragments to compose stimuli, and we will attempt to model the responses to those fragments in terms of their curvature. The work by Yamane et al. [65] has uncovered an IT coding for three-dimensional object shape and spatial configurations. While the limitations of fMRI mean that a parametrization as wide as Yamane et al.'s would be too ambitious for our study (as we shall discuss in this chapter), we have also integrated findings from their work in an attempt to address spatial variations in fragments. We will include fragments in different polar positions with respect to the fixation point in our model. Hung et al. [25] have provided a framework that demonstrates medial axis shape coding in IT. While finer parametrizations based on globular shape curvature (as per the previous studies dis-

cussed) would have been too complex to discern with fMRI, medial axis shapes are very well-defined and can be discretized much more effectively. As such, our work has also adopted a medial axis stimulus design.

2.3 Experiment Setup And Equipment

Many of the design choices in this chapter are based on limitations and capabilities of fMRI. In order to justify those choices, we need to discuss the physical setup of the experiment, the equipment used including the details of the fMRI scanner, the setup for displaying the stimuli to the subject, and the image acquisition parameters. This section will also include information regarding fMRI image acquisition principles and how those relate to design choices made in the experiment.

Scanning was conducted at the F. M. Kirby Center for Functional Brain Imaging located in the Kennedy Krieger Institute, Baltimore, MD. fMRI data were acquired using a 3-Tesla Philips Gyroscan MRI scanner equipped with a 32-channel SENSE head coil using both higher-order shims and parallel acceleration techniques. Functional images were collected using a T2*-weighted high-speed echo-planar single-shot pulse sequence. T2 weighting requires a longer time resolution (TR), which is the time period in which the slices in the image are excited and allowed to demagnetize. Neuron response times are in the order of milliseconds, whereas the Haemodynamic Response Function (HRF) that is measured by fMRI peaks at 4-6 seconds and sta-

bilizes at after 10 seconds. Due to the fact that the HRF is a well-defined, smooth, continuous function, increasing the time resolution does not make a significant difference in signal quality, unless stimuli are being presented very rapidly. While there are fMRI experiments with block designs that show many stimuli in quick succession, in fact our functional localizers do this, such experiments are generally designed to observe activity in an area and not to characterize said activity. Since our objective was to isolate and model the response to each stimulus, we did not use an experiment design with very fast stimulus presentations. As a result, we used a TR of 2000ms. The scans used a field of view of 96×96mm, flip angle of 70°, SENSE factor of 2, TE of 30 ms, and resolution of 3 mm isotropic for the initial experiment, then 1.5mm for subsequent experiments. Each EPI pulse sequence began with 4 dummy pulses that were not recorded in order to allow magnetization to reach steady-state.

In the original experiment, we had acquired full-brain images with a voxel resolution of $3mm \times 3mm \times 3mm = 27mm^3$. We also used a randomly-staggered stimulus presentation timing in order to prevent the subject from predicting the presentation time of the next stimulus. The staggering used was a 2/4/6/8 second timing between each presentation (with an average of 4), with 1.5 seconds of each presentation showing the stimulus, and the rest of the time a blank screen. The reason for the blanking in between stimuli is tied to our experiment design, which will be discussed in section 2.5. After conclusions drawn from the initial study (detailed further in Section 4.3), we realized that the stimulus timing could be set to a constant value because of the

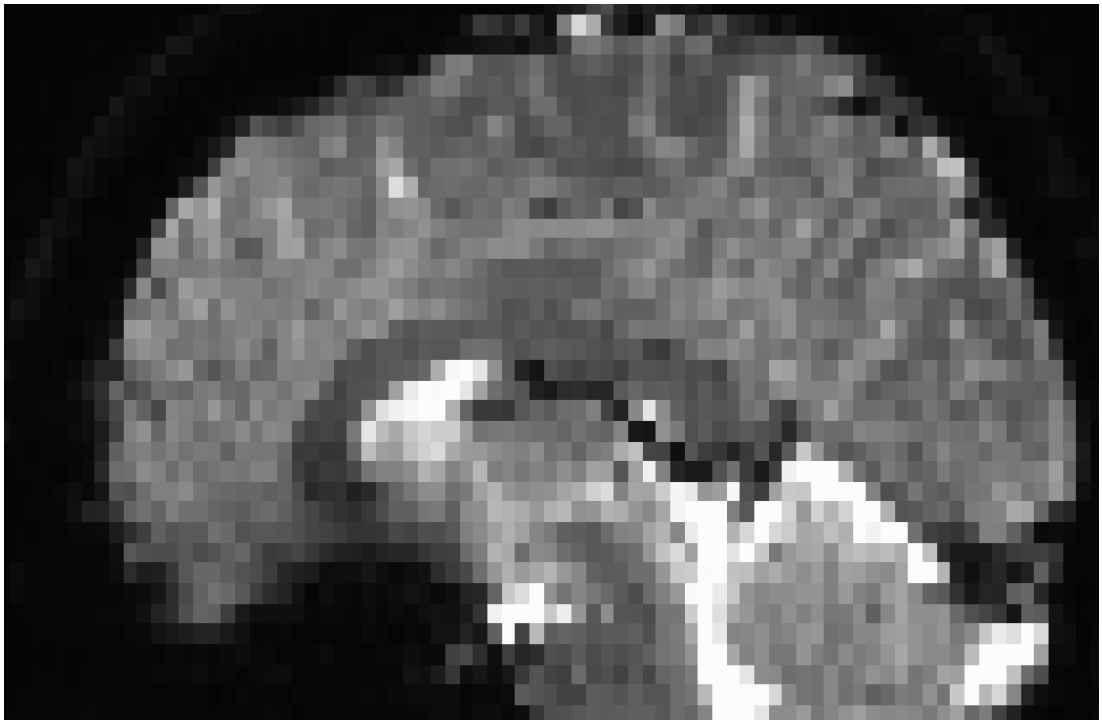
way the data is analyzed, so we set it to 4 seconds for every stimulus.

Results from the initial study (discussed in Section 4.3) showed that we needed more resolution from our images. Without changing to a higher-Tesla scanner, which causes more fatigue in subjects and sometimes incites minor discomfort like numbness in the extremities, our options within the same scanner were limited. Inspired by the technique Yassa et al. [66] use for imaging the hippocampus, we designed a novel imaging scheme. While the overall amount of information acquired by the scanner is fixed, one can reduce the field of view. The effect of this is that, instead of imaging the whole brain we can image part of it, and that would let us have smaller voxels, which in turn meant more information per voxel. For further scans, we used $1.5mm \times 1.5mm \times 1.5mm = 3.375mm^3$ voxels. However, this required precisely positioning the field of view to the areas we were interested in imaging. In this imaging scheme, 26 axial slices were first aligned to the AC-PC plane and then positioned to include the inferior-most portions of occipital cortex. The resulting volumes thus provided only partial acquisitions of the brain but were targeted to include visual areas V1-V4 and LOC. This is a non-trivial change to the hippocampal acquisition technique. Since the visual cortex is larger than the hippocampus, it requires careful consideration of imaging parameters and FOV alignment on a per subject basis to capture it in its entirety. While the hippocampal acquisition can be used without obliquely aligning the FOV, our approach only works when aligned with the axis of the visual cortex, which creates additional challenges in the registration step that will be further explored in Section

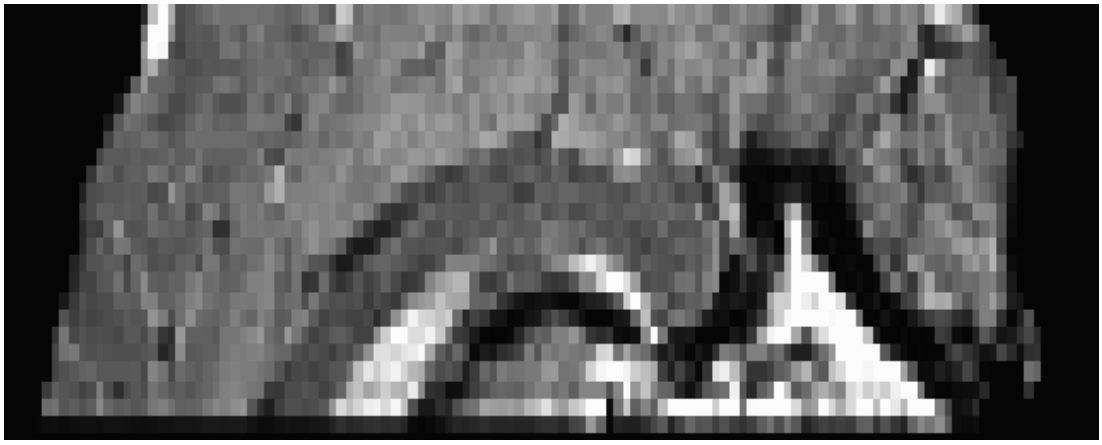
3.4. An image comparing a regular scan to our novel acquisition paradigm can be seen in Figure 2.1.

The visual stimuli were displayed using an Epson PowerLite 7600p projector with a custom zoom lens onto a screen mounted at the end of the magnet bore behind the participant's head. Participants viewed the screen by way of a mirror mounted to the head coil at a distance of about 72cm from the screen and 11.5cm from the eyes. Responses were recorded using a custom-built, fiber-optic push button box. The experiment was set up so that the subject would focus on a fixation point at the center of the screen at all times. Stimuli were displayed at exactly the size to fit in the fovea, since more neurons are dedicated to representing the fovea than any other part of the visual field. As a result, stimuli were sized to be 2 deg of visual field around the fixation point. Images were displayed via a Python image presentation software we built. In addition to displaying stimuli, the software also recorded button presses from the subject as demanded by the cognitive task. We additionally collected a high-resolution structural MPRAGE scan with a field of view of 240×240mm, flip angle of 9°, TR of 12 s, TE of 5.9 s, matrix size of 384×384, and 0.65 mm isotropic resolution. SENSE parallel imaging was used in two directions (2×1.5), yielding an overall volume of 231 slices. This scan was used to target our high-resolution-limited-FOV scans, and also to anatomically register the fMRI runs to statistical atlases.

Four neurologically healthy Johns Hopkins University graduate students (2 female)



(a) Image acquired with a 3 Tesla scanner using voxels of size $3mm \times 3mm \times 3mm$.



(b) Image acquired with a 3 Tesla scanner using voxels of size $1.5mm \times 1.5mm \times 1.5mm$.

Figure 2.1: Comparison of a regular fMRI image with voxels of size $3mm \times 3mm \times 3mm$ acquired with a 3 Tesla scanner to our paradigm that uses voxels of size $1.5mm \times 1.5mm \times 1.5mm$ acquired with the same scanner. The field of view is aligned along the visual cortex. Note the oblique angle of the high resolution image.

with normal or corrected-to-normal visual acuity participated in the study (one in the initial study, three in the finalized study). Participants were compensated at a rate of $25 per hour and provided written informed consent. All procedures were approved by the Johns Hopkins School of Medicine Institutional Review Board and were conducted in accordance with the principles outlined in the Declaration of Helsinki.

2.4 Stimulus Shape Generation

Pasupathy and Connor [49]'s work in monkeys shows that inferotemporal cortex codes for object shape in both letterlike medial axis stimuli and surfaces of globular shapes. However, such work, being based on neurophysiological recording experiments with monkeys, has the luxury of being able to present hundreds of stimuli in quick succession, which is not the case with fMRI. As a result of limitations described in section 2.3, we were limited by time in terms of how quickly we could display stimuli. Since fMRI experiments can only last so long before subjects get fatigued and lose focus, the minimum time between stimuli inherently caps how many stimuli we can possibly display.

Feedback from subjects and prior experience dictated that 90 to 120 minutes was the longest time we could feasibly run an fMRI experiment that requires attention, depending on how the experiment is structured. Depending on how conservative the timing between stimuli is set to be and how much rest time the subjects are allowed,

this means somewhere between 1000 and 1500 images could be presented to the subject. To counteract fMRI noise issues and other artifacts, several presentations of the same stimulus image were necessary. Originally, we were optimistic about being able to obtain good signals from our experiments, so we decided to use 200 unique stimuli presented five times each. While this later proved to be too ambitious, being limited to 200 stimuli still was a constraint on stimulus variety and design.

Due to the limitation on stimulus set size, we chose letter-like medial axis stimuli, as per Figure 2.2. Whereas globular stimuli would have to be parametrized by splines with free parameters, medial axis stimuli could simply be broken down into parts with a few discrete parameters which could be more easily extracted from fewer presentations. Medial axis stimuli we used were combinations of straight lines and 90-degree arcs, connecting at the tips. The stimuli were designed to have a center, and lines or arcs coming out of that center.

As mentioned before, the IT in monkeys is the homolog for LOC. Since previous work has shown that object shape, be it globular or medial axis, is coded for in the IT in terms of polar position, orientation and curvature, we wanted our parametrization to reflect this. Straight line fragments going out from the origin at four cardinal directions represent the polar positions at 0, 90, 180 and 270 degrees. Arc fragments going out from the origin at the four cardinal directions, due to being a quarter-arc, end at the polar positions of 45, 135, 215 and 305 degrees and represent those positions. For curvature, we had three values. Straight line fragments represent zero

curvature, arcs that start from the origin and go left represent negative curvature, and arcs that go right represent positive curvature. Finally, the directions that the outside of the arcs face represent the orientations of 45, 135, 215 and 305 degrees, and the directions each straight line faces represent the orientations of 0, 90, 180 and 270 degrees. This gives us a three-dimensional discrete parameter space.

The process used to generate the stimuli can be seen in Algorithms 1 and 2. Algorithm 1 takes as input the desired number of stimuli in the stimulus set, the "depth" of the stimuli (which denotes how many levels the farthest stimuli can be away from the center), and the desired size of a fragment. The last parameter will be addressed in Section 2.8. The first algorithm simply calls the second algorithm four times (once for each quadrant), checks if the generated stimulus contains at least two fragments, then repeats this process until enough stimuli are generated.

The second algorithm recursively generates fragments while checking for conflicts. It uses a global two-dimensional array *end_points* to keep track of locations visited in the 2D grid. Whenever a new fragment is placed, an entry is created on this grid at the coordinates of the end point of the fragment, and the value is set to the depth of the fragment. Whenever a new fragment is to be placed, the algorithm checks this global grid to see if the end point of the candidate fragment clashes with a pre-existing fragment. The algorithm works by choosing a random set of directions to explore based on the direction of the parent fragment. For each of these directions, it randomly considers adding zero, one or multiple of the following: a straight line, a left

curved arc or a right curved arc. It checks for conflicts for these candidate fragments, and if they are placeable, it places them and calls itself on the new position, reducing the amount of depth allowed for its child. At the end, once all children have been placed, it recursively returns a list of all child fragments. This list can then be parsed by the experiment presentation software in Section 2.7 and displayed to the subjects. This software was written with Python 3.2.

Algorithm 1: Stimulus set generation

```
Function generate_stimuli(num_stimuli, max_depth,part_size)
    valid_stimuli ← 0;
    stimuli ← [];
    while valid_stimuli < num_stimuli do
        global end_points[][] ← ∅;
        end_points[0][0] ← 1;
        fragments ← [];
        fragments+ = place_fragment(0, 0, 0, max_depth, part_size);
        fragments+ = place_fragment(0, 0, 90, max_depth, part_size);
        fragments+ = place_fragment(0, 0, 180, max_depth, part_size);
        fragments+ = place_fragment(0, 0, 270, max_depth, part_size);
        if max(end_points) > 1 ∧ size(fragments) > 1 then
            valid_stimuli ← valid_stimuli + 1;
            stimuli+ = [fragments];
    return stimuli;
```

Algorithm 2: Recursive stimulus fragment generation

Function `place_fragment`(x, y, $parent\_facing$,$depth\_left$,$part\_size$)**:**

- $depth\_left \leftarrow depth\_left - 1$;
- **if** $depth\_left == 0$ **then**
  - **return** [];
- $children \leftarrow []$;
- $directions \leftarrow choose\_random\_subset(parent\_facing, parent\_facing - 90, parent\_facing + 90)$;
- **foreach** *facing in directions* **do**
  - $valid\_fragments \leftarrow [end\_point]$;
  - $[straight\_x, straight\_y] \leftarrow [x + part\_size * cos(facing), y + part\_size * sin(facing)]$;
  - $[left\_x, left\_y] \leftarrow [x+part\_size*\sqrt{2}*cos(facing+45), y+part\_size*\sqrt{2}*sin(facing+45)]$;
  - $[right\_x, right\_y] \leftarrow [x+part\_size*\sqrt{2}*cos(facing-45), y+part\_size*\sqrt{2}*sin(facing-45)]$;
  - $choices \leftarrow [random\_bit(), random\_bit(), random\_bit()]$;
  - **if** $choices(0) \wedge end\_points[straight\_x, straight\_y] == \emptyset$ **then**
    - $end\_points[straight\_x, straight\_y] \leftarrow end\_points[x, y] + 1$;
    - $valid\_fragments \leftarrow valid\_fragments \| [straight\_line]$;
  - **if** $choices(1) \wedge end\_points[left\_x, left\_y] == \emptyset$ **then**
    - $end\_points[left\_x, left\_y] \leftarrow end\_points[x, y] + 1$;
    - $valid\_fragments \leftarrow valid\_fragments \| [left\_arc]$;
  - **if** $choices(2) \wedge end\_points[right\_x, right\_y] == \emptyset$ **then**
    - $end\_points[left\_x, left\_y] \leftarrow end\_points[x, y] + 1$;
    - $valid\_fragments \leftarrow valid\_fragments \| [right\_arc]$;
  - **foreach** *fragment in valid_fragments* **do**
    - **if** $fragment == straight\_line$ **then**
      - $children+ = place\_fragment(straight\_x, straight\_y, facing, depth\_left, part\_size)$;
    - **else if** $fragment == left\_arc$ **then**
      - $children+ = place\_fragment([left\_x, left\_y, facing + 90, depth\_left, part\_size)$;
    - **else if** $fragment == right\_arc$ **then**
      - $children+ = place\_fragment([right\_x, right\_y, facing - 90, depth\_left, part\_size)$;
- **return** *children*

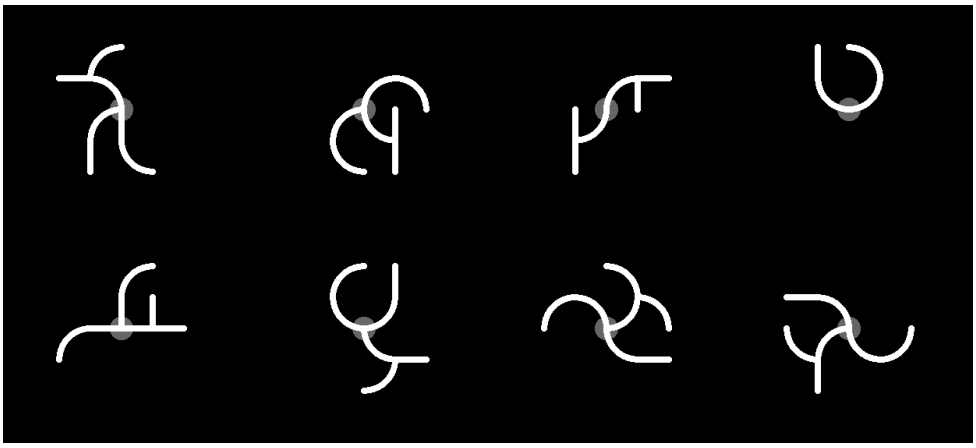


Figure 2.2: Samples from the original stimulus set. Fixation point is displayed as it was presented to the subjects, and it highlights the fact that many stimuli had quadrants with no fragments in them. The original stimulus set had stimuli that went at most two fragments deep from the fixation point.

The original stimulus design we used consisted of fragments coming out of the center at one of four cardinal directions, then as a second level, fragments coming out of the center also had a chance to have fragments coming out from their other edge, with the same parameters. In the original experiment we used stimuli that could go up to two levels deep, however subsequent analyses revealed that these stimuli were too complex to extract signal from the brain images they correlate with, so for future experiments we reduced stimulus complexity to one level deep.

Stimuli were parametrized by listing each fragment in the stimulus as a point in this 3D space in a vector whose length depended on the complexity of the stimulus. When we were using two-level-deep stimuli, there were more possibilities for the values each stimulus could represent in the 3D space, as the second level of fragments could end up in different polar positions or orientations, however that representation

ended up being too complex based on our results from the initial experiment. The end result using one-level-deep stimuli was a 8-by-3-by-8 set of possibilities in the position-curvature-orientation space. Given that we had rules against fragments starting at the same point to also end at the same point in order to prevent closed shapes, the space was not fully populated by our stimulus set.

Even with the restrictions, this meant more than a hundred unique stimuli were possible. We imposed a further set of rules on stimulus complexity, as we learned from our initial experiment that stimuli that had too many fragments did not drive good signal in the brain. As a result, we limited the stimuli to have at lest two and at most three fragments. We were able to control these values by changing the parameters of Algorithm 1. Due to timing limitations discussed in Section 2.5, we ended up deciding to use 60 stimuli, so we used 15 unique stimuli and their four cardinal rotations. Since the parametrization is not rotationally invariant, this filled our quota of 60 stimuli. Some stimuli, when rotated 180 degrees, were identical to themselves, but we kept both rotations in for consistency. We also included blank presentations to use as contrast in our statistical analyses.

2.5 Cognitive Task Design

The way an fMRI experiment is designed is critical to how useful the data extracted from it will be. Improper timing of stimulus presentations, tasks that do not

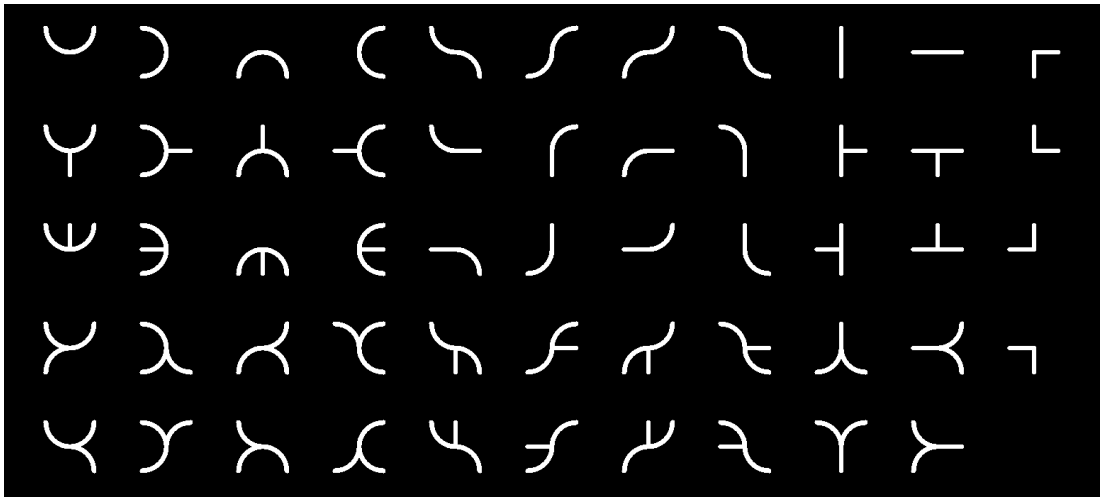


Figure 2.3: The entire stimulus set used for the final experiment.

direct the subject's attention properly, and many other factors can severely hamper the effectiveness of an experiment. In this section we will explain the choices we've made regarding stimulus presentation timing, stimulus size, experiment structuring and the cognitive task that the subjects were asked to perform. These choices will inform the requirements of the experiment presentation software we will describe in Section 2.7.

The goal of the experiment is to model the behavior of LOC, which is believed to perform a visual task but one beyond pure image processing. If LOC worked like earlier visual areas, for example V1, there would be no cognitive task demanded of the subject. But due to the fact that there is a higher level of processing at work in the LOC, the subjects need to attend to and visually examine the stimuli. Our objective was to design a task would make the subjects attend to the images. Conversely, we did not want the subject to pay too much attention to the stimuli and have their

minds wander, thus we wanted to obtain a balance between presenting the stimulus long enough for subjects to visualize them, but short enough for them to not think too much about them.

The task design we chose to fulfill these criteria is a "one-back matching task". An n-back task is a commonly used cognitive neuroscience experiment, and for our purposes we used $n = 1$. In this type of task, subjects are presented with a stream of stimuli, then at any point in time, a query stimulus is presented, then the subject is asked to determine whether the n^{th} previous stimulus was identical to the query stimulus. While this type of design is used in many experiments, it required some modifying for the experiment we needed to conduct.

Since we need subjects to attend to the structure of the images being presented, we wanted to eliminate purely V1-based strategies where the brain can easily do image-matching between two pictures viewed in rapid succession. Specifically, there are several strategies we needed to eliminate so that the subjects were forced to compare the shapes of the stimuli presented and not simply the appearances:

1. **Matching the overall luminance of the two stimuli**

2. **Matching the general structure of the two stimuli**

3. **Mentally overlaying the two stimuli to find small differences**

Traditionally, the last strategy is easily eliminated by moving the stimuli around the field of view while the subject is asked to focus on a fixed point. This allows

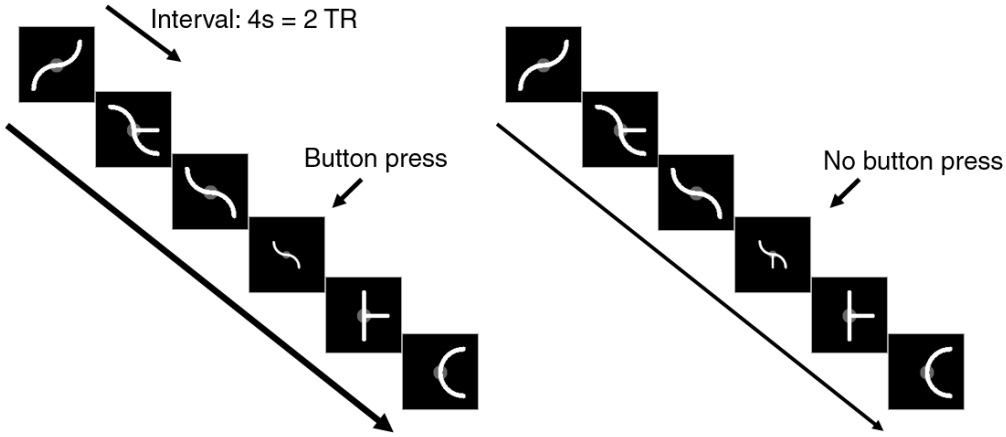
researchers to have stimuli appear in visual areas connected to different neurons, which means visual-matching strategies wouldn't work. However, since what we believe about the functionality of LOC is that it's not invariant to position, this means that stimulus positioning would be an additional factor to introduce to the experiment, which we wanted to avoid. Instead, we decided to keep the fixation point unchanged at all times and introduce variation in stimulus sizing, as IT, the monkey homolog for LOC, is invariant to size - even the position of object parts is represented in polar coordinates and thus don't change as the object changes size.

In addition to size changes, we used a scheme of displaying the stimuli for 1.5 seconds and then blanking the screen for the rest of the interval, which both helped prevent visual matching strategies and also prevents subjects from spending too much time thinking about the stimulus which can cause extraneous brain activity in higher level areas, causing distraction. Having the query stimulus differ in size also eliminates the problem of luminance-based strategies, since if the background is constant and only the stimulus changes size, the overall luminance of the image will be different regardless of whether the query stimulus is identical to the previous stimulus or not. Our query stimuli were 20% the size of regular stimuli, which meant they were still easily discernible yet visibly different. This requirement was why we designed the stimulus size as a parameter of Algorithm 2.

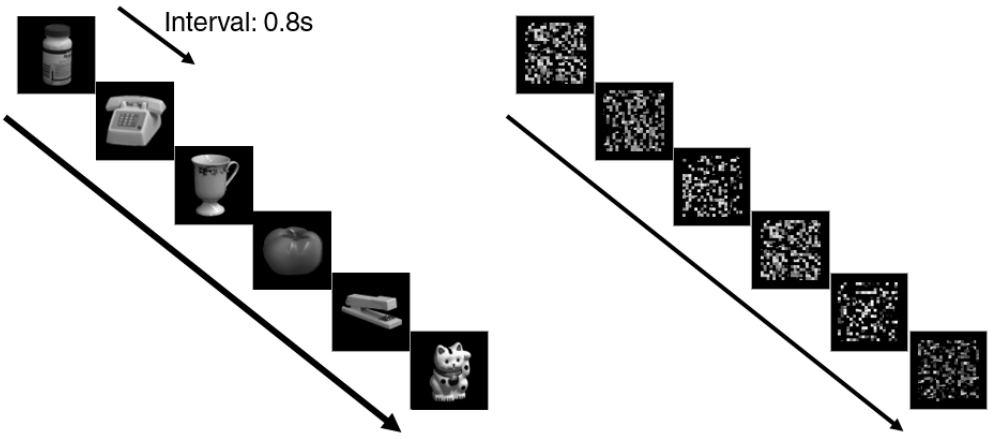
This only leaves the problem of subjects matching the general shape of the stimuli and not paying attention to the structural details, which we combated by having the

query stimuli vary only slightly. As described in section 2.4, our stimuli have 1-3 fragments per quadrant, and 4 quadrants per stimulus. Each fragment can be one of the following three shapes: straight line, left-curving-arc, and right-curving arc. As previously mentioned, each of those fragments are assigned a curvature value of 0, -1 and 1. The variation we introduce in the query stimuli that should be responded to as "different" is modifying a single fragment's curvature by 1 point, or adding a single fragment with a curvature that is 1 point off from one of the existing fragments. In practice, this was done by going into the data generated by Algorithm 1, and selecting a maximum-depth fragment, and changing its value, based on the availability of the *end_points* grid. If the fragment was a straight line fragment, it was changed to be an arc, and if it was an arc, it was changed to be a straight line. 10% of our stimuli were designed as query stimuli, and were randomly inserted into the experiment. Half of those queries were positive (query stimulus identical to previous), and the rest were negative (query stimulus different from previous).

Since the cognitive task of responding to queries is not the true factor being measured in the experiment, we wanted to eliminate the complexities of operating a button response box with multiple buttons. We used a single-button response box that sent its response over the serial port, and subjects were asked to press the button when the query stimulus was identical to before, and they were asked to not respond if the query was different. Our experimentation confirmed this approach to be a valid one for capturing the attention of the subjects, as over the four experiments, the



(a) The one-back task for the main experiment. On the left, we have a positive match, where the trial stimulus is a smaller version of the previous stimulus, and the subject is required to press the button. On the right, the trial stimulus is different by a single fragment, thus the subject is not required to press the button.



(b) The viewing task for the LOC localizer. The subject is not required to press a button and instead they are to passively view the images that appear in quick succession. On the left, we have the intact images. On the right, we have the scrambled images.

Figure 2.4: Demonstration of the one-back task for main experiment and the viewing task of the localizer runs. Stimuli are presented at a time interval, and the subject is required to attend to them.

amount of total mistakes made was 3. Subjects were not told about the true nature of the experiment (that the queries were irrelevant and it was a scheme designed to get them to pay attention to the shapes), as we wanted them to pay attention to the task and be ready for a query stimulus at any time. The performance of each subject can be seen in Table 2.1.

Subject	TP Presses	FP Presses
1	48/50	0
2	49/50	0
3	50/50	1

Table 2.1: The performance of each subject on the one-back task. True Positive and False Positive presses from each subject are denoted.

As mentioned in section 2.4, the HRF is a significant constraint on stimulus timing. Given that the more subjects get fatigued in scanners, the more their performance decreases and the more they move (which results in registration errors), keeping the experiment as short as possible was our goal. Our initial experiment included 1000 stimulus presentations (200 unique, 5 repetitions) at an average of 4 seconds per stimulus. Including the 100 queries (10% of total presentations), the total number of presentations goes up to 1100. That meant and experiment that lasts at least 4400 seconds (approximately an hour and 15 minutes) not including the time between runs, dummy pulses, localizers and structural scans. The design of the first experiment was to have approximately 4 and a half minute runs, with a 30 second break in between. That meant 68 presentations per run (including 4 dummy acquisitions that take 8 seconds), totaling at 15 runs. With 15 runs at 5 minutes each, and 15 minutes for

the localizer (see section 2.6.1) and the structural scan, that meant an experiment that would last over 90 minutes. We considered this to be an acceptable length for an experiment.

As we will discuss in Chapter 4, the initial experiment was not successful in acquiring good enough data to perform statistical analyses, so we decided to modify our experiment design significantly. We decided to include a visual word-form area (VWFA) localizer as described in section 2.6.2. We moved to a design with 60 stimuli presented 20 times, which meant 1200 stimulus presentations. Adding 120 query presentations as 10% of total stimuli brought the number to 1320, and we added 20 blank presentations to use as contrast in our statistical analysis. That brought up the total number of presentations to 1340. With 4 seconds per stimulus, the pure presentation time was 5360 seconds, or 1 hour and 30 minutes. If we were to divide this experiment up into runs the same way as the first experiment, it would take 20 runs (100 minutes), and adding 15 minutes for the structural scan and LOC localizer, then 7 more minutes for the VWFA localizer, the experiment would take well over 2 hours. Including approximately 15 minutes of setup time, the experiment would last over 2 hours and 30 minutes. This was undesirable for fatigue and cost reasons, since scanner time was bookable in 30 minute increments, which would lead to us having to book 3 hour scans. Feedback from our initial subject indicated that we could have longer runs with less breaks in between, and the subject even expressed preference for a design like this. As a result, we decided to modify our runs so that

they last 9 minutes each, with 10 seconds of rest in between. This meant 10 runs with 134 presentations each, which meant the overall experiment would be 1 hour and 45 minutes long. With 15 minutes of setup time, the experiment's total length would be two hours long, and this was the ceiling for how long we wanted the experiments to take, thus we decided this paradigm would be the one to be used for our three final experiments.

2.6 Localizer Experiment Designs

Whole-brain images are often not very useful for determining a parametrization for the function of highly specific areas. While whole brain imaging is used in works where a model is being built based on a general percept without a specific brain area targeted, works that try to determine the functionality of a single area need to specifically target those areas. While certain areas, especially earlier areas like V1 can be defined anatomically, purely anatomical markers aren't generally the sole basis upon which fMRI analyses depend. Even for clearly defined areas, a data-driven approach is more reliable on individual brains and helps verify structural assumptions made by researchers.

In the case of this work, we are looking at the LOC, which is an area that is defined functionally, thus a functional localizer is necessary to isolate voxels that belong to this area. LOC localizers are well-documented in literature, and involve

contrasting the brain signals induced by images of objects against signals induced by pixel-scrambled versions of those images. In order to verify that our parametrization of stimuli is represented uniquely in LOC, we also decided to localize voxels belonging to the VWFA and to show that the parametrization doesn't work on those voxels. VWFA codes for words and letterlike objects, which is what our stimuli are based on, but instead of parametrizing such shapes in a structural breakdown, it works based on identifying shapes as a whole. Showing that the same shape parametrization works in LOC but not in VWFA would increase the strength of our argument, hence our decision to localize both areas.

In this section we will discuss the implementation of both localizers using the same imaging hardware and setup as discussed in section 2.3. The requirements of these experiments will inform the design of our software tool, discussed in Section 2.7.

2.6.1 LOC Localizer Design

In order to be able to identify voxels in the LOC, we used a localizer approach established by Kourtzi and Kanwisher [31]. The structure of our experiment was mostly similar to theirs, with the main difference being the images used and the scrambling process. This section will consist of highlighting the relevant parts of Kourtzi and Kanwisher [31] and Grill-Spector et al. [19] and our changes to their paradigm.

The LOC's functionality is known to be related to object images, and the localizer

exploits that knowledge. This area of the brain responds to images of objects, but images containing the same pixel content rearranged spatially so that no discernible object is present does not elicit a response. The localizer is computed by contrasting images of the brain while the subject is looking at objects, against images of block-scrambled images of the same objects. We used the Columbia University COIL-100 image dataset by Nene et al. [46] for pictures of small household objects with no background. 20 object images were converted to grayscale, then scaled to 160x160 pixels after cropping them to a similar ratio of image-size-to-object-size as in [19]. The images were divided into a 20x20 grid, and each block in the grid was randomly relocated to another position using MATLAB. We ensured that no block would be in the same location as it was before, and also no two blocks who were originally adjacent were adjacent in the final image.

Using the Python software described in section 2.5, we displayed localizer stimuli to the subjects. As per the design in Kourtzi and Kanwisher [31], we used a block design where images of objects were shown in quick succession, followed by scrambled images in quick succession. More specifically, images of household objects were shown with a 0.2 second presentation time followed by a 0.6 second blank time in 16-second blocks. Stimuli were presented in a random order within that block. Following the intact-image block, a scrambled-image block of 16 seconds was displayed immediately after. The third block was a 16-second-long blank. These blocks were repeated, with the scrambled and intact blocks switching places every repetition, for a total of 8

times each. With each group of blocks taking 48 seconds, 8 groups, and startup time, the localizer runs took about 7 minutes.

2.6.2 VWFA Localizer Design

In order to determine that our model of object structure is unique to the LOC and doesn't work in other brain areas, we wanted to investigate other areas that could similarly code for images like ours. It is suspected that the VWFA codes for words and letters [57] (further discussion in Section 2.2), thus we decided to use a VWFA localizer as a sanity check for our models. Our localizer is based on the work of [57] and our LOC localizer described in section 2.6.1. This section will discuss the changes made to the localizer paradigm based on previous work and the implementation of the localizer.

Prior work in VWFA localization used images of words, generated from letters that have 45% of their luminance removed by deleting unsalient segments of them. Two approaches have been used by Szwed et al. [57]: removing the midsegments of letters, or removing the vertices. We decided that removing the midsegments was more appropriate for our ideas, as the edges provided sharp curvature changes which we thought would be more appropriate as a comparison point for the functionality of LOC. Szwed et al. used a French corpus for their approach. For our purposes, we needed to use English words instead. Following their approach, words were selected from a database of contemporary American English [62]. Neutral nouns with 6-8

letter length that had an occurrence rate of over a million were considered. The final selection of words can be seen in Table 2.2.

PEOPLE	SCHOOL	SYSTEM	NUMBER	MOTHER	WINDOW
FRIEND	MEMBER	MINUTE	HEALTH	MOMENT	WORKER
POLICY	MARKET	EFFECT	RESULT	COURSE	LEADER
MATTER	STREET	CHURCH	LETTER	THEORY	ANIMAL

Table 2.2: The set of nouns used for the VWFA localizer. 6-letter neutral English nouns with an occurence rate of over a million in the database of contemporary American English [62]

Images of these words were created using the font provided by Szwed et al. [57]. Those images were then loaded into MATLAB. We developed a bounding box based approach to scramble the stimuli. Stimuli were to be scrambled so that no fragment was in its original place, and no fragment was touching any other fragment. This was achieved by segmenting fragments out of the image with using connected components, defining bounding boxes for each fragment, randomly jittering fragment positions, then shuffling fragment positions until a non-conflicting arrangement is achieved (conflicts included fragments that were too close to each other, which was resolved by setting bounding boxes 5 pixels larger than the actual shape in each direction).

Similar to the LOC localizer, the Python software described in section 2.7 was used to display VWFA localizer stimuli to the subjects. Combining the experiment design of the LOC localizer and in Szwed et al. [57] As per the LOC localizer, we used a block design where intact words were shown in quick succession, followed by

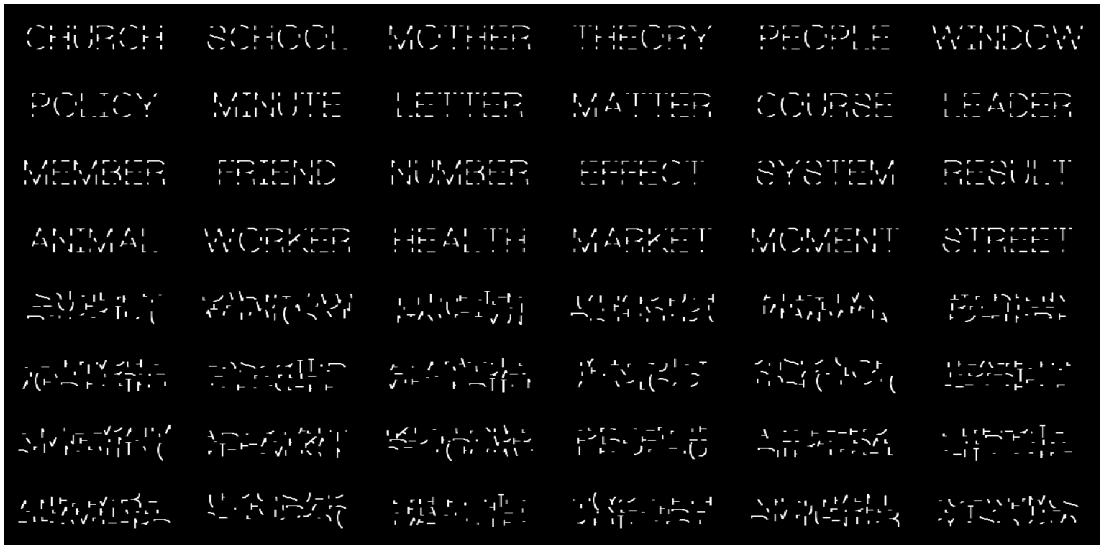


Figure 2.5: The entire stimulus set used for the VWFA localizer. Note that inside the scanner these stimuli were legible, which does not come across as well in this format. Subjects were asked post-experiment if they were able to read all the words, and they responded positively.

scrambled words in quick succession. More specifically, the intact words were shown with a 0.4 second presentation time followed by a 0.1 second blank time in 12-second blocks. Stimuli were presented in a random order within that block. Following the intact-word block, a scrambled-word block of 12 seconds was displayed immediately after. The third block was a 12-second-long blank. These blocks were repeated, with the scrambled and intact blocks switching places every repetition, for a total of 8 times each. With each group of blocks taking 36 seconds, 8 groups, and startup time, the VWFA localizer runs took about 5 minutes.

2.7 Experiment Software

So far, in this chapter, we have described several different experiment paradigms. We will build a software tool that will facilitate our experiments, informed by these requirements. Specifically, here are the design considerations that our software tool will need to address:

1. **Interfacing with the experiment hardware.**

   This includes the fMRI scanner, response box and projector

2. **Precision**

   The software must be able present both fast-paced and slower stimuli over short and long runs without time drift.

3. **Flexibility**

   The software must be able to accommodate the variety of experiment designs we have proposed throughout this chapter.

We will now discuss these considerations and the steps we took in our design to address them. We will also describe the overall structure of the experiment, and consequently the flow of the software. The Python program was built using pyglet [23], an OpenGL wrapper.

The fMRI scanner, subject response button and the projector were all controlled through a Cedrus RB-830 response box as seen in Figure 2.6. This device has several

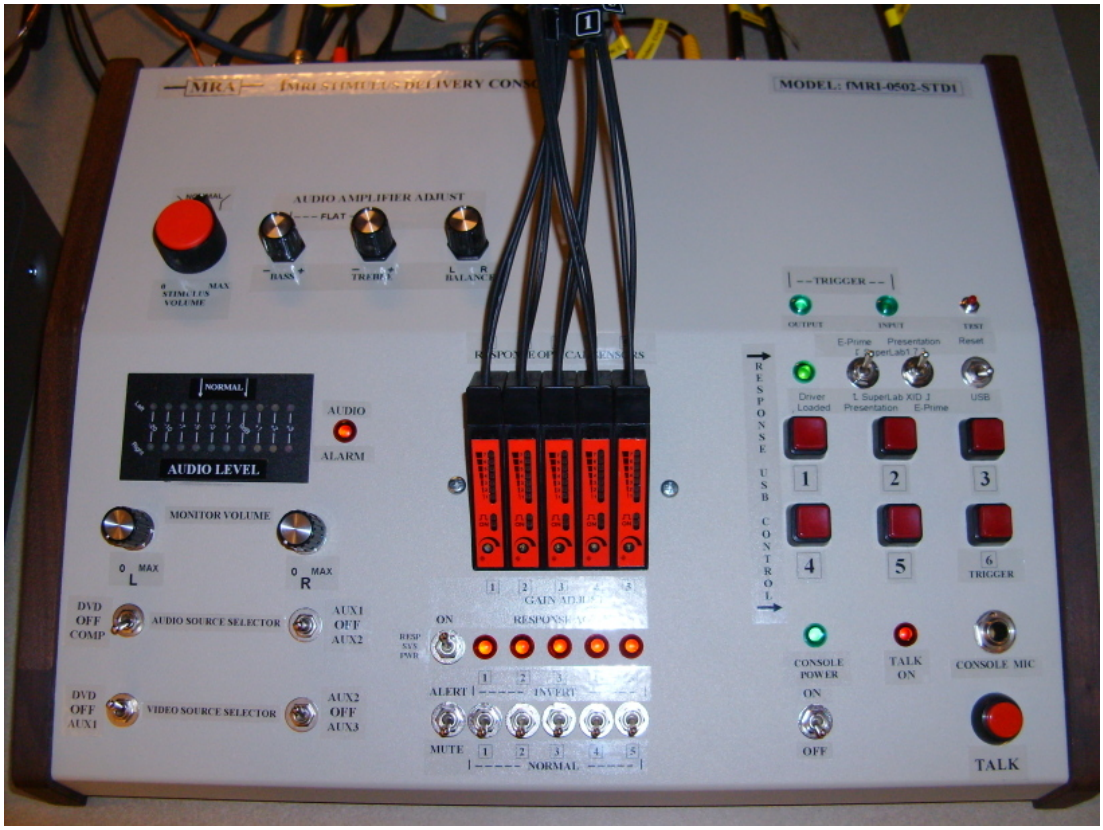


Figure 2.6: The Cedrus RB-830 response box used in our experiments. This device sends all fMRI-related signals over the serial port.

modes of communication through the serial port. We used "Presentation" mode. In this mode, the scanner sends a pulse to the response box, which is sent over the serial port as the single ASCII character '6'. This also light up the button 6 on the box, which can be used to verify synchronization with the software. Our subject response button was a single-button device, and whenever the subject pressed the button, it sent a single ASCII character '1' through the serial port. This immediately creates a design concern. Polling the serial port too often would slow the software, and potentially block. Polling it too rarely would mean the possibility of missing the scanner start pulse on time, which would cause desynchronization. Interfacing with the scanner to synchronize runs and receive button input was done with PySerial. While PySerial has recently added supprot for asynchronous I/O, it is considered an unstable experimental features, thus we did not rely on it. Instead, we used pyglet's handle-based update loop. Pyglet programs run based on an event dispatcher. This dispatcher can be configured to run handle functions based on either a trigger, or a schedule. We solved this problem by introducing a state machine and a variable update rate based on the state of the program. When polling for the scanner start pulse, the software will update at a half millisecond rate, otherwise it will update at a 1-millisecond rate. This gives us high precision when we need it, and performance when we don't need as much precision. This addresses both the concern of interfacting with the hardware and the concern of precision.

Displaying of the stimuli is handled by a combination of regular updates and a

schedule. The presentation software receives the pulse from the fMRI machine to start an experiment run, waits 8 seconds for the 4 dummy pulses, then starts displaying stimuli with the predetermined presentation times. Using test runs on the dummy scanner (see Section 2.8), we had designed the software such that whenever a stimulus is displayed, it would schedule the display of the next stimulus. We discovered that over hundreds of presentations throughout two hours, this would accumulate latency and end up several seconds off by the end of the scan. This was unacceptable to us. Instead, we used the time between the sync pulse and actual scans afforded by the dummy pulses to pre-schedule every single presentation. We calculated the timing of every stimulus, and set up a handler to display, hide and switch out each stimulus ahead of the actual presentations. Through several timing tests on the dummy scanner (that each lasted two hours), we were able to achieve millisecond-accuracy with this approach. The design of this state machine can be seen in Figure 2.7.

During a functional run, while the software is in the stimulus displaying state, we use pyglet's OpenGL drawing primitives and the output of our stimulus generation algorithm described in Section 2.4 to dynamically generate image stimuli. The advantage of dynamically drawing the stimuli is that it is faster to use OpenGL routines to draw polygons based on data already loaded in memory, than it is to load images from disk, then rasterize them as bit maps on a pixel-per-pixel basis. This also gives us the ability to generate new stimuli and immediately display them on the fly. This also makes the software have a small space footprint.

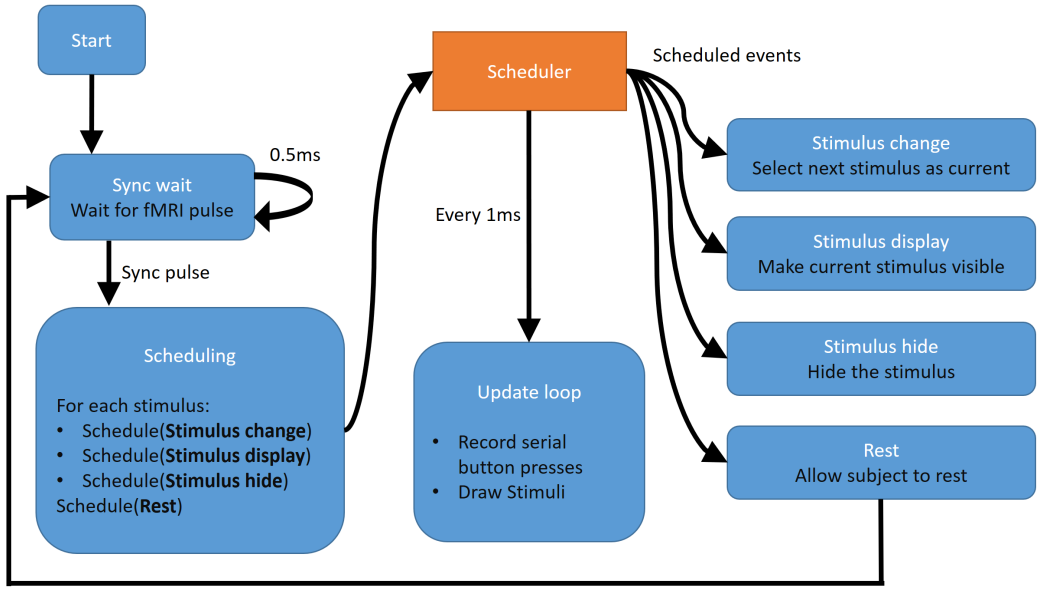


Figure 2.7: The state machine used by the software. Initially, the program runs a loop as it waits for the scanner pulse. Then, when the pulse is received, it schedules all stimulus-related activities, and gives the scheduler control. The scheduler ensures a regular update loop, in addition to calling stimulus-related event handles when needed. When the run is over, the subject is given a brief period to rest, after which the software starts waiting for a pulse again.

Our design of the state machine is inherent flexibility. This paradigm can be used to implement both the functional run, and also the localizers. By changing the set of stimuli used to either the medial axis stimuli we generated in Section 2.4, the LOC localizer stimuli we generated in Section 2.6.1 or the VWFA localizer stimuli we generated in Section 2.6.2, and setting the timing values as described in each respective section, our program can perform either task with ease. This addresses the flexibility concern.

While this state machine can handle individual functional runs or localizer runs, we also needed the software to handle the entire experiment. As such, we used a larger scale state machine based on the experiment designs we have discussed throughout this chapter. Figure 2.8 demonstrates the overall flow of the software. Each localizer or functional block within this figure represents an instance of Figure 2.7.

This software tool we have designed is flexible, precise, cross-platform, efficient and portable. Our tool can easily be extended to other types of experiments and even non-visual stimuli. The design of this software is motivated by all the factors described in this chapter, which are all building towards the end goal of developing a computational model for the activity of the LOC. These presentation approaches are intertwined with the design of the stimuli, which will be the most important facet of Chapters 4 and 5.

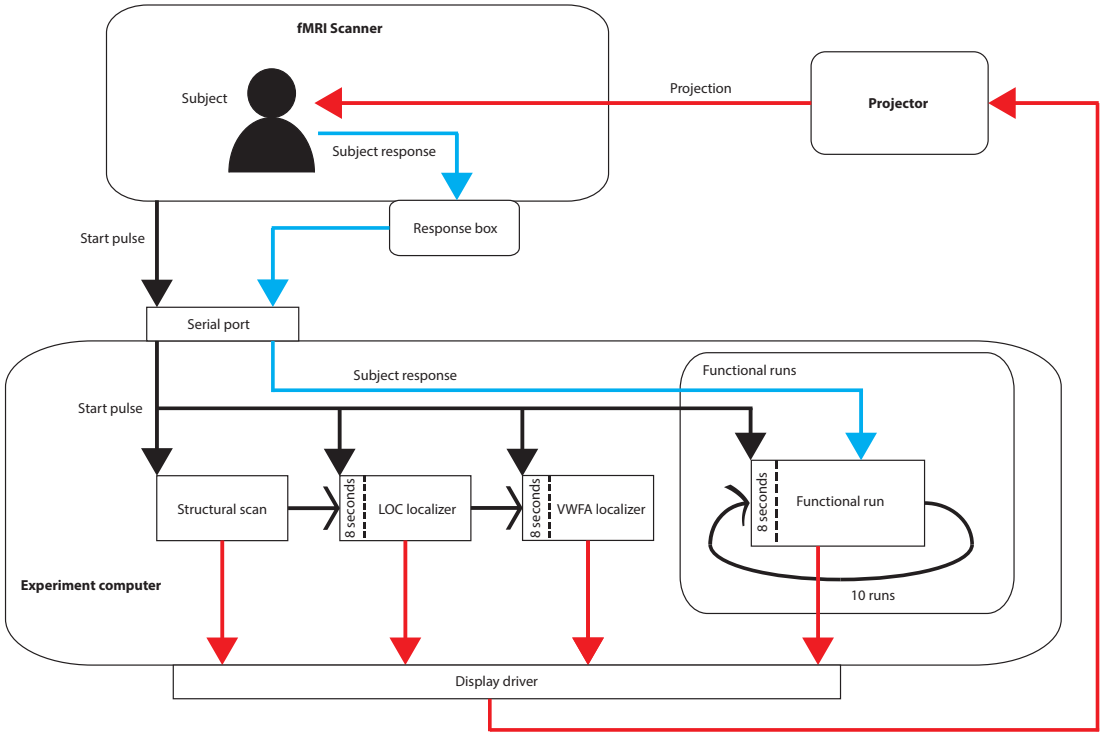


Figure 2.8: Flow of the experiment. Black lines indicate the signal's start pulses, which trigger each part of the run. Red lines are the stimuli being displayed through the projector. Blue lines are the subject's responses, which are a button press depending on a query stimulus presented.

2.8 Complications And Configurations Of Experiments

While the experiment design is quite concrete in theory, certain aspects of the experiments needed a significant numberof offline tests without subjects and on-the-fly adjustments to certain parameters. In this section we will discuss the parameters of the experiments that weren't discussed in previous sections, the dry runs without subjects, and any other configurations or changes to the software made during the experiments, either proactively, or in response to issues that arose during the process.

The Kennnedy-Krieger Institute has a dummy scanner with a button box and projector where the experiment setup was tested before any live runs were performed. This dummy scanner is a replica of the real scanning environment including projection dimensions and response box, however the scanner itself does not work or make any noise. Instead, one can simulate the scanner's behavior by pressing the button '6' on the response box. While the code was originally displaying images in 1920x1080 pixel resolution, test runs on this setup revealed that the projector had issues with resolutions above 1024x768, so the experiment software was reconfigured to output images at that resolution. Considering we wanted our stimuli to fit within 2 deg of visual angle, the distance of the scanner from the subject, and the depth of the stimulus set, the part size of fragments (the parameter in Algorithm 1) was set to 34.13 pixels. Consequently, the "trial" stimuli discussed in Section 2.5 were of the

size 6.826 pixels.

Another issue discovered during the course of the dry runs was that if there was an exception in the code or a situation that required the experiment to be suspended, the software could not be started at an arbitrary point in the run. This was rectified by adding to the software the ability to signify a starting point through the command line.

The first attempt at our initial experiment failed due to technical reasons beyond our control. It was during this experiment that we discovered the stimulus scheduling issue mentioned in Section 2.3. It would have been possible to recover the data by retroactively computing the timing delay and aligning the brain images temporally accordingly, however that ended up not being necessary. Due to inclement weather conditions the day before the scan, air humidity in the scanning area had changed significantly, throwing the magnetic calibration of the scanner off. As a result, all scans performed on that day had checkerboard artifacts and were effectively unusable for any analysis. It still served as a useful trial run for the setup and technical aspects of the experiment. Another experiment was run shortly after that, which is the experiment that is referred to as the "initial experiment" throughout this work.

After concluding from the analysis of the first experiment (as explained in section 2.3 that full-brain images with $27mm^3$ voxels weren't sufficient in their resolution, we decided to use the switch to the partial-brain images with $3.375mm^3$ voxels. We ran a trial study with no experiment just to test out the feasibility of thos image

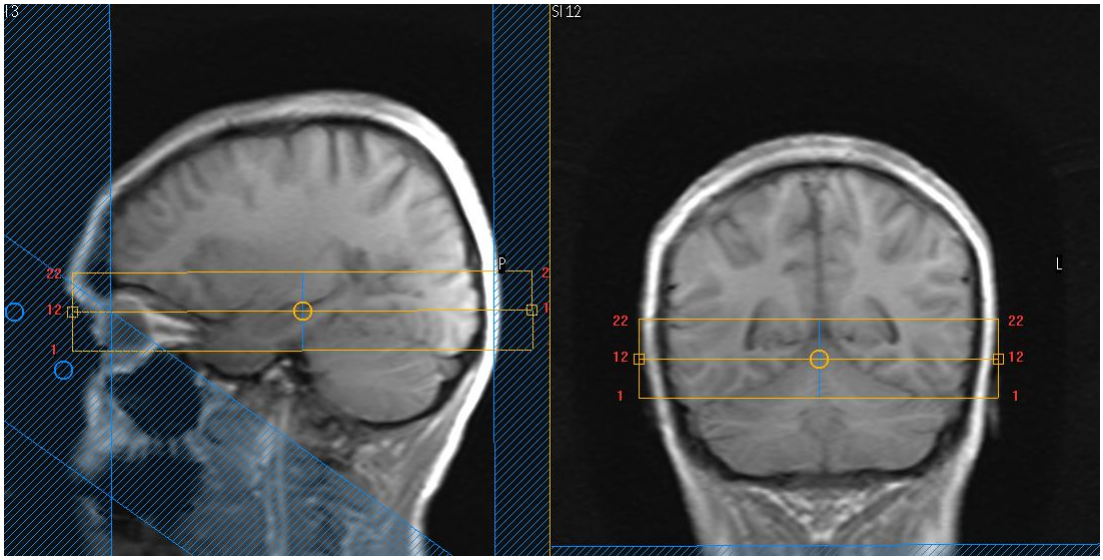


Figure 2.9: Image of the acquisition FOV. The yellow box indicates the limits of the FOV. In this instance the angle of the visual cortex wasn't very oblique, but we had other scans where the field was up to -45 off degrees from the X-axis of the left image. The blue areas were used as virtual fixtures to constrain the FOV placement and weren't used for purposes of the analyses.

acquisition method. This method was originally used to image the hippocampus by Yassa et al. [66], and we weren't certain if it would be able to encapsulate the visual areas we needed to image. We ran the study and were able to visually determine that the field of view was large enough to contain most visual areas, especially ones we were interested in, given some manipulation.

The next experiment was the first after the changes to the paradigm, however we encountered another potential problem this time. While the narrow field of view was large enough to image the visual areas of the brain sufficiently, the head of the participant who we ran the pilot study for the new imaging technique was significantly smaller than the subject that participated in this experiment. We were eventually able to manipulate the positioning of the scanning window to include all relevant

areas. The VWFA localizer was also included in our experiment design after this particular study, thus this subject does not have any data regarding that paradigm.

The next two experiments went without any issues and we were able to acquire all the data we needed successfully.

2.9 Conclusions

A key and often underestimated aspect of fMRI experiments is the experiment design. The quality of the experiment directly affects the quality of the data. In this chapter we have tackled problems related to designing our fMRI experiment. The first challenge we faced was determining fMRI scanner parameters and paradigms to get the highest possible image quality and the most data out of a two-hour window. We achieved this by running an initial experiment, then using information from that. Using this initial experiment we determined the best length for an individual run within a scan, and how many images to acquire during that window. We also implemented a novel imaging paradigm for acquiring high-resolution images of the visual cortex based on prior hippocampal imaging techniques. We built a Python software tool to interface with the scanner and deliver stimuli for the subjects to view. There were unforeseen complications during the experiments, like scanner failures and challenges with implementing our scanning paradigm, but we were able to overcome them.

The second challenge tackled within this chapter was the design of the stimuli. No

prior studies have demonstrated a parametrization of the responses of the human LOC to visual stimuli. With no prior human studies to rely on, we designed a stimulus set to achieve this goal. Inspired by neuron studies on monkeys, we used medial axis line fragments to compose letter-like shapes. We developed an algorithm for generating sets of stimuli based on the monkey work. Using the results of that algorithm we performed an initial study to test the viability of our stimuli, then improved our stimulus set based on findings from the initial study.

The next challenge to be addressed was the means by which the subjects' attention was to be directed and maintained. Since the experiments are nearly two hours long, keeping subjects attending is non-trivial. We have designed a non-intrusive one-back task to direct the attention of subjects as they attend to visual stimuli. The task itself was used as a distraction to keep subjects paying attention to every stimulus.

The last experiment design issue in this chapter was using functional localizers. In order to derive a parametrization for activity in the LOC, we needed to identify voxels in that region. For this, we integrated existing localizing techniques into our experiment design. Similarly, as a way to demonstrate the exclusivity of our parametrization to the LOC, we used a VWFA localizer. Part of the design of the VWFA localizer involved adapting the original experiment that was designed with French letters to English.

Finally, we have created an experiment paradigm encompassing all these aspects, and built a software suite to implement the paradigm. The design of this suite takes

into account all of the challenges we have faced throughout this chapter.

The specific contributions of this chapter are as follows:

1. A stimulus parametrization composed of medial axis fragments, designed to model lateral occipital complex activity.

2. An algorithm to generate stimuli based on the parametrization.

3. An experiment structure for delivering those stimuli to subjects and maintaining their attention.

4. An efficient and flexible software for conducting fMRI experiments based on our design principles.

5. A high-resolution fMRI imaging paradigm to view the visual cortex with 80 times the per-voxel resolution.

6. An English adaptation of the the visual word-form area functional localizer by Szwed et al. [57].

Using our parametrization, we will show evidence for structural coding in the LOC by developing a computational in further chapters. This stimulus set and the algorithm that generates it can be used in future studies to derive more information about the functionality of the LOC. The same stimuli can be used in monkey studies to show a correlation between coding in the inferotemporal cortex and human LOC. Future studies can also build upon our generation process to seek even finer or broader

parametrizations. Our experiment structure can be utilized by others as well, as it is not unique to the LOC. Any visual stimulus study on higher visual areas can employ the same paradigm to deliver a large amount of stimuli in a two-hour period. Our design is flexible enough to implement modules for our software for other types of experiments easily as well. Our high-resolution imaging of the visual cortex can be used in the future by other studies that seek to map out the functionality of visual areas. Most studies in the visual cortex have used scans that are lower resolution than our paradigm allows, and this technique provide an improved look at areas that have already been explored. As scanning technology improves, the base resolution of scanners will increase as well, which means more complex parametrizations of the LOC that build upon our stimuli can be achieved. Finally, while there are other localizers for the VWFA, the approach proposed by Szwed et al. [57] and refined by us is very simple and robust, and can be used by future studies with ease.

Chapter 3

Preprocessing of fMRI Data

In Chapter 2 we created an fMRI experiment paradigm in order to determine the location of the LOC and try to develop a model for how it represents object structure. The next logical step in this approach is analyzing the data. However, before being able to conduct any modeling of the activity of the LOC, the first step is to pre-process the fMRI data. This step involves taking the raw data output of the fMRI machine and turning it into information viable for statistical processing. The processing for each participant was done independently of other participants. For the purposes of this chapter, the steps taken will be described for a single participant. Unless otherwise noted, the process is identical for each subject.

3.1 Overview

There are many challenges that need to be addressed before fMRI data becomes interpretable. The raw file output from the scanners is in a sequential form, which means it does not carry any spatial or temporal connotations. Performing analyses on the data in this format would be infeasible. The data outputby the scanner includes header files, which include information on the parameters of the scanner during acquisition. The header files also include information on how each image in the scan was acquired, so these files can be used to transform the data into a format that is more suited for processing that takes into account spatiotemporal factors. However, simply transforming the image into a voxel format doesn't solve all challenges involved with fMRI preprocessing. As described in Chapter 2, our experiments take nearly two hours. Within this duration, subjects inevitably move their heads. In addition to subject movement, the superconducting magnet of the fMRI machine has its magnetic field drift, which affects the signal over the course of time [24]. These factors, coupled with baseline fMRI noise, inconsistency in the brain's response to individual stimuli, and other fMRI properties mean that several additional steps are necessary even for basic fMRI experiments. Typically, the steps to address these issues would include registering the brain images to each other, then correcting for factors like drift and noise via fitting a linear model. Additionally, the Hemodynamic Response Function, which dictates how fMRI values change on the onset of a stimulus, needs to be modeled together with stimulus presentation timings

to extract out activation values for individual stimuli. The goal of this chapter is to detail the approaches taken to prepare raw fMRI data for analysis in following chapters.

As mentioned in Section 2.3, we use a novel fMRI acquisition paradigm that results in partial-brain images with unconventional positioning. This makes registration more complicated than usual. This chapter will discuss the techniques used for cros-modally registering oblique, narrow-FOV fMRI images to whole-brain MRI images. Section 3.3 will discuss the format of the data as produced by the scanner, and how that data is transformed into voxel coordinates. Section 3.4 details our contributions in the approach used to register images acquired with our novel visual cortex imaging paradigm (as described in Section 2.3) to a baseline.

Even with a partial-brain acquisition like our approach, fMRI analyses that target specific brain areas don't use the entire set of voxels that have been acquired from the scanner, as there are a lot of voxels outside the target area in these images. Such extraneous voxels do not contribute meaningful information and confound legitimate analyses, thus it is common to use functional localizers to narrow down the set of voxels to ones that are within the targeted brain area. In Section 3.5 we describe the approach we use to model the activity of the areas we are looking to analyze in this work with localizers. We also describe how we use these models to eliminate noise, drift and other factors. In Section 3.6 we discuss how we use the data prepared throughout this chapter to model each voxel's response to each of our stimuli described

in Section 2.4. The results of this model will be used in the following chapters. Finally, we discuss the lessons learned while developing the approaches in this chapter and how they can be applied to future analyses.

3.2 Related Work

There is an extensive body of work dedicated to pre-processing functional magnetic resonance images. Similarly to 2, in this section we will establish some basic principles that are commonly accepted in the field, then proceed to discuss sources that are closer in application to our experiments.

The textbooks by Huettel et al. [24], Smith et al. [56] and Ashby [2] have all expansively described the canonical steps to be taken to acquire, register, pre-process and prepare fMRI images for analysis. While this work will follow many of the key steps from those works, the unique nature of our image acquisition and several other factors require extra steps to be taken. This chapter will focus on the relevant parts from these works, and contribute to improve upon those steps.

The work of Robert Cox is very important to the field of fMRI analysis. The software package, Analysis of Functional NeuroImages (AFNI) [11, 13] has facilitated all aspects of fMRI processing. Over the years, it has evolved with the addition of many scripts and other functionality. It is still the de facto set of tools for analyzing fMRI images [12]. While the entirety of our preprocessing, and consequently the entirety of

this chapter was performed with AFNI, the tools aren't capable of handling all kinds of data with no external manipulation. Especially given our limited field of view, oblique-acquisition technique described in Section 2.3, AFNI's out-of-the-box algorithms failed. This chapter will focus on how we used the set of capabilities provided by AFNI to achieve the steps of preprocessing that is considered inconsequential in more conventional studies. Specifically, AFNI's implementation of Saad et al. [53] is used to register functional brain images to structural scans multi-modally, and it even has a parameter for partial registration, however our scan is too oblique and narrow for that tool to work without tweaking.

As it is the case with Chapter 2, we rely on functional localizers to identify voxels in the LOC and the visual word-form area. The process of extracting the localizer information from the voxel data using a general linear model is intertwined with the experiment design. As such, for the LOC localizer we will refer to the works of Kourtzi and Kanwisher [31], Kourtzi and Kanwisher [32], Kourtzi et al. [33], Grill-Spector et al. [19], Amedi et al. [1] once more. Specifically, for the localizer experiment we followed the design established by Kourtzi and Kanwisher [31], and for the calculation of the localizer we shall follow their work as well.

Similarly to localizing the LOC, our visual word-form area localizer calculations are also intertwined with the way the experiment is designed. We used the localizer approach described by Szwed et al. [57] in Chapter 2, and as such we will be following their work for the process of localization as well.

We follow guidelines for proper fMRI research practice described in Poldrack et al. [50]. Many fMRI studies report Talairach coordinates for their data, which are used for describing locations of brain areas using common atlases. These coordinates are often ambiguous and potentially unreliable. As per Poldrack et al., we use Brodmann areas as defined by the AFNI toolbox to report our localizer findings. We are also explicit about how our localizer ROIs are determined.

3.3 fMRI Data Format

As described in Section 2.3, fMRI data were acquired using a 3-Tesla Philips Gyroscan MRI scanner. The data format outputby Philips scanners is PAR/REC. Each run of the scanner produces a pair of files, one .par file, and its accompanying .rec file. The REC file contains the raw data from the scanner, and the PAR file is the header containing details about the data, the scanning and other auxiliary information. REC files are difficult to interpret on their own, as the data is simply stored in a linear fashion, regardless of the acquisition's dimensions, direction and timing. The information contained in the PAR file, which is generated from the parameters of the scanner, is necessary to convert the data into a readable format.

Specifically, the PAR file contains the following key information, among others that are not relevant to the preprocessing step:

1. **Number of slices**

fMRI images are acquired in sequences of two-dimensional slices along the Z-axis that form a 3D image when composited together. Each 2D image is a slice.

2. **Repetition time (TR)**

   Time Resolution is the time period in which the slices in the image are excited and allowed to demagnetize.

3. **Number of dynamics (how many TRs)**

   How many 3D images in total were acquired. This number, multiplied with TR, gives the total length of the run.

4. **Scan resolution in voxels, X and Y dimensions**

   The size of each 2D image in terms of voxels.

5. **Field of view in all three directions, in millimeters**

   The total physical size of the acquisition volume.

6. **Slice position in millimeters**

   This is the 3D location of the center of the scanning volume with respect to the scanner.

7. **Slice angulation in degrees**

   This is the 3D angle of the scanning volume with respect to the scanner.

The file also contains individual entries for each image acquired, detailing its position and other parameters. Other information contained in the file includes details of the experiment, like subject name and date, data types for numeric values found in the file, and more. However, the parameters enumerated above are the requisite information necessary for converting the raw fMRI data into a more readable format.

The Analysis of Functional NeuroImages (AFNI) software package (Cox [11]) is used for most steps detailed in this chapter. AFNI is a set of C scripts with command line arguments that is commonly used to process and visualize fMRI images.

Each of the 10 functional runs, the structural scan, the Lateral Occipital Complex localizer and the Visual Word Form Area localizer (for subjects where it was used) were converted into AFNI's BRIK/HEAD format pair. For each run, the BRIK file contains voxel data, and the HEAD file contains metadata. The AFNI script *to3d* was used for this. The above parameters found in the PAR file are sufficient for *to3d*, so a Python script was written to parse PAR files and call the *to3d* program with the appropriate parameters. The script takes the linear hexadecimal data outputted by the scanner and converts it into three-dimensional voxel data.

3.4 Registration

fMRI data can not be interpreted as-is directly out of the scanner. There are several factors that prevent proper analysis. These factors must be accounted for and fixed before the data is usable. This section will address these issues and how they were dealt with.

The prime problems that arise in fMRI experiments are:

1. **Subject head movement throughout scan**
2. **Baseline noise of the scanner**
3. **Scanner signal drift over time**
4. **Inconsistency of individual voxel activation values**
5. **The Hemodynamic Response Function**

The first issue to be addressed is the movement of the subject's head throughout the scan. Since our scan takes nearly two hours, it is natural that the subject's head will move during this process even though they are stabilized inside the head coil with pieces of foam. To track the activity of a voxel throughout the scan, we need to be able to either track its movement throughout the scan, or register every image acquired throughout the experiment to a base image. In addition, in order to verify our localizer results a posteriori via anatomical knowledge, we need to be able to transform our brain images to a statistical atlas. The elegant solution to this

problem is registering every brain image to a base image that is registered to an atlas, and re-sampling every voxel from every image to fit the voxels in our base image.

Normally the task of registering brain images to an atlas is rather trivial. AFNI has a script named *align_epi_anat.py* that takes two datasets, one structural and one functional and aligns them using Local Pearson Correlation [53]. Typically, the structural data would be an atlas or the base structural scan acquired along with the functional images, and the functional data would be the fMRI runs themselves. This script works very well for traditional purposes; however, in our case it fails.

Since we use partial brain acquisitions with the FOV being rotated from the base acquisition axes of the scanner in all 3 directions, and the inherent difficulty associated with registering images across modalities, every automated functional-to-structural registration algorithm we tried failed. These algorithms aren't typically designed to deal with partial brain images, and even those that have parameters designed to account for this case don't deal with a FOV as limited as ours. Because of this, we needed to develop our own method for registering our images.

As mentioned in Section 3.3, the parameters of the FOVs alignment can be extracted from the PAR files. While these do not account for the head movement between the structural scan and the first functional scan, they can be used as a starting point. Using these coordinates as a starting point to bootstrap the registration process, the functional image closest in time to the structural scan was registered to the whole-brain structural image. Then every other functional image was adjusted

with the FOV parameters, and subsequently registered to the first functional image. This registration was also verified by qualitative interpretation in addition to quantitative metrics. This process is demonstrated in Figure 3.1.

All 10 runs of the experiment were concatenated into a single run. These post-registration images are used for the purposes of the analyses, but for anatomical comparison, they still need to be mapped to an atlas. This is achieved by first registering the structural image to the atlas, then applying the transformation obtained from that registration to the first functional image registered to the structural scan as per above. Following the registration of the functional initial image to the structural image, every functional image was registered to the initial functional image to correct for subject head movement and resampled. This gives us a set of functional images all aligned to an atlas. For the purposes of the analyses in Chapter 4, we did not warp our data to an atlas as we didn't want to introduce nonlinear transformations to the data. However, this approach was essential for verifying the accuracy of the voxels selected by the localizers as described in the following section.

Further steps of processing include skull stripping, which is removing the parts of the image that are not brain matter. The voxels that correspond to skull tissue introduce a lot of noise to fMRI, and they can confound statistical analyses, so they are removed using the AFNI script $3dSkullStrip$. The script operates in three steps. First, it does pre-processes the volume to remove artifacts and repositions the brain image. Then it uses the BET algorithm by Smith [55] to expand a spherical surface

iteratively until it envelops the brain volume. Finally, it uses masks to extract the brain matter volume from the skull volume it separately detects.

A comparison of the skull-stripped and unprocessed structural image can be seen in Figure 3.1. As expected, the skull stripping algorithm failed with our narrow functional images, so we computed the strip on the structural image, then applied the same mask to the functional images post-registration.

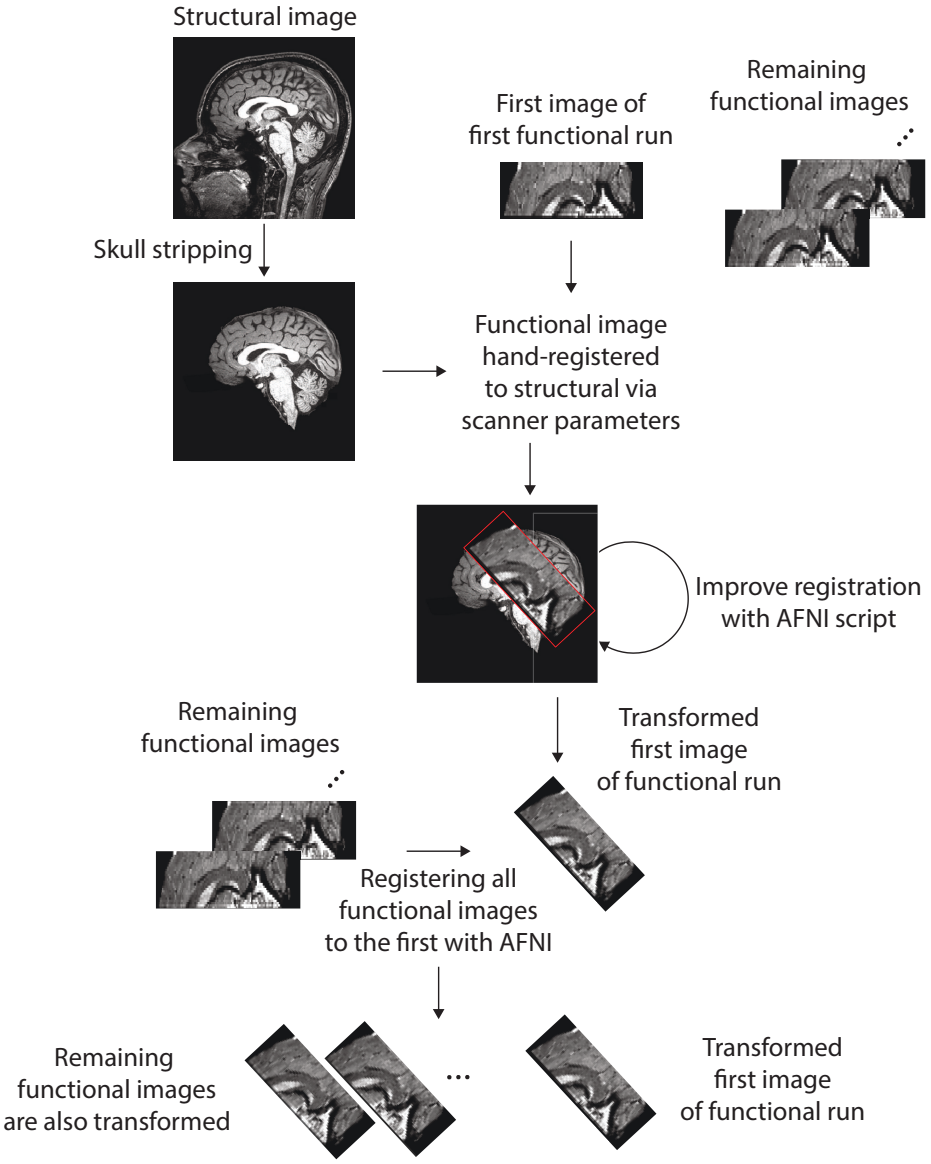


Figure 3.1: The registration process.

3.5 Localizers And Noise

In order to narrow down the voxels we will be conducting our analysis on and only select voxels that produce useful information, we will be using an LOC localizer. In addition, to ensure the uniqueness of our parametrization to LOC, we will be using a VFWA localizer. The implementation of these localizers is discussed in detail in Section 2.6. In this section we will discuss how the data acquired during the localizer runs is interpreted and how the voxels associated with the localizer are identified.

The main principle behind functional localizers is contrasting the activity of the brain when it is viewing images that the area one is looking for is believed to be coding against activity of the brain while it is viewing images that go against that characterization. Voxels that respond strongly to the desired images and do not respond to the negative images are selected. In other words, the localizer is a statistical model that uses the Hemodynamic Response Function as a kernel, the presentations of each image as a regressor with target images having positive weights and the "scrambled" images as negative weights.

More specifically, the localizers were modeled with a generalized linear model (GLM), where the data in each voxel were modeled as the time series:

$$Z(t) = K(t) \star S(t) + b + \epsilon \tag{3.1}$$

where t is time, $Z(t)$ is the BOLD time series, $K(t)$ is the HRF kernel, $S(t)$ is the

stimulus time series, b is the set of regressors that are of no interest (null hypothesis, constant, scanner signal drift, etc.) and ϵ is noise, and $\star$ is the convolution operator. This equation is to be solved for $K(t)$, resulting in obtaining coefficients for the sets of regressors. The localizer data is modeled with Ordinary Least Squares (OLSQ) regression, with $S(t)$ being defined by localizer stimulus presentation times, where intact stimuli are represented with positive sign and scrambled stimuli are represented with negative sign. Alternatively, we can write this equation as:

$$Z = X \times \beta \tag{3.2}$$

by absorbing the convolution and extra regressors into X, and solving for β. See Figure 3.2 for a visual representation of this equation. The AFNI tool $3dDeconvolve$ was used for this. Following standard AFNI practices and using their methods, a second-degree polynomial was used to fit the null hypothesis, which was used to account for noise and drift. Voxel values were z-scored to account for temporal inconsistencies like magnetic drift. Voxels that have an F-statistic with $p \leq 0.01$ are selected, then are clustered spatially using 6-nearest neighbors, and clusters with a size of less than 20 voxels were eliminated. This procedure results in a voxel mask that is used to functionally identify the targeted brain areas.

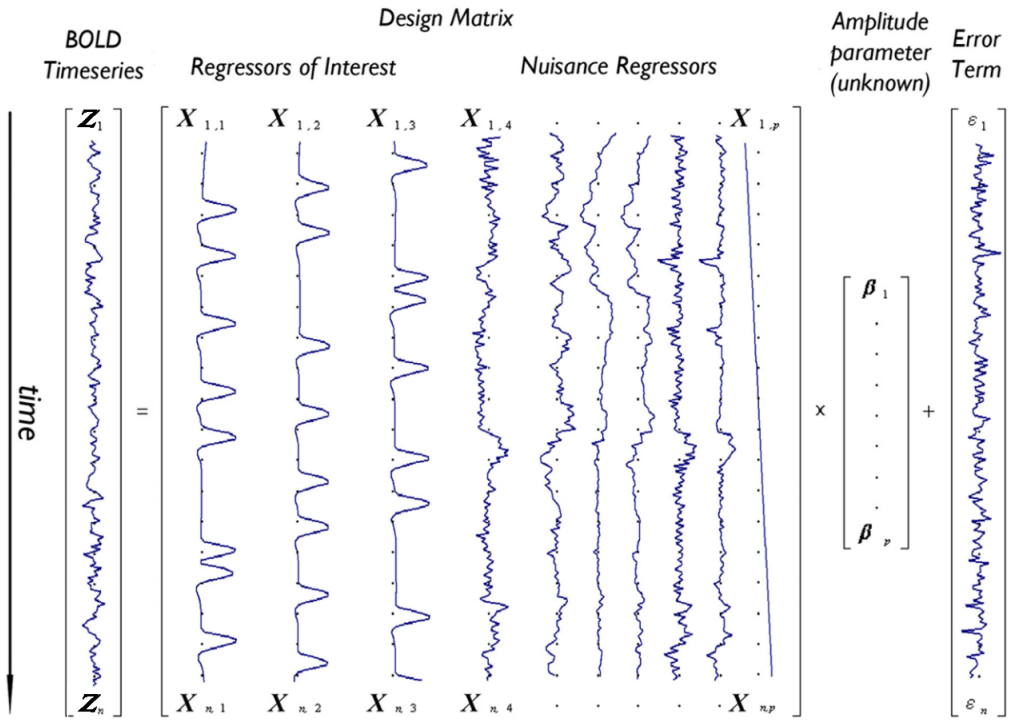
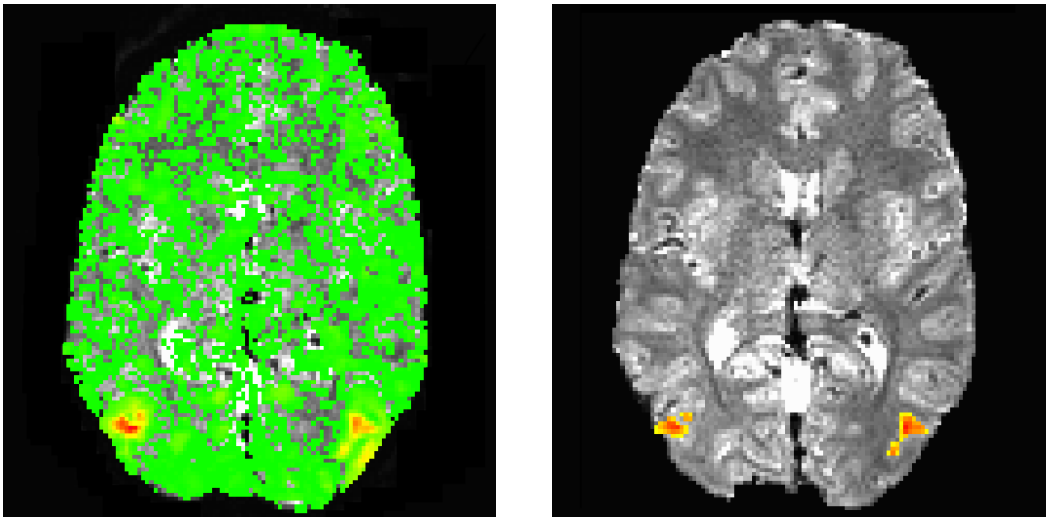


Figure 3.2: Depiction of the GLM model for an imaginary voxel with time-series Z predicted by a design matrix X including regressors of interest e.g., 3 tasks task regressors and seven nuisance regressors e.g., six motion parameters and one linear drift of unknown amplitude, and an error term. Figure adapted from [43]

3.5.1 LOC Localizer Results

For the LOC localizer, as mentioned in Section 2.6.1, 20 object images were used as positive regressors in the localizer, and 20 scrambled object images were used as negative regressors. The null hypothesis is the polynomial fit that accounts for the noise and other nuisance regressors. The f-statistic of the GLM in each voxel was thresholded for $p \leq 0.01$ and a 6-nearest-neighbor clustering scheme in three dimensions was used with a minimum size of 20 voxels. The purpose of this clustering was to eliminate noise by filtering out clusters of activity with less than 20 voxels,

which is a commonly used technique in fMRI analysis. That means the smallest area that our approach allows for would be a cube with each side being $\sqrt[3]{20} \approx 2.71$ voxels wide, with our 1.5mm voxel size that would mean the smallest possible area would be $4mm^3$. This is slightly larger than what most LOC studies have used as the size of an individual voxel ($3mm^3$) so it is not a stringent criterion. The three subjects each had unilateral areas of the brain selected by the localizer, as can be seen in Figure 3.3. The subjects had regions of size 107, 83 and 124 voxels selected. These areas were also verified both through visual assessment and AFNI's brain area localization tool. AFNI confirmed these areas to be in Brodmann Area 19, which is where the LOC is located [37].



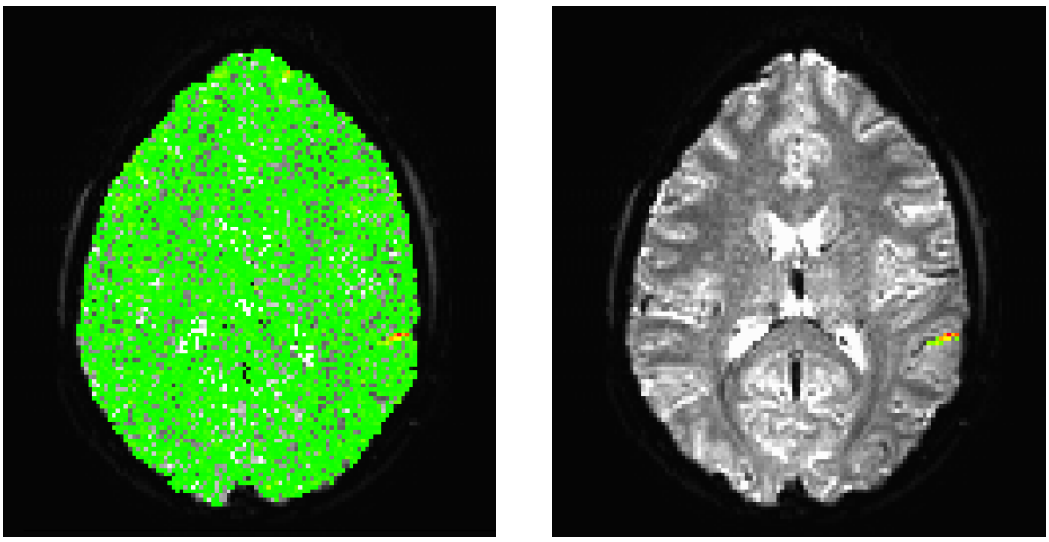
(a) The raw result of the GLM before thresholding.

(b) The f-statistic of the GLM results thresholded by $p \leq 0.01$, then clustered with a minimum cluster size of 20 voxels.

Figure 3.3: LOC localizer GLM results for Subject 1. The color map from green to red indicates the f-statistic of the GLM.

3.5.2 VWFA Localizer Results

For the VWFA localizer, as mentioned in Section 2.6.2, 24 word images were used as positive regressors in the localizer, and 24 scrambled word images were used as negative regressors. Similar to the aproach for the LOC localizer, the f-statistic of the GLM was thresholded for $p \leq 0.01$ and a clustering scheme was used with a minimum size of 20 voxels. As it is the case with the LOC localizer, this means the smallest area that our approach allows for would be a cube with each side being $\sqrt[3]{20} \approx 2.71$ voxels wide, with our 1.5mm voxel size that would mean the smallest possible area would be $4mm^3$. This is slightly larger than what most VWFA studies have used as the size of an individual voxel ($3mm^3$) so it is not a stringent criterion. With the VWFA localizer, we came up with the idea to include it after we ran the experiment with the first subject. As such, only the second and third subjects have their VWFA localized. The three subjects each had unilateral areas of the brain selected by the localizer, as can be seen in Figure 3.4. The subjects had regions of size 72 and 93 voxels selected. These areas were also verified both through visual assessment and AFNI's brain area localization tool. AFNI confirmed these areas to be in Brodmann Area 37, which is where the VWFA is located [22].



(a) The raw result of the GLM before thresholding.

(b) The f-statistic of the GLM results thresholded by $p \leq 0.01$, then clustered with a minimum cluster size of 20 voxels.

Figure 3.4: VWFA localizer GLM results for Subject 2. Note that the images for Subject 2 were acquired at a very oblique angle (as per Section 2.8) and thus are being viewed from an unconventional perspective. The color map from green to red indicates the f-statistic of the GLM.

3.6 Extracting Stimuli

Since fMRI is inherently noisy and unstable, stimuli need to be presented several times to obtain a consistent model of the response to any given stimulus. Due to this, as explained in Section 2.5, we display each stimulus 20 times to the subject. This introduces the issue of modeling the brain's response to each individual stimulus. Similar to Section 3.5, we model each voxel's response to each stimulus using a GLM. However, the difference in this section is that we use only voxels selected by the localizers for this analysis. Using the clustering results from each localizer, we create a mask of voxels for this GLM to be applied to. The process is otherwise identical, and Figure 3.2 also applies here. After the GLM is fit, the coefficients for each stimulus are extracted. AFNI's $3dmaskdump$ function is used to create a plaintext file containing a $\#\_voxels \times \#\_stimuli$ matrix with each voxel's activation value (the coefficient assigned by the GLM) for each stimulus.

Once this process is completed for the voxels selected by the LOC localizer, it is repeated for the voxels selected by the VWFA localizer, and it is saved in a separate file for the analysis in Section 4.5.2.

3.7 Conclusions

Our goal with this chapter was to take the raw fMRI images acquired using the approaches of Chapter 2 and prepare them for the statistical analyses described in

Chapters 4 and 5. In more conventional fMRI experiments, these steps are relatively simple. The functional images are registered to a structural image acquired during the same scanning run, motion-corrected, then a general linear model is fit onto the blood-oxygen-level dependent signal provided by the scanner, modeling the hemodynamic response function and eliminating noise, drift and other irrelevant features. Studies that require functional localizers simply include an additional GLM to identify voxels that respond specifically to the stimuli being presented. In this work, however there was a complication that we needed to address. Since our high-resolution scan that we proposed in Chapter 2 had a very narrow field of view and a very oblique acquisition, none of the registration algorithms that were commonly used worked. As a result, we had to improve the registration process to account for these images. We also adapted a French visual word-form area localizer into English, and validated its performance on identifying VWFA voxels.

The contributions in this chapter are directly tied to those from Chapter 2, as they are consequences of those methods. As such, this chapter serves less as a series of contributions, and more as a validation of the contributions introduced in the previous chapter, and steps tying those results to the following chapter. Future work that seeks to build on the high-resolution acquisition can refer to the technique used in this work to register fMRI images. Additionally, this chapter served to show that our adaptation of the VWFA localizer proposed by Szwed et al. [57] to English is viable.

Chapter 4

Statistical Modeling Of Structural Representation In LOC

As has been mentioned, the ultimate goal of this thesis is to derive a parametrization of shapes that is in line with how the lateral occipital complex models visual stimuli based on structural parameters. Everything so far in this thesis has been leading up to this analysis. In Chapter 2 we designed and conducted an fMRI experiment to obtain brain images that such a parametrization will be derived from and validated by. These experiments focused on highlighting nuances in medial axis shaped objects based on their structural information. Following those experiments, we used a variety of techniques in Chapter 3 to take the raw fMRI data and pre-process it into a format where stimulus-based information can be harnessed. We then described approaches we followed to extract stimulus values from the transformed images by modeling the

brain's response patterns to presentations and localizers. In this chapter, we will address the true challenge, that of developing a computational model for the voxel activations derived from the previous chapters in a manner consistent with a structural parametrization of the visual stimuli presented to the subjects. Furthermore, we will seek to show that this parametrization is applicable exclusively to the LOC, by showing that it's not applicable to other brain areas, and that parametrizations that describe the functionality of other areas do not apply to the LOC.

4.1 Overview

As a result of the experiments in Chapter 2 and the pre-processing in Chapter 3, we have a set of voxel activation values for each subject. Specifically, for each subject we have a $\#\_voxels - by - \#\_stimuli$ matrix of values that describe the response of each LOC voxel to each stimulus. For the secondary analysis, we also have a similar matrix for Visual Word-Form Area voxels, but our primary concern for the majority of this chapter is the LOC. These matrices will be used for all analyses within this chapter. We seek to derive a parametrization of object shapes that, with the fitting of a statistical model, can be used to describe the activity of the LOC voxels in response to the corresponding visual stimuli. While there has been previous work that has shown that the LOC has a role in recognizing objects, no previous work has attempted to actually model the behavior of the LOC in terms of mapping its

responses directly to parameters derived from visual stimuli. In monkeys, work has shown that the inferotemporal cortex, which is the homolog of the LOC, represents visual object stimuli in terms of medial axis and surface component parameters. The goal of this chapter, and ultimately this work is to establish a link between the work done on monkeys in the IT and our experiments by showing evidence for the LOC describing object parameters in a manner consistent with the functionality of IT.

This chapter is specifically concerned with modeling the activity of voxels found within the LOC in terms of medial axis shape components that represent parametric breakdowns of the visual stimuli used in Chapter 2. Chapter 5 will be focused on the inverse approach of reconstructing stimulus images from voxel activations. As a whole, these two chapters serve to provide evidence of structural coding of object shape in the LOC. Within the scope of this chapter, the process was to ascertain the validity of the data acquired and prepared throughout this work, then trying to derive a model that establishes a link between the stimuli and the activations elicited by them. This process was non-trivial, as conventional statistical modeling approaches failed for a variety of reasons. We will discuss the steps taken to arrive at the final parametrization and modeling method. We will also show that this parametrization is specific to the LOC, and it is a parametrization that describes the activity of the LOC better than other parametrizations of visual stimuli.

Section 4.3 will be concerned with demonstrating the viability of the data obtained as a result of Chapter 3's approach. More specifically, Section 4.3.1 will describe the

attempts at extracting information from the stimulus set used in the initial experiment (described in Chapter 2) using signal detection theory, and the lessons learned from the failure of these attempts. Section 4.3.2 will then show the improvements in results in the final studies compared to the previous section. Section 4.4 will detail the steps taken to arrive at the model that is the most significant computational novel contribution of this chapter. This will include models tried as candidates for the process, their results, and discussion of why they may have failed. Section 4.4.1 will provide a novel computational model for fitting our data, and use an independent data set to verify the performance of the model. Following the derivation of a model Section 4.5 will describe the approaches taken to demonstrate the statistical effectiveness and uniqueness of the model. In particular, Section 4.5.1 will describe the several cross-validation techniques used to validate the model and their results. Section 4.5.2 will compare the effectiveness of the developed model when applied to voxels identified by the LOC localizer to its effectiveness when applied to voxels identified by the VWFA localizer. Section 4.5.3 will demonstrate the performance of a Gabor Wavelet Pyramid model, which is the de facto technique for modeling the activity of voxels in the early visual system, on LOC voxels, and compare it to our method. Finally, we will discuss the novel approach we have developed for modeling, the results obtained within the chapter, the significance of these findings, and how this model and these findings can be applied to future work.

4.2 Related Work

There is a large body of work concerned with statistically modeling functional magnetic resonance images. As these modeling techniques span a vast variety of studies with different goals, it is difficult and possibly superfluous to attempt to address them all. The relevant studies fall into several categories. First, we will go over works that establish guidelines for fMRI analyses in all domains. Then, we will discuss the primate neurophysiology work that was the inspiration for and set the goals for our approach. Following that, we will examine fMRI studies on other relevant visual areas of the human brain that either help us justify our approach or provide a measure to test against. Finally, we will address work in statistics that has guided our approach.

4.2.1 General Guidelines For fMRI Analyses

As it has been the case with chapters 2 and 3, the key textbooks on fMRI analyses are important in this chapter as well. Many of the key techniques for statistical analysis of fMRI have been outlined by the seminal work of Huettel et al. [24], Smith et al. [56] and Ashby [2]. These texts provide guidelines for the kinds of statistical models that are best used with fMRI images, how to avoid pitfalls that are specific to the domain, and bets practices. Many modern fMRI studies apply Multi-Voxel Pattern Analysis, established by Norman et al. [47]. MVPA is a set of approaches

focusing on pattern-classification from a multi-variate perspective, focusing on groups of voxels as a whole instead of individual voxels. While we apply some of the wisdom from MVPA studies, our end goal is parametrization and not classification. As such, we need to look at our data with a finer lens than whole-brain or whole-region approaches. Combined with the neurophysiology work that motivates our approach (discussed in Section 4.2.2), this makes the MVPA approach unsuitable for this work. The Science paper by Mitchell et al. [41] is an important work in the field of fMRI for several reasons. The authors used whole-brain fMRI images to derive a fit a model to features learned from nouns. They were able to predict brain images and used a leave-two-out scheme for validating their results, which we will also use and compare our results to. While their approach is based on multi-voxel pattern analysis and whole-brain images, it is still worth considering their approach. Our early approaches based on linear regression and its variants were inspired by theirs, even though our final model was different.

Even though our analysis is unconventional in some ways, we still follow guidelines for proper fMRI research practice described in Poldrack et al. [50]. We use several statistical tests to support our conclusions, and as per previous chapters, we have been explicit about the steps taken to arrive at our results.

4.2.2 Neurophysiology Studies On Object Structure

As explained in Chapter 2, the primary motivations for our work come from monkey studies concerned with describing how the inferotemporal cortex codes for object structure. Since deriving a similar parametrization for LOC is the primary goal of this chapter, it is important to reiterate some of the background work here. Since the inferotemporal cortex is considered the homolog of the LOC in monkeys [37], the coding we will propose for the LOC is dependent on understanding how IT neurons code for object structure.

Studies as early as Pribram and Barry [51] and Wilson [64] have shown via ablation of inferior temporal regions in monkeys that these areas hold a visual function. Dubner and Zeki [17] have shown that the IT receives inputs from the primary visual area (V1) relayed by V2 and V4. Further studies by Dean [15], Gross et al. [20] and Mishkin [40] have shown via ablation that the IT specifically affects visual discrimination or recognition of objects. Perhaps the most important finding that establishes the groundwork for this work is by Tanaka et al. [58], demonstrating a specific coding for objects of differing structural parameters in IT.

Recent work in monkey IT has shown evidence of parametrization for specific characteristics of objects based on structural components. Parametric coding of object shapes in monkeys has been shown in Pasupathy and Connor [49]. Neuron recod-

ings have enabled the authors to reconstruct stimuli being viewed, and this work will also follow a similar design, but with fMRI voxels instead. This chapter will use the information for modeling voxel activity, but Chapter 5 will further detail reconstruction. The IT has been shown to code for combinations of straight and curved shape fragments by Brincat and Connor [4]. The authors have used a model to describe the response to stimuli as parametrized by such fragments, which this chapter will also focus on. We will attempt to model the responses to those fragments in terms of their curvature. Yamane et al. [65] have developed a coding for three-dimensional object shape and spatial configurations based on IT neuron responses. Due to fMRI limitations a parametrization with the complexity of Yamane et al.'s is out of reach (see Chapter 2 for discussion of fMRI limitations), we have taken into account the spatial variations of fragments as a factor in our parametrization, specifically polar positions of fragments around the visual fixation point. The most directly relevant work to our is Hung et al. [25], who have used medial axis stimuli to demonstrate that the tuning function of IT neurons can be modeled using fragments of such stimuli. While globular shape curvature (as per the previous studies discussed) based parametrizations would have been too complex to gather enough data for an fMRI parametrization, medial axis shapes are much more suited for a discrete analysis like ours. As a result, our work has uses medial axis stimuli as described in Chapter 2.

4.2.3 fMRI Studies On The Visual Cortex

4.2.3.1 Early Visual Areas

In the past decade or so, there has been an increasing focus on trying to parametrize the activity of voxels in earlier visual areas. The V1 has been the focus of most of these studies, as the functionality of the V1 is both well-documented and similar between monkeys and humans. Many studies have shown that the V1 codes for visual stimuli in terms of spatial frequencies, color and orientation [61]. Specifically, Gabor Wavelets have been used to characterize the way V1 neurons (and consequently voxels) respond to images with a sparse coding [48]. Singh et al. [54] established using fMRI on the primary visual area (V1) with their seminal work. They demonstrated spatial frequency based coding using noninvasive imaging, correlating what neuron recording studies have shown. This justifies our approach in modeling LOC voxels with models that are based on IT neurons.

Several recent fMRI studies have used Gabor wavelet based approaches to parametrize the functionality of voxels in early visual areas. As mentioned earlier, a large amount of fMRI work is based on classifying brain activity. Unconstrained reconstruction is a more challenging task. Miyawaki et al. [42] have reconstructed binary image patches using such an approach. Looking at early visual areas and parametrizing image stimuli with Gabor wavelets, they fit a model to voxel activations to predict images associated with them. Their approach is relevant to this work

in two avenues. Firstly, it demonstrates the possibility of using a parametric approach to modeling voxel activities, and secondly, it provides an early visual cortex model for which to compare our approach against. By comparing the parametrization this work proposes against one that targets a different visual area, we will show that the LOC does not code for images in the same fashion as the V1. In addition, Miyawaki et al. have used letter-like stimuli similar to our medial axis shapes, which makes the comparison between the performance of the two models more compelling and convincing. Similarly, Naselaris et al. [45] and Vu et al. [63] have demonstrated the further viability of Gabor-based models on early visual areas.

Another group of ground-breaking work is the research of Jack Gallant and collaborators. Their work, specifically Kay et al. [28], Kay and Gallant [27], Naselaris et al. [44], Naselaris et al. [45] and Vu et al. [63] has used fMRI of early visual areas to reconstruct natural images being viewed by subjects. These works have used Gabor wavelet decompositions to parametrize images, then use a bag-of-words inspired model to reconstruct images being viewed by the subject. These works have helped us refine our approach greatly, and we have used their data as a measuring tool to compare our results to. They have also demonstrated the power of fMRI as a tool that can be used to gain significant understanding of voxel activities with statistical models.

4.2.3.2 Lateral Occipital Complex

The Lateral Occipital Complex is a relatively under-explored area of the brain in fMRI when compared to earlier areas like V1. The area is smaller compared to the V1, and its functionality is of a higher order, which means that it is more complex and harder to analyze. Despite this, there have been a few studies in this area, ascertaining its function. One of the first fMRI studies to explore the area was Malach et al. [37], showing that the area does not respond to spatial frequencies, but instead responds to images of objects. Similarly, Grill-Spector et al. [18], Kourtzi and Kanwisher [31], Grill-Spector et al. [19], Kourtzi and Kanwisher [32], Amedi et al. [1] and Kourtzi et al. [33] have all shown evidence for LOC responding specifically to images of intact objects. This is consistent with our understanding of how IT works, which gives us justification to explore a parametrization of LOC similar to those of IT.

4.2.3.3 Visual Word-Form Area

The visual word-form area is also a relatively under-explored area of the brain. Since it is considered to be exclusive to humans, monkey studies do not provide much insight into its functionality. Cohen et al. [9] was one of the early fMRI studies to explore the area. Price and Devlin [52] disputed the existence of the area citing their functional imaging study where the area was activated even when not viewing words, however further studies have erased this doubt. Examples of such studies include Hillis et al. [22], Cohen and Dehaene [8], Kronbichler et al. [34], Dehaene and Cohen

[16] and many others. Szwed et al. [57] have provided a robust and comprehensive exploration of the area, and we follow their methodology. These studies have shown that the VWFA codes for letters with meaning, and does not respond to simple letter-like fragments. We will be looking at the VWFA to validate our model, in order to show that the parametrization that we derive is exclusive to the LOC and does not apply to other higher visual areas that respond to similar shapes.

4.2.4 Statistical Modeling Techniques

The novel modeling approach we will propose is based on expanding concepts introduced by stepwise regression. Even though stepwise regression is commonly used as a way to model data ([3, 7, 10, 35] and many more) Jennrich and Sampson [26] and Thompson [60] have pointed out that it can be prone to overfitting data, and can lead to greedy solutions. As a result, we will be extremely careful to avoid this case, setting strict constraints for our convergence, and performing extensive cross-validation tests to verify our results. Additionally, we will propose a novel algorithm based on stepwise regression that addresses some of its shortcomings by training on two groups kept independent of each other to prevent overfitting either.

4.3 Ascertaining The Distinguishability Of Stimuli

While the generalized linear model obtained in Section 3.6 represents the activations of each voxel in response to each stimulus, we have yet to determine whether those activations contain any useful information or are statistically distinguishable from each other. The goal of this section is to analyze this set of data to determine its usability for future analyses that form the core of this work.

4.3.1 Initial Study

This part of the process was originally developed when we had our initial data, as referenced in Chapter 2. When we acquired that data, we skipped this step and went straight to trying to model the voxel activations based on our parametrization of the stimuli. We tried many approaches to modeling; however none of them yielded meaningful results, statistically or otherwise. We began to suspect that the data itself wasn't good enough, that there might be some factor that was a barrier to this modeling. We set out to examine this data to great extent, in order to discern the cause of the issue and improve the stimulus set and/or experiment design for the following studies.

Before delving into any type of parametrization, we wanted to see if the stimuli were differentiable in terms of the voxel activation values each of them evoked. We

used several approaches to determine this. First, we approached this problem from an analysis of variance (ANOVA) perspective. We treated each stimulus label as a group, and each voxel's response to that stimulus a sample of that group. Our aim in this analysis was to see if the groups had distributions that were distinguishable, which would mean we could easily model the response of the brain to each stimulus. Unfortunately, this ANOVA yielded a p-value of $p = 0.41$, which was not conclusive. In light of this, we decided to proceed to further, more nuanced analyses.

Our initial idea for parametrization was to map every stimulus to a set of points in the three-dimensional space of curvature, orientation and polar position as envisioned in Section 2.4. This experiment was done with the two-level-deep stimuli, so each stimulus could have two to a dozen fragments. With the results of the stimulus-based approach being inconclusive, we decided to examine whether the signals were differentiable on a fragment basis, considering our parametrization was based on fragments. We measured the sensitivity index (d') of the presence of each fragment. More specifically, for each fragment, we separated the stimulus set into two groups. A group of stimuli that contain the fragment and a group of stimuli that don't contain it. Using the d', we compute a Receiver Operating Characteristic curve, and computer the area under the ROC curve. We use the area under ROC as a measure of how well the stimuli are distinguishable based on this parametrization.

As this analysis didn't yield conclusive results, we began to suspect that the signals for each stimulus or fragment potentially weren't easily differentiable, but

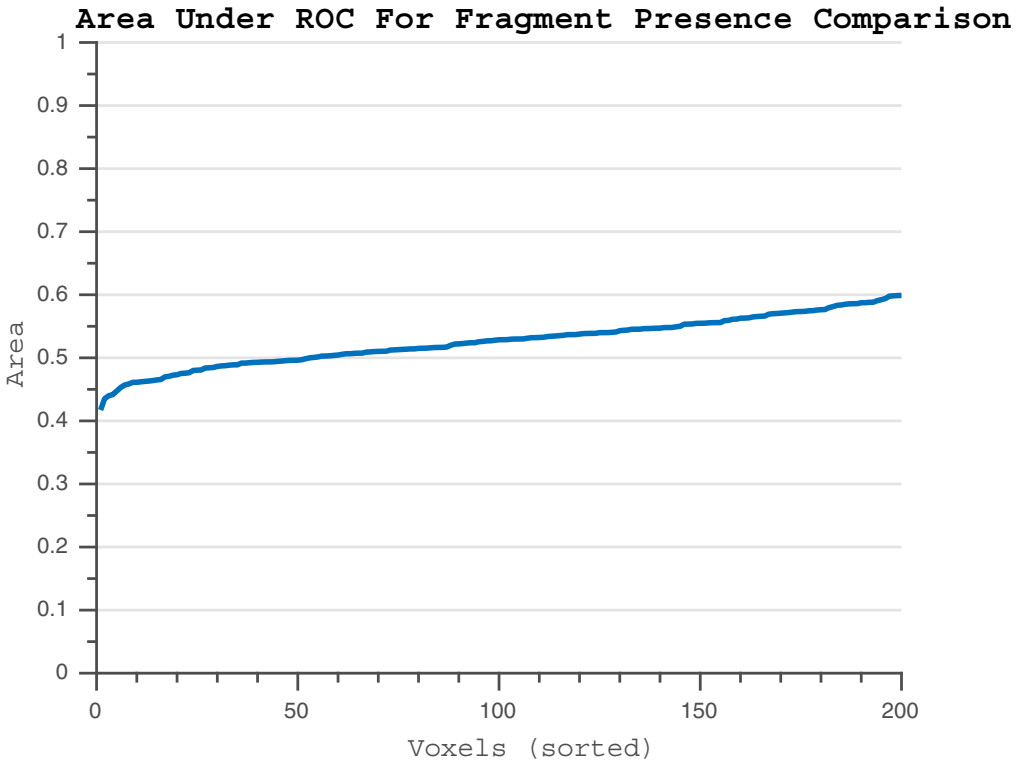


Figure 4.1: Plot of the area under the ROC curve for each fragment from the initial experiment. The fragments are sorted by the area under ROC value.

they could be modeled instead. We turned to the monkey inferotemporal cortex work by Pasupathy and Connor [49] for guidance. In that work, they derive a tuning function for each neuron, which is a function for describing how the neuron responds to any stimulus. The resulting tuning maps they derive for each neuron resemble combinations of Gaussian distributions, thus we decided to replicate their work by using a Gaussian Mixture Model to describe each voxel's activity. For each voxel, we aim to solve:

$$p(\mathrm{x} \mid \lambda) = \sum_{i=1}^{192} w_i \, \mathcal{N}(\mathrm{x} \mid \mu_i, \Sigma_i) \tag{4.1}$$

$$\textbf{w.r.t} \sum_{i=1}^{192} w_i = 1 \tag{4.2}$$

$$\lambda = \{w_i, \ \mu_i, \ \Sigma_i\} \quad i = 1, \cdots, 192 \tag{4.3}$$

Where x is the vector of the voxel's responses to each stimulus, $w_i, i = 1, \cdots, 192$ are the mixture weights, one for each possible position in our feature space (as described in Section 2.4), and $\mathcal{N}(\mathrm{x} \mid \mu_i, \Sigma_i)$ the component Gaussian density functions:

$$\mathcal{N}(\mathrm{x} \mid \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} \mid \Sigma_i \mid^{1/2}} \exp\{-0.5(\mathrm{x} - \mu_i)'\} \tag{4.4}$$

with D the number of stimuli, mean vector μ_i and covariance matrix Σ_i.

We solve Equation 4.3 for λ using the Expectation Maximization algorithm. The EM Algorithm works by iterating between two steps. In the expectation step, we

compute w_{ij} using Equation 4.5 where w_{ij} is the membership weight of point j in cluster i. Then, in the maximization step, we compute new values for μ_i, Σ_i using Equation 4.6. Where N is the number of voxels.

$$w_{ij} = \frac{p_i(x_j \| \mu_i, \Sigma_i) w_i}{\sum_{m=1}^{192} p_m(x_j \| \mu_i, \Sigma_i) w_j} \tag{4.5}$$

$$\hat{w_i} = \frac{\sum_{j=1}^{N} w_{ij}}{N} \hat{\mu_i} = (\frac{1}{N_i}) \sum_{j=1}^{N} w_{ij} x_j \hat{\Sigma_i} = (\frac{1}{N_i}) \sum_{j=1}^{N} w_{ij} (x_j - \hat{\mu_i})(x_j - \hat{\mu_i})' \tag{4.6}$$

Given the weights and parameters for each voxel's model, we are now tasked with validating these models. We choose two approaches to measure the validity of this model: leave-one-out cross-validation and leave-two-out cross validation. For the first approach, we separately leave out each stimulus from the set of stimuli, then use models learned from the remaining stimuli to try to predict each voxel's response to each left-out stimulus. In this method, we use Normalized Root-Mean-Square Error as the metric. The results from the leave-one-out test were not encouraging.

As can be seen in Figure 4.2, the residual error percentage was very high, with a vast majority of cases being having somewhere around 80%-90% residual error.

The second test employed was leave-two-out cross-validation. As used in Mitchell et al. [41], this analysis is a benchmark for validating the predictive power of models learned from fMRI data. Following Mitchell et al. we use the following scheme, where we leave out two stimuli from the set, train the models on the left-in stimuli, then

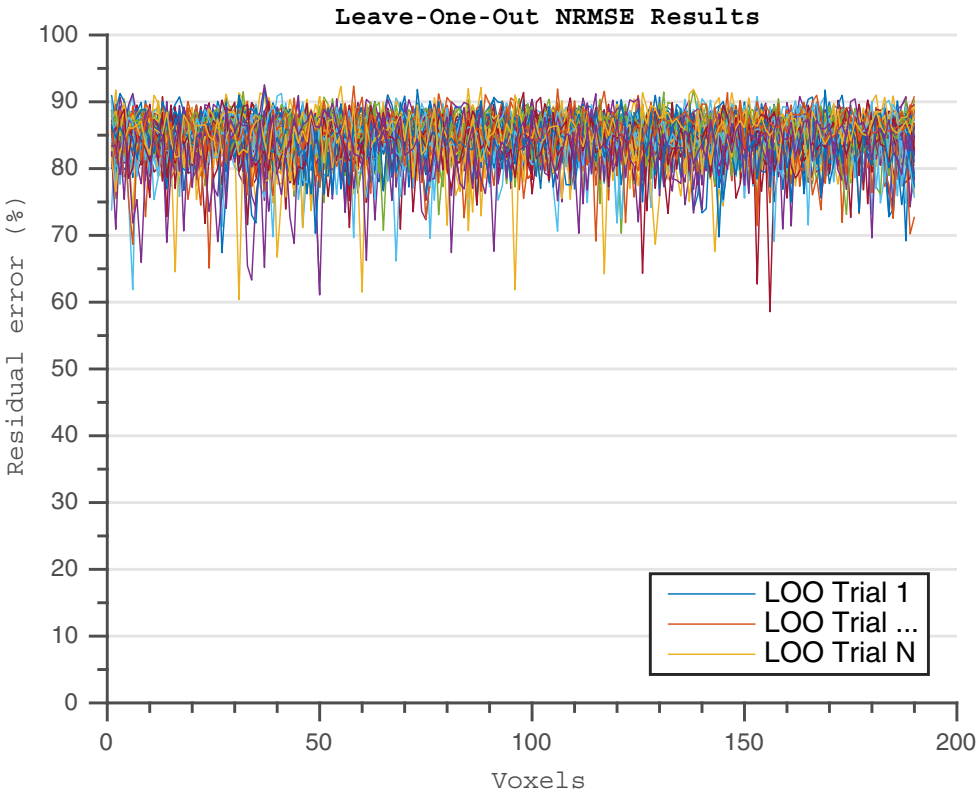


Figure 4.2: Leave-One-Out results. Each color represents a different LOO trial, and their residual error percentage according to NRMSE is displayed. Voxels are not sorted. The error percentage is quite high for all trials.

use the models to predict the voxel activation patterns for the left-out stimuli. We compare the predicted values for the left-out stimuli to their observed values:

$$cossim(A_1, P_1) + cossim(A_2, P_2) \overset{?}{>} cossim(A_1, P_2) + cossim(A_2, P_1) \tag{4.7}$$

Where $A_{i=1,2}$ are the actual voxel activation patterns for the left out stimuli, and $P_{i=1,2}$ are the predicted voxel activation patterns for the left out stimuli based on the model. $cossim(X, Y) = \frac{X \cdot Y}{\|X\|\|Y\|}$ is the cosine similarity. If the inequality holds, the model was marked as "correct". With this scheme, choosing by chance is at 50%. Combining the accuracy of all leave-out trials, the result was 50.8%, which was a poor result.

All of our efforts to validate the parametrization we derived were unsuccessful. We suspected several potential causes for the failures. The resolution of our scans was likely not high enough to get enough specificity to determine a parametrization as complex as we designed. Our stimulus set was quite large and didn't span the set of possible configurations well enough despite using a sizable amount of stimuli. We didn't have enough repetitions for each stimuli displayed to the subjects in our experiment design. As a result, we redesigned our experiment, our stimulus set and our parametrization. We increased the scanning resolution, reduced stimulus complexity and increased number of presentations per stimulus to increase the quality of

signal we obtain for each stimulus. This process is detailed in Chapter 2. With the improved design, we were able to acquire data that is more robust, which is detailed in the next section.

4.3.2 Final Studies

With the second set of experiments, which also ended up being the final experiments, our aim was to boil down the parametrization to a simpler scheme so that we could get more useful signal from our scans. How the second experiments differ from the initial experiment is explained in more detail in Chapter 2. Once we have this new set of data and pre-processed it as described in Chapter 3, we moved on to similar analyses to those in Section 4.3.1. Since our stimulus set uses a fragment alphabet that is significantly simpler compared to the initial experiment, we decided against trying to parametrize exact values of curvature, orientation and polar position. Instead, we turned our parametrization into a binary presence descriptor. Stimuli are each represented by a binary vector with the length of the fragment alphabet, where each element of the vector denotes whether a given fragment is a part of the stimulus or not. This simplifies our model greatly without sacrificing any descriptive power given the set of stimuli we have, which increases the usefulness of the signal we obtain. This change in parametrization was motivated by the d' analysis of the previous experiment, and the fact that our stimulus set spans a much smaller space of possible stimuli, which would restrict the rank of any model we would derive based on actual

angle values. Limiting our parametrization in this manner allowed us to use a model that doesn't have to account for variability in fragment terms per stimulus, so we can use regression-based models.

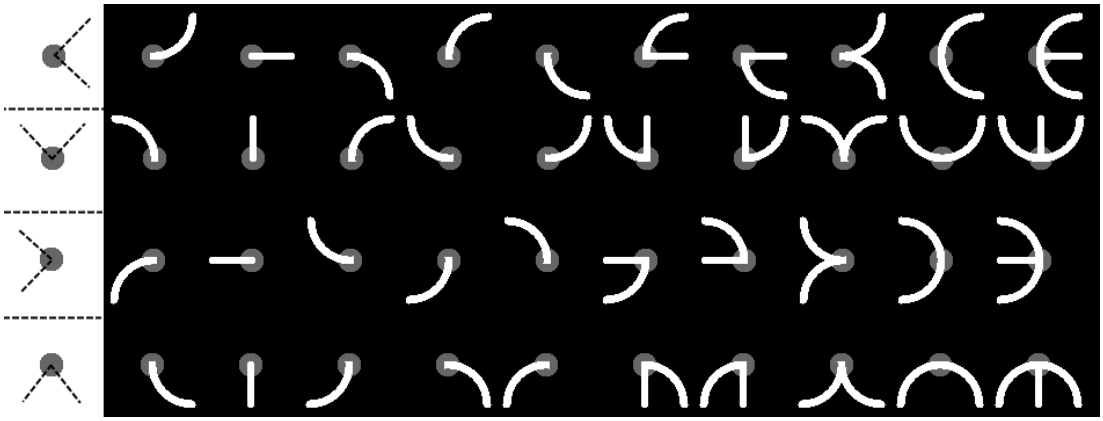


Figure 4.3: The set of all fragments used in our final model, except for the "blank" fragments. The figure is arranged such that every row is one quadrant (demonstrated by the graphic on the first column) and the fragments in each column are consecutive 90 deg rotations of each other. The first five fragments in each row are the individual fragments, the rest are compound fragments.

Following the new experiments and parametrization, we performed the previous d'-based ROC analysis on each of our new experiments. The results, as seen in Figure 4.4 were notably more promising than the initial experiment.

4.4 Developing a Computational Model

After validating our data and shown that it contains discernibly different signal for fragments, our next task was to develop a model for the activations of LOC voxels based on the fragment composition of stimuli causing those activations. Following the neurological results in Pasupathy and Connor [49], we decided to model the

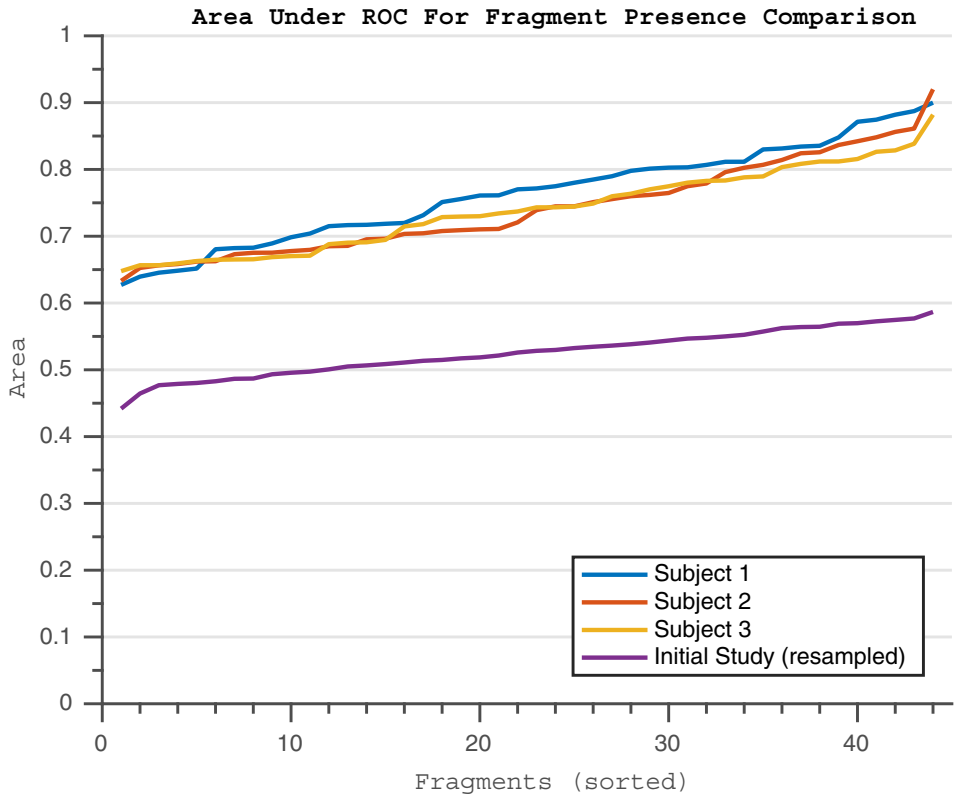


Figure 4.4: Plot of the area under the ROC curve for each fragment from the final experiment, compared to the initial experiment. The fragments are sorted by the area under ROC value. Note that the initial study had 200 fragments, and the final study has 44, so the initial study's curve is resampled down to 44 points for the purposes of this comparison using linear interpolation. The original curve can be seen in Figure 4.1

activations of each voxel in response to the stimuli in terms of a linear combination of the fragments contained within that voxel as such:

$$Z = \beta\mathcal{X} + \epsilon \tag{4.8}$$

Where Z is the $N \times 60$ activation pattern of N voxels as per section 3.6, β is the $N \times F$ matrix of coefficients to be estimated for fragments, and $\mathcal{X}$ the $F \times 60$ binary matrix

of fragment presences per stimulus as described in section 4.3.2, and ϵ is noise. N is the number of voxels selected by the localizer, and F is the number of fragments that are used in the alphabet. Since each voxel is treated as independent, this equation can be broken down into $Z_i = \beta_i \mathcal{X}$ where Z_i is the response pattern of voxel i to each stimulus, and β_i is the vector of coefficients for that voxel.

The simplest way to solve Equation 4.8 would be linear regression. However, there is a complication that arises when trying to set up the regression. In the parametrization, there are five unique single-component fragments per quadrant, and their four rotations around the origin. However, in the stimulus set, we have allowed multiple fragments to occupy the same quadrant, creating compound fragments. Since we envision our model working on a per-quadrant basis, if we are to use a vector's value to binarily denote whether a quadrant contains a fragment, that model encounters issues when multiple fragments are present in the same quadrant. This introduces a nonlinearity to the model. There are two solutions to this issue. We could either use a nonlinear model with the base fragments to account for the compound fragments, or we could use a linear model and introduce extra terms for the compound fragments. We ended up choosing the latter approach for several reasons. It's a simpler model, which is more elegant and has grounding in the neurology work; it's computationally easier to formulate and solve and most models in fMRI have been linear models; and it makes the reconstruction process, which we will address in Chapter 5 more elegant as well. While we initially chose the linear-model-nonlinear-fragments approach, once

the analysis was done, we also went back after the fact and tried the other approach, which can be seen in Section 4.5. The results were not distinctly different, so for the sake of elegance and consistency with literature, we converged on the linear model. In addition to compound fragments, we have also added an "empty fragment" for each quadrant signifying the absence of a fragment in that quadrant.

We have tried several approaches towards solving Equation 4.8. While many of our approaches have failed, it is still important to detail their failures here in order to justify the final approach.

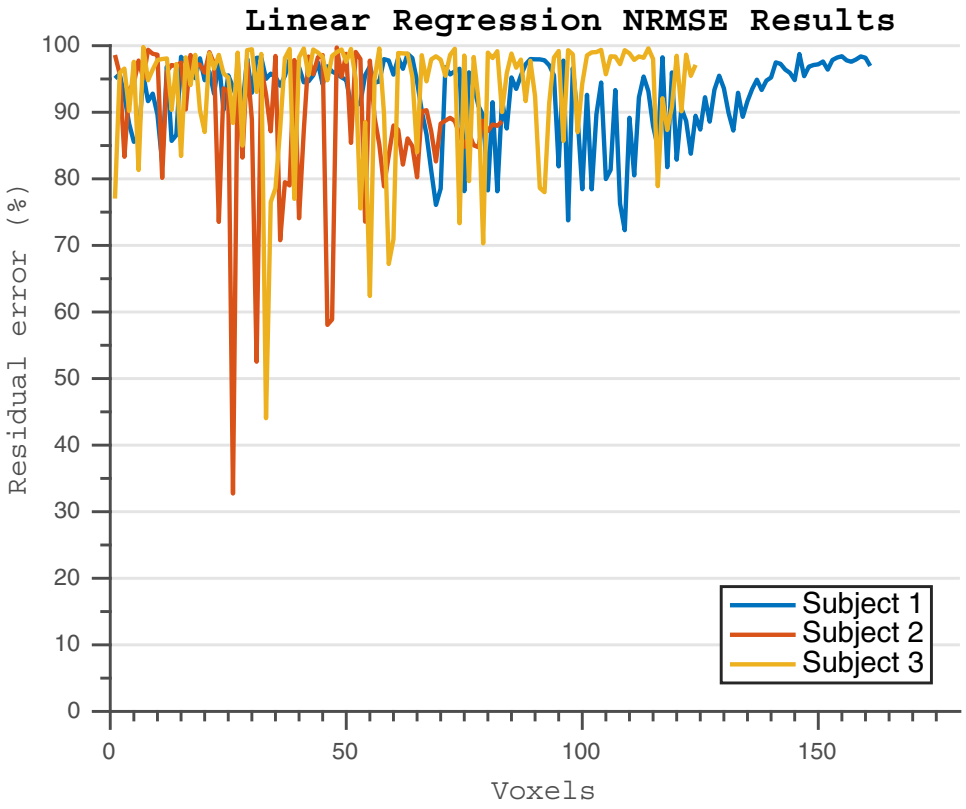


Figure 4.5: The residual error percentage according to NRMSE of applying a linear regression solution to Equation 4.8. Voxels are not sorted.

As mentioned earlier, the first approach we tried was linear regression. We attempted to solve Equation 4.8 using MATLAB's linear least squares solver with the QR decomposition. As can be seen in Figure 4.5, the results were not very strong. We initially suspected that this might be due the the problem being ill-posed and the variables correlated (which is the care for our compound fragments), hence we opted to try Tikhonov regularization, also known as ridge regression as such:

$$\underset{\beta}{\text{minimize}} \ \|Z - \beta\mathcal{X}\|^2 + \|\Gamma\beta\|^2 \tag{4.9}$$

$$\hat{\beta} = (\mathcal{X}'\mathcal{X} + \Gamma'\Gamma)^{-1}\mathcal{X}'Z \tag{4.10}$$

Where $\Gamma = \alpha\mathbf{I}$ is the Tikhonov matrix, α the regularization term and $\mathbf{I}$ the identity matrix. $\hat{\beta}$ is the solution to the minimization problem. Note that for $\alpha = 0$ this solution is identical to the least squares solution. We have tried different values for α to exploit different assumptions about the data, but these attempts proved to be unfruitful as well.

To understand why these approaches failed, we went back and examined the assumptions we have made about the data. In the monkey work by Pasupathy and Connor [49], IT neurons each code for a small selection of orientations, positions and curvatures. Trying to fit a model with the assumption that each neuron will explain each stimulus does not work. Based on that, we supposed the same might be applicable to LOC voxels. Namely, expecting each voxel to respond to every fragment, and

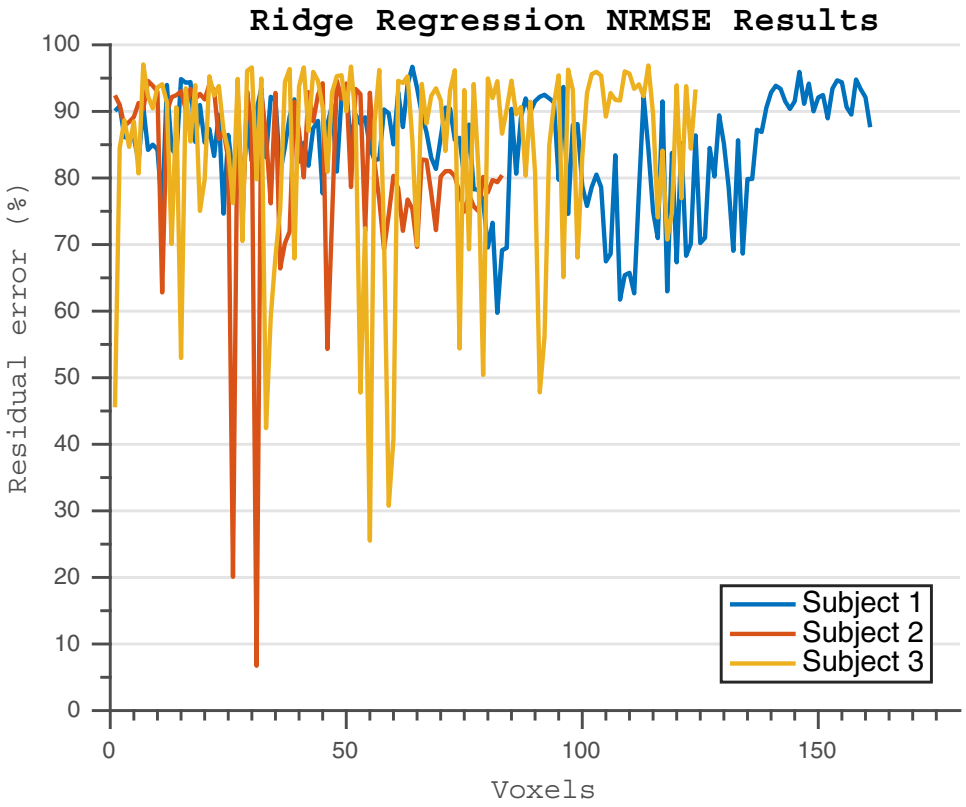


Figure 4.6: The residual error percentage according to NRMSE of applying a ridge regression solution to Equation 4.8. Voxels are not sorted.

by extension every stimulus is not consistent with neurological knowledge. Considering approaches like ridge regression are geared against setting coefficients to zero, they do not fit well with our approach. Thus we turned to approaches that seek solutions with small amounts of coefficients.

The most commonly used technique for regression that seeks to reduce the number of elements in the model is the least absolute shrinkage and selection operator, also known as the lasso. In its most general form, the lasso operates by adding the $L1$-norm of the coefficient matrix as a penalty into the minimization. This penalty term works

to ensure the sparseness of the model. This approach was inspired by neuroscience work that shows early visual cortex using a sparse coding scheme for stimuli [48].

$$\underset{\beta}{\text{minimize}} \; \|Z - \beta\mathcal{X}\|^2 + \lambda|\beta|_1 \tag{4.11}$$

Where λ is the weight parameter. One issue for consideration with the lasso is the weighting of the penalty term compared to emphasizing the fit of the model. Thus, we experimented with many different values for λ.

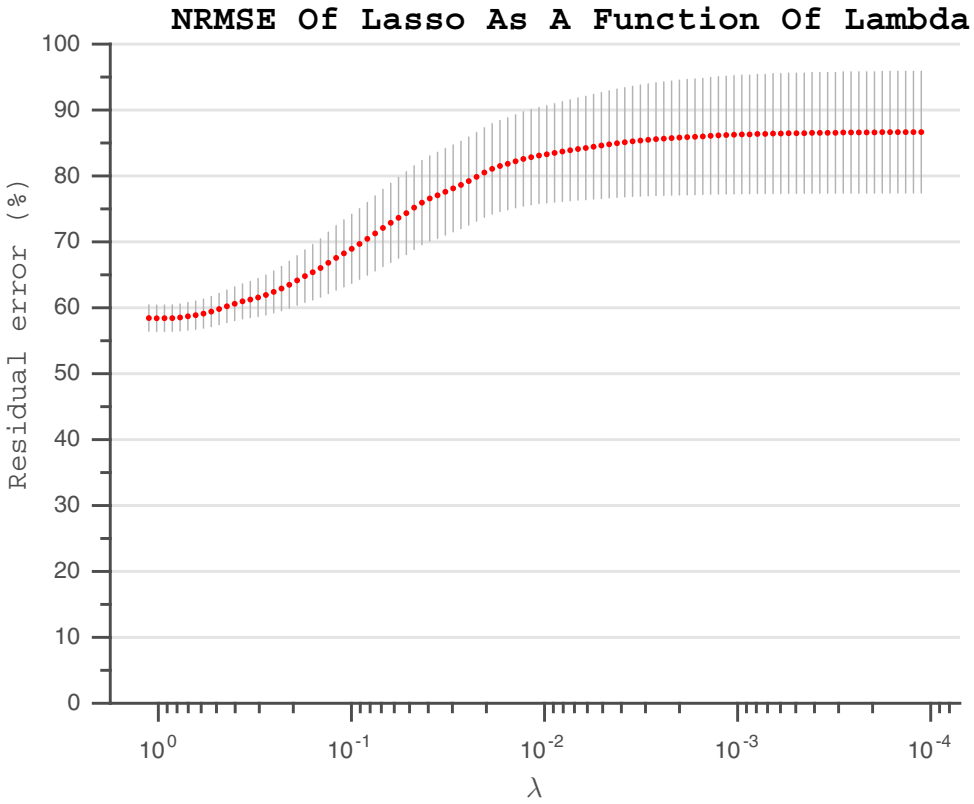


Figure 4.7: The residual error percentage according to NRMSE of applying lasso. Only Subject 1's results are displayed as they were the best. The mean NRMSE across all voxels is calculated, with the error bars showing cross-validation performance.

As can be seen in Figure 4.7, lasso did not perform significantly better than other approaches. It also performs poorly when using cross-validation, (our cross-validation approach discussed in Section 4.5). Considering our data is still quite noisy, it is expected that the lasso isn't robust enough to model it. Thus we developed several alternative approaches.

One particular method of calculating a regression with a small number of components is stepwise regression. This is an iterative procedure that starts with a linear fit, then adds to or removes from the model's terms by forcibly settingcoefficients to zero, then re-fits a new model based on the new set of terms. It can be used to find an approximate solution for Equation 4.11 using Algorithm 3.

Algorithm 3: Backwards Stepwise Regression

Data: Voxel activation pattern for voxel i: Z_i;
Binary fragment presence matrix: $\mathcal{X}$;
Threshold for convergence: τ;
Result: Fragment coefficients: β_i
while *model not converged* **do**
  Solve $Z_i = \beta_i \mathcal{X} + \epsilon$ for β_i using OLSQ;
  $S \leftarrow$ Sum-squared-error (SSE) of current fit;
  for $f \leftarrow 1$ **to** *44* **do**
    $\hat{\beta}_{if} \leftarrow \beta_i$;
    $\hat{\beta}_{if}[f] \leftarrow 0$;
    Solve $Z_i = \hat{\beta_{if}} \mathcal{X} + \epsilon$ for $\hat{\beta}_{if}$ using OLSQ;
    $\hat{S}_f \leftarrow$ SSE of current fit;
  find f such that $\hat{S}_f$ is smallest;
  if $\tau > abs(S - \hat{S}_f)$ **then**
    $\beta_i \leftarrow \hat{\beta}_{if}$;
  else model has converged;

The parameter τ in Algorithm 3 gives a more explicit degree of control over

selection of the number of variables than λ in Equation 4.11. However, Algorithm 3 is greedy, as it selects the best model achievable by removing a single coefficient at each step. The advantage of stepwise regression is its iterative nature, where we can use each step to make adjustments that would otherwise be difficult to make with other models. Since our input data is in binary format, it is quite degenerate. This is why gradient-based optimization solutions don't perform well with our data. Since Algorithm 3 iteratively applies a model selection approach, it is suited for dealing with degenerate data like ours.

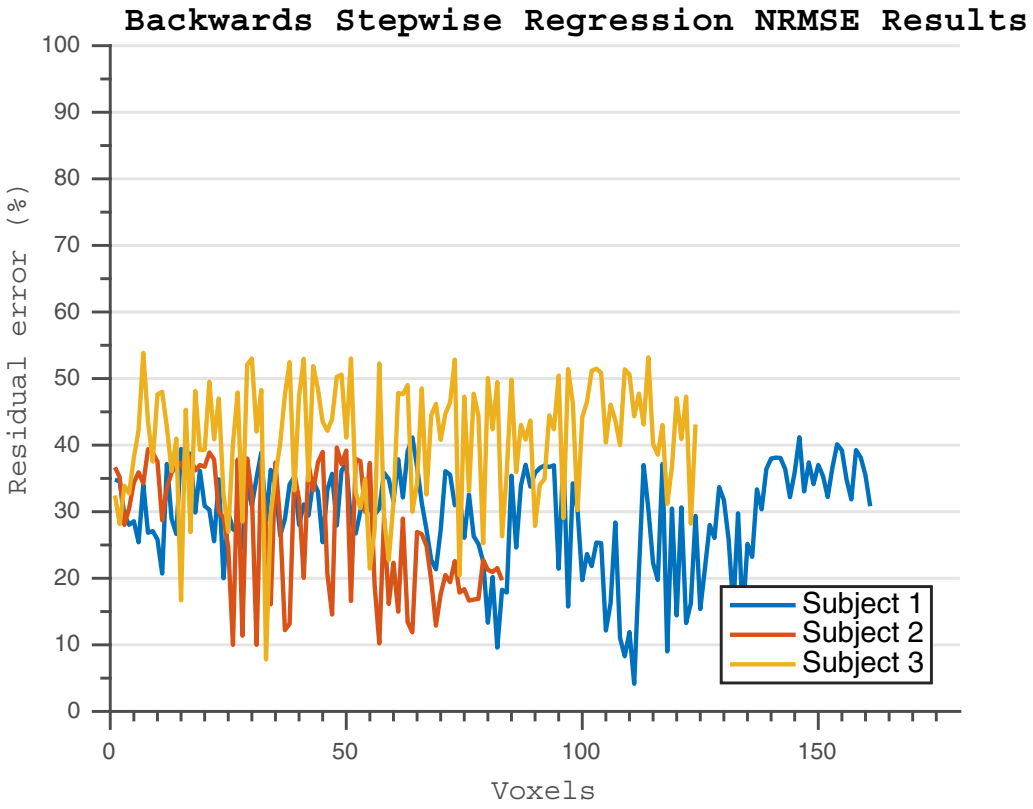


Figure 4.8: The residual error percentage according to NRMSE of applying the BSR approach in Algorithm 3

4.4.1 Developing A Novel Modeling Approach

Up until this point, we have used existing modeling approaches to characterize the parametrization of the stimuli into the fragments. Since overfitting was a concern with using stepwise regression, we wanted to improve the ability of the model to represent groups of stimuli separated from each other. In other words, we wanted to split the data into two groups, learn a model on one group, then improve the model based the performance of the model on the other group. More specifically, we took Algorithm 3 and changed the check in its update step. While the model learned in each step is still based on the inside group, the variable selection is based on outside-group performance, balancing the ability of variables to fit the data well and the ability to explain variance in novel variables. The formulation of the problem would essentially be:

$$Z = \{Z_1, Z_2\} \text{ where } Z_i \text{ are the groups of data} \tag{4.12}$$

$$\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2\} \text{ where } \mathcal{X}_i \text{ are the variables} \tag{4.13}$$

$$\underset{\beta}{\text{minimize}} \ \|Z_1 - \beta\mathcal{X}_1\|^2 + \lambda|\beta|_1 - \eta\mathbf{R^2}(Z_2, \mathcal{X}_2, \beta) \tag{4.14}$$

$$\text{where } \mathbf{R^2}(Z, \mathcal{X}, \beta) = 1 - \frac{\|Z - \beta\mathcal{X}\|^2}{\|Z - \mathcal{Z}\|^2} \tag{4.15}$$

where $\mathcal{Z}$ is the mean of Z, η is the weighting for the fitting of the outside group, $\mathbf{R^2}$ is the coefficient of determination. If we adapt Algorithm 3 to solve this equation iteratively, we would get Algorithm 4.

Algorithm 4: Two-Group Fitting

Data: Voxel activation patterns for voxel i: Z_{1i}, Z_{2i};
Binary fragment presence matrices: $\mathcal{X}_1, \mathcal{X}_2$;
Result: Fragment coefficients: β_i
while *model not converged* **do**
  Solve $Z_{1i} = \beta_i \mathcal{X} + \epsilon$ for β_i using OLSQ;
  $R^2 \leftarrow \mathbf{R^2}(Z_{2i}, \mathcal{X}_2, \beta_i)$;
  for $f \leftarrow 1$ **to** *44* **do**
    $\hat{\beta}_{if} \leftarrow \beta_i$;
    $\hat{\beta}_{if}[f] \leftarrow 0$;
    Solve $Z_{1i} = \hat{\beta_{if}} \mathcal{X}_1 + \epsilon$ for $\hat{\beta}_{if}$ using OLSQ;
    $\hat{R^2}_f \leftarrow \mathbf{R^2}(Z_{2i}, \mathcal{X}_2, \beta_i)$;
  find f such that $R_d \leftarrow R^2 - \hat{R^2}_f$ is smallest;
  if $\tau_{lower} < R_d$ **then**
    $\beta_i \leftarrow \hat{\beta}_{if}$;
  else model has converged;

In our case, we have chosen eight of the most complex stimuli in our set as the "outside" group Z_{2i}. The difference of Algorithm 4 from Algorithm 3 is that the variable selection step is decided by the value of β that best explains the fit in the outside group, while the fit itself is still computed from the inside group. This process still computes the fitting based entirely on the inside group, but it eliminates components from the model based on their ability to explain not just data in the same group but also data in the outside group. If the outside group were also included in the model to begin with, the algorithm would easily overfit to the parameters of the entire data set. Its fit quality in training data is comparable to that of Algorithm 3, but it performs noticeably better on test data. Note that we don't actually have a parameter η in our algorithm. Similar to how our BSR algorithm eliminates the

need to choose a λ parameter while estimating a solution to the L_1 regularized regression problem, this algorithm eliminates the need for η. Instead, the role of those parameters is mimicked by our convergence parameter τ. This eliminates the need to find an optimal parameter to balance several constraints, instead focusing on a convergence threshold. Algorithm 4 will pick the same set of coefficients in the same order regardless of τ, the only thing τ changes is when the algorithm stops. Another implicit parameter is to choose how to separate the data into two groups. In practice, we've found that separating out about 10% of the data works well, and that 10% should ideally be the most nonlinearly represented samples in the data set. This has a double advantage, as such features often destabilize the linear regression part of the algorithm, and putting them in the R^2 clause allows them to affect the model's quality without making the linear solution account for outliers in its linear solver step. Our approach is to pick these samples by their variance.

Figure 4.9 demonstrates the fitting quality of our approach, and Figure 4.10 demonstrates a comparison of all approaches discusses so far, averages across subjects. Since the subjects have a different amount of voxels selected, their NRMSE values from their corresponding plots have been resampled via linear interpolation for the purposes of visualization.

It is worth noting that the final models selected by our algorithm have in the range of 3-6 coefficients selected per voxel. This is relevant for drawing parallels to monkey work. Pasupathy and Connor [49] have shown that neurons in a cluster code

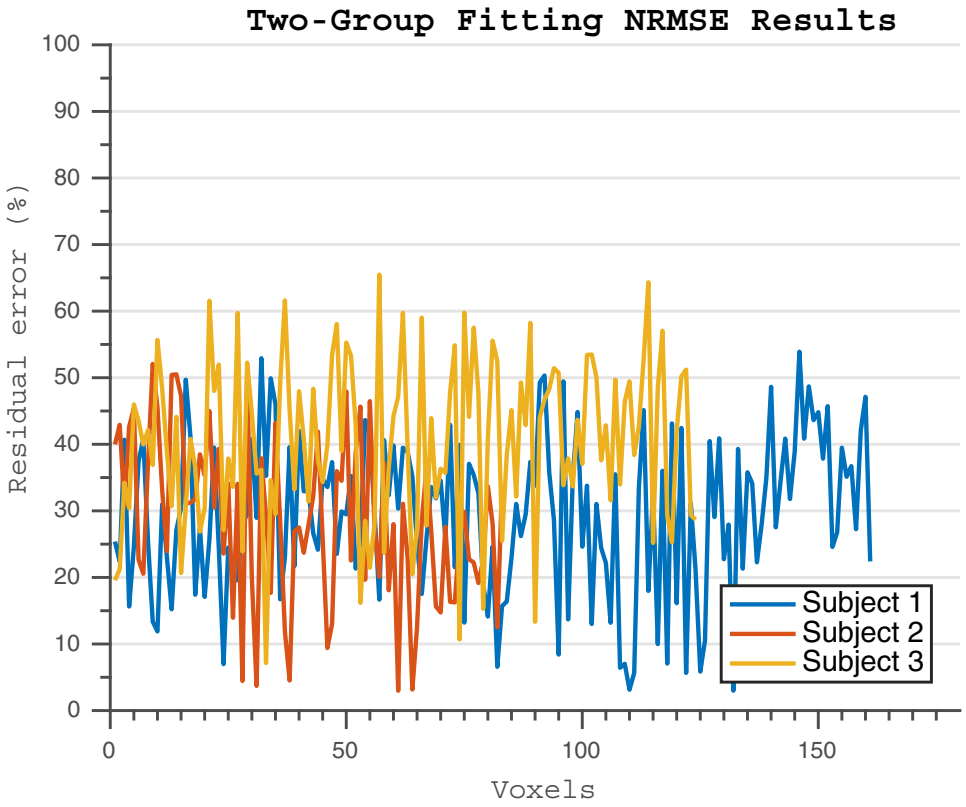


Figure 4.9: The residual error percentage according to NRMSE of applying the TGF approach in Algorithm 4

for around 2-4 specific curvature/orientation/position tunings. This comparable level of sparsity provides a neurological justification for the representative nature of our models. A sample operation of the algorithm can be seen in Figure 4.11.

4.4.2 Advantages of Two-Group Fitting

Our algorithm developed in Section 4.4.1 is designed to prevent overfitting by separating the data into two groups and fitting only on one of them while improving performance on the other. In this way, certain variables are completely left out

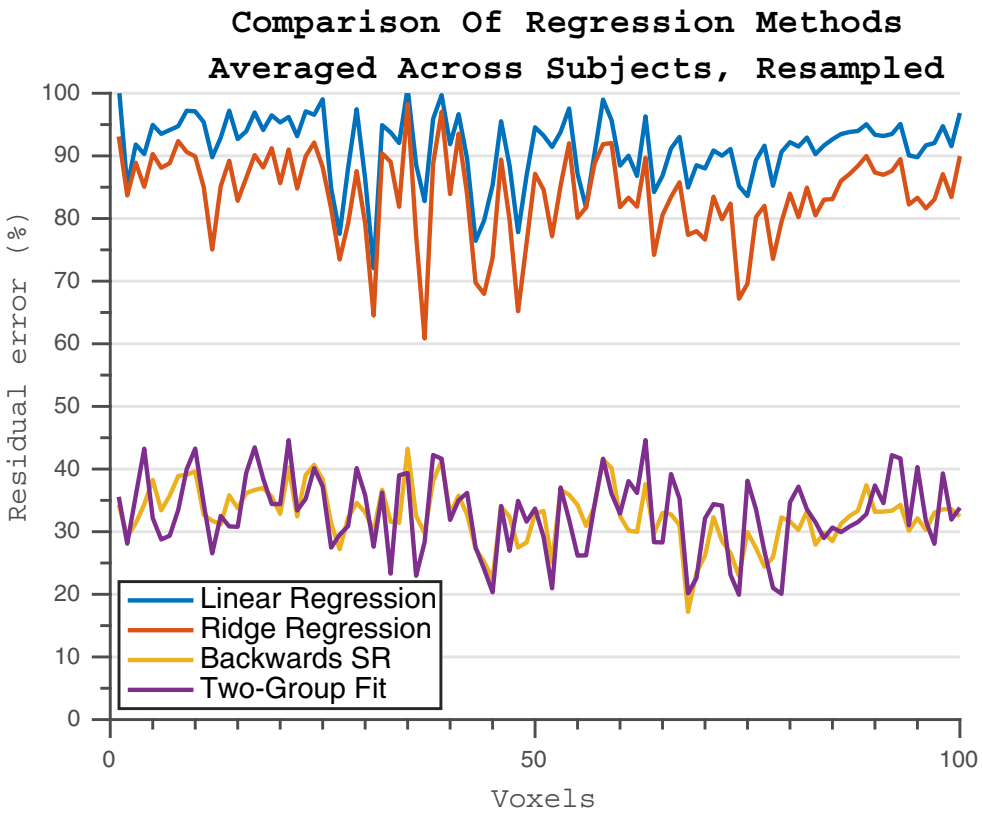


Figure 4.10: The residual error percentage according to NRMSE of applying each approach. The NRMSE curves are averaged across each subject via resampling, then displayed against each other. Note that our TGF algorithm performs comparably to Backwards Stepwise Regression.

of the fitting process, eliminating their influence into the regression, yet they still have a degree of control over the process via variable selection. We have tested our data with a publicly available data set separated into training and test groups to validate our claim. We have trained linear regression, Algorithm 3 and Algorithm 4 on the BlogFeedback Data Set by Buza [6], and used the models to predict each testing example. This data set has 280 features derived from blog comments, being mapped to the value of how many comments will be received in the next 24 hours.

Since this data set is also highly discontinuous, it is a good test of our algorithm's capabilities. We left out 20 features that had the highest variance among the 280. Figure 4.12 shows the performance of each algorithm measured by R^2 comparison between actual and predicted test data. We have both shown the actual comparisons and the comparison of each curve sorted among itself to demonstrate performance.

4.5 Validating The Model

A key aspect of our approach in this chapter is that we aren't trying to find the best fit for our data, we are trying to show that this parametrization of stimuli validates our hypothesis about the functionality of the LOC. As such, cross-validation is an essential tool for our aims. If we can train a model that, given two separate sets of data, can perform well on the left-out data, then we can say with confidence that our parametrization is valid. Towards this end, we will take several steps. The first step is to develop a model based on the ability to explain left-out data. The next step is to show that our parametrization is unique to the LOC, which we shall demonstrate by applying it to a brain area similar in function. By demonstrating that the parametrization is unique to LOC, we are able to make the claim that we are identifying functionality specific to that brain area instead of whole-brain functionality. Finally, we also need to show that other parametrizations don't work for LOC, so we will show that parametrizations established for the V1, whose functionality has

been mapped, don't explain our stimuli. With these, we will be able to confirm the thesis statement and the confirm the link between monkey IT models, showing that the LOC indeed codes for object structure in a manner that other brain areas don't.

4.5.1 Cross-Validation Analyses

A good test of a model's ability to represent the inherent patterns of the data and a safe guard against overfitting is cross-validation. Leave-N-Out cross-validation is a very common technique used for model evaluation. For the purposes of this work, we will use several values for N, specifically 1, 2 and 6. With leave-one-out and leave-six-out, we will assess the ability of the model to predict what left-out voxel patterns look like, and with leave-two-out we will assess how well can models trained on the rest of the data can decide between left-out stimuli pairs, as per Mitchell et al. [41].

4.5.1.1 Leave-One-Out Cross-Validation

For leave-one-out (LOO) cross-validation, we leave out one of 60 stimuli, train on the 59 left-in stimuli, then assess the quality of the predicted voxel pattern used NRMSE. This gives us 60 tests to evaluate, one for each stimulus. We measure the quality of the model by leaving each stimulus out, which gives us a set of 60 predicted values. We then compare the set of actual stimulus values to the predicted stimulus values for each voxel. We do this by computing the R^2 correlation coefficient between the predicted and actual value of the stimulus set for each voxel. We also performed

a randomization test to assess significance, randomly scrambling the stimulus labels 10,000 times for each LOO trial and compared the resulting R^2 value to $p = 0.05$.

4.5.1.2 Leave-Two-Out Cross-Validation

With leave-two-out (L2O) cross-validation, we leave out two stimuli, train on the 58 left-in stimuli, then assess the model's ability to make a binary choice between the actual values of the left-out stimuli and the predicted values using cosine similarity, repeating the approach in Section 4.3.1. The decision is made using Equation 4.7. There are $\binom{60}{2} = 1770$ L2O trials in this case. Since this is a binary choice, chance is at 50%, and through 10,000 randomized stimulus label assignments for each trial, we found the significance point ($p = 0.05$) to be at 62.3%. The overall accuracy of all binary choices across all L2O trials was 65.65%, which is comparable to the results of Mitchell et al. [41] and above significance.

Subject	Accuracy
1	65.40%
2	66.10%
3	65.45%
Average	65.65%
Significance	62.30%
Chance	50.00%

Table 4.1: The accuracy of choosing between the predicted values of two left-out stimuli.

4.5.1.3 Leave-Six-Out Cross-Validation

For leave-six-out (L6O) cross-validation, we leave out groups of six stimuli out of 60, train on the 54 left-in stimuli, then assess the quality of the predicted voxel patterns used NRMSE. The reason for leaving six stimuli out was have groups large enough to eliminate every appearance of certain compound fragments, and assess whether the models can work even if left-out stimuli have fragments within them that the left-in data set does not contain. In this case, we would normally have $\binom{60}{6} = 50063860$ trials, which would be quite intractable, especially with randomization tests using 10,000 different scramblings. Instead, we decided to intelligently leave out stimuli, leaving stimuli out in groups in a specific pattern. Our scheme was to leave out the same rotation of six stimuli, for each rotation and every possible combination of stimuli. We did not leave out different rotations of different stimuli, or all rotations of a single stimulus (since our model is fragment-based, it doesn't have a particular preference to every instance of a single stimulus being removed). This led to $4 \times \binom{15}{6} = 5 \times 5005 = 20020$ trials, which was a lot more manageable. Similar to the L1O case, we also performed a randomization test to assess significance, randomly scrambling the stimulus labels 10,000 times for each trial and compared the resulting R^2 value to $p = 0.05$.

4.5.2 Applying Our Model To The Visual Word-Form Area

Considering that our aim is to derive a model for the functionality of the LOC, the model should be unique to the LOC and not work in other, similar brain areas. Otherwise, the model derived could not be claimed to specifically describe the role of the LOC. Instead the derived parametrization would be a general-purpose model, which would go against the hypothesis of structural coding being the functional definition of the LOC. The role of the V1 is already well-established to be different. Another area similar in function to the LOC is the Visual Word-Form Area. If we were to show that our model, based on a parametrization of shapes using curvature, orientation and polar position, works on LOC voxels and does not work on VWFA voxels, it would strengthen our argument that this model is a depiction of the role of the LOC. Without this, it could be argued that this model does not uniquely identify the capabilities of LOC voxels, which would imply that VWFA also has a role in structural object coding. Such a result would be inconsistent with current understanding of the roles of these areas based on neuroscience research as per section 4.2.2, which could imply that our model is erroneous. Eliminating this possibility by showing that the model does not explain the behavior of VWFA voxels is critical. As such, we have identified voxels in the VWFA in Section 3.5. Using the same parameters, stimuli and design as per Section 4.4.1, we try to fit a model to the VWFA voxels. The results

of this approach can be seen in Figure 4.16. As it is quite clear, our parametrization performs very poorly in VWFA voxels, it does not go beyond significance levels.

4.5.3 Applying V1 Models To The LOC

Similar to using our parametrization on VWFA voxels to show that it is unique to the LOC, we also need to show that our parametrization does a better job of describing LOC voxel activity than models for other brain areas. Out of similar visual areas, only V1 has a significant body of work in terms of describing its functionality. As discussed in Section 4.2, it is commonly accepted that V1 voxels perform a Gabor Wavelet decomposition on the visual signal. Specifically, Naselaris et al. [44] have characterized the activity of V1 voxels using a Gabor Wavelet Pyramid to parametrize images, then learned a mapping between coefficients of the wavelets and the fMRI images. We applied the same approach to our stimulus images, breaking them down with a GWP decomposition following the work of Naselaris et al., and applied our model to learn a mapping. This model performed poorly in comparison to the fragment-based model that we have proposed, not going beyond significance levels. These results can be seen in Figure 4.17.

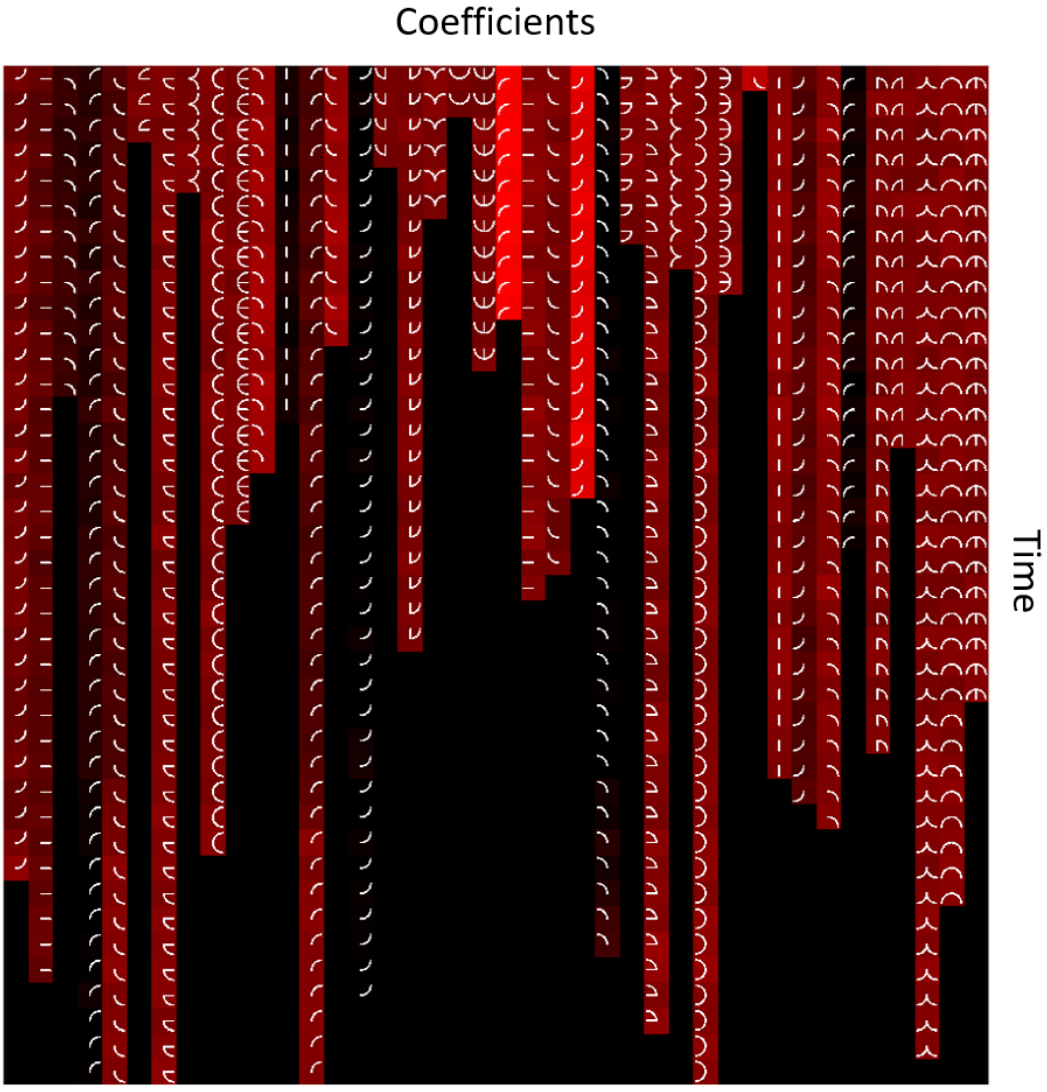
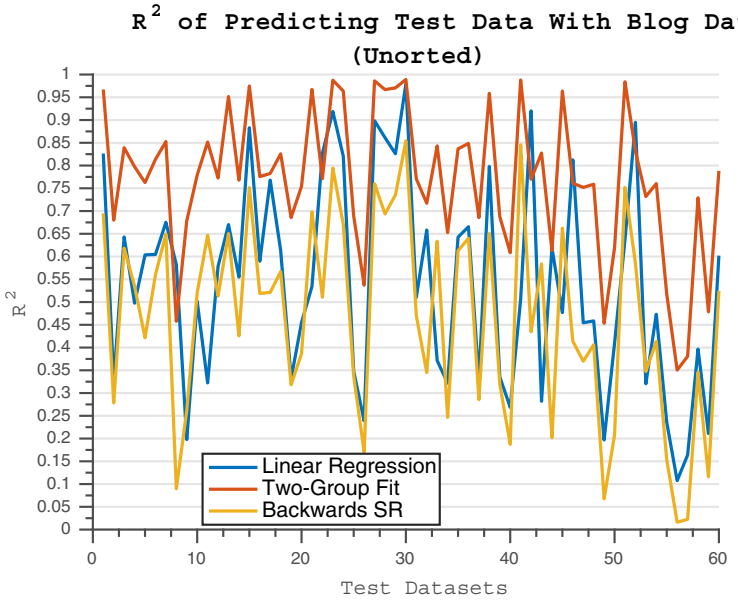
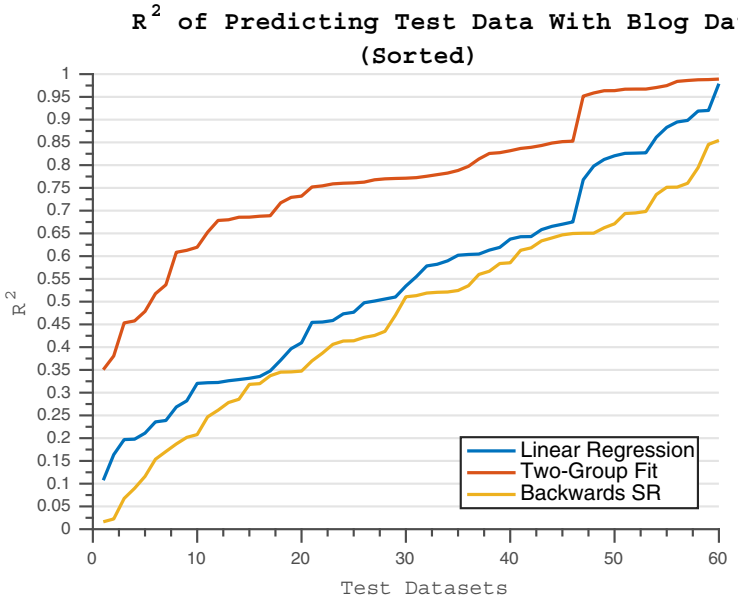


Figure 4.11: A demonstration of the Two-Group Fitting algorithm for a sample voxel. Every row represents the weights assigned to each fragment at a particular iteration, with a gradient from red to black denoting normalized values between 1 and 0. At each step, one fragment is removed from the model by setting its weight to zero.



(a) The correlation coefficient between the predicted and actual value of the blog testing set according, compared across linear regression, backwards stepwise regression and two-group fitting.



(b) The correlation coefficient between the predicted and actual value of the blog testing set according, compared across linear regression, backwards stepwise regression and two-group fitting. The curves are each sorted within themselves to demonstrate performance differential between approaches.

Figure 4.12: The correlation coefficients between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for each subject.

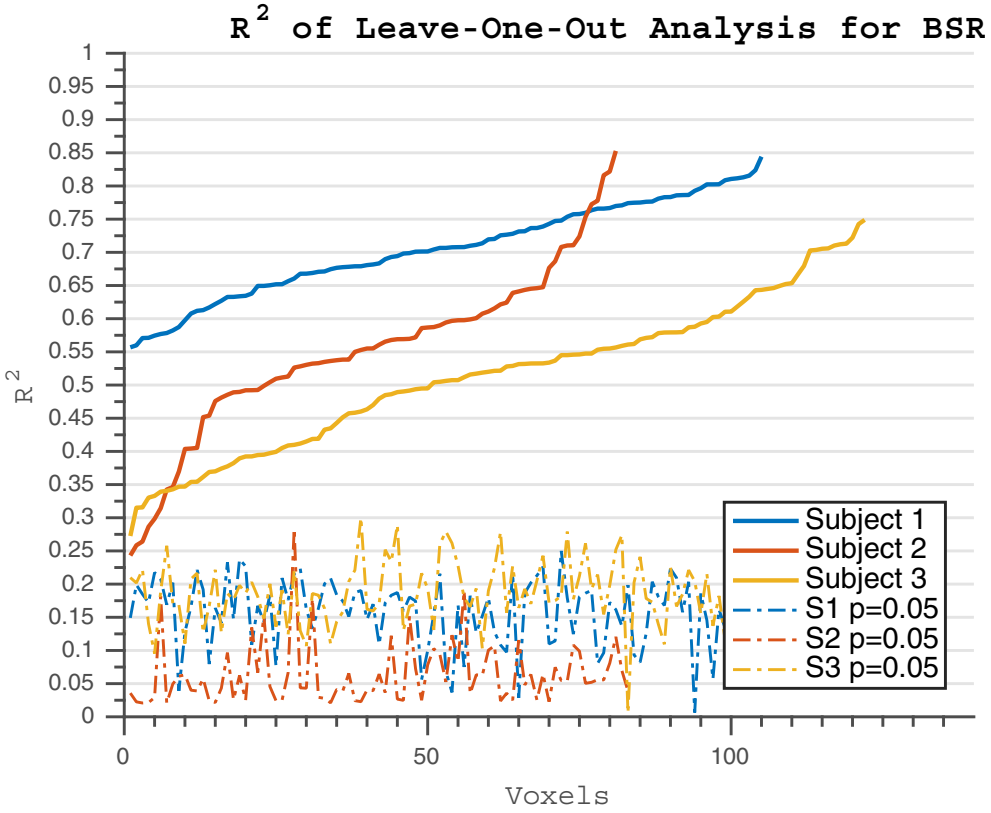
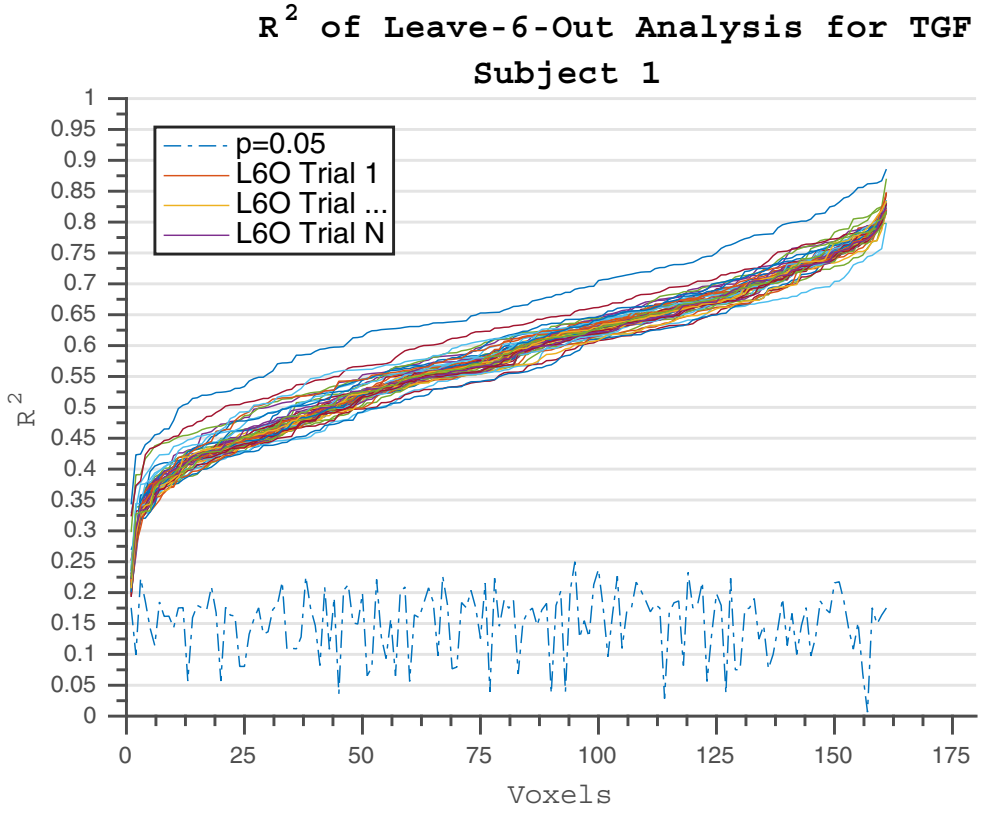
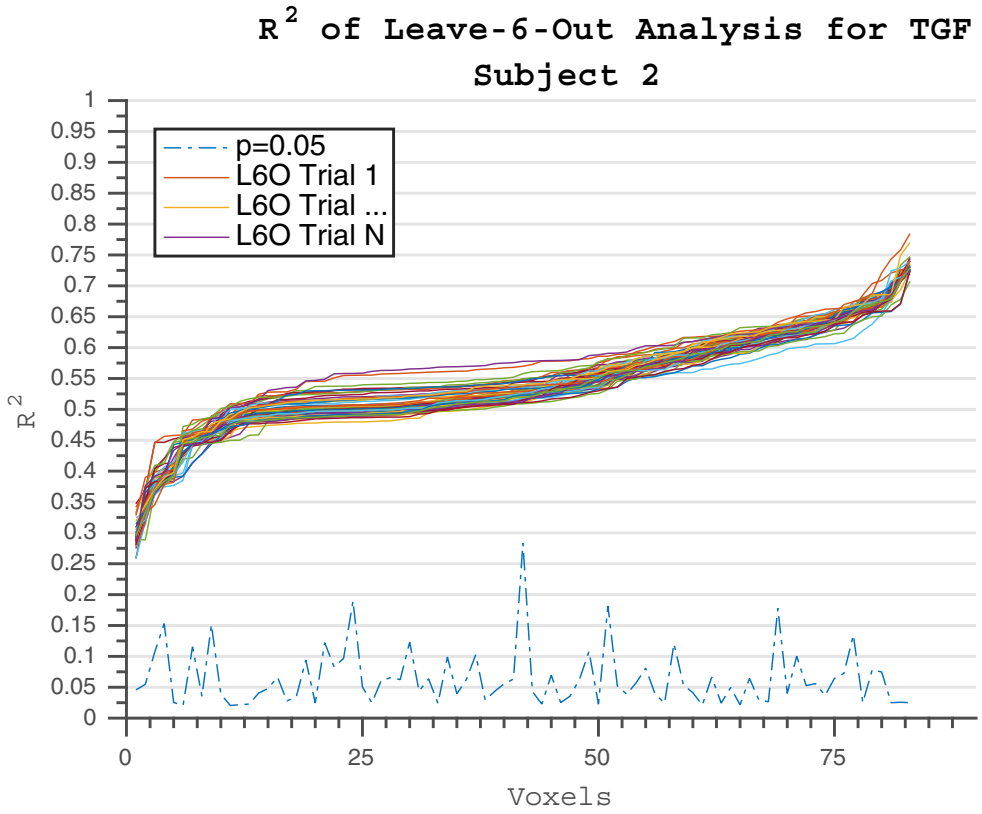


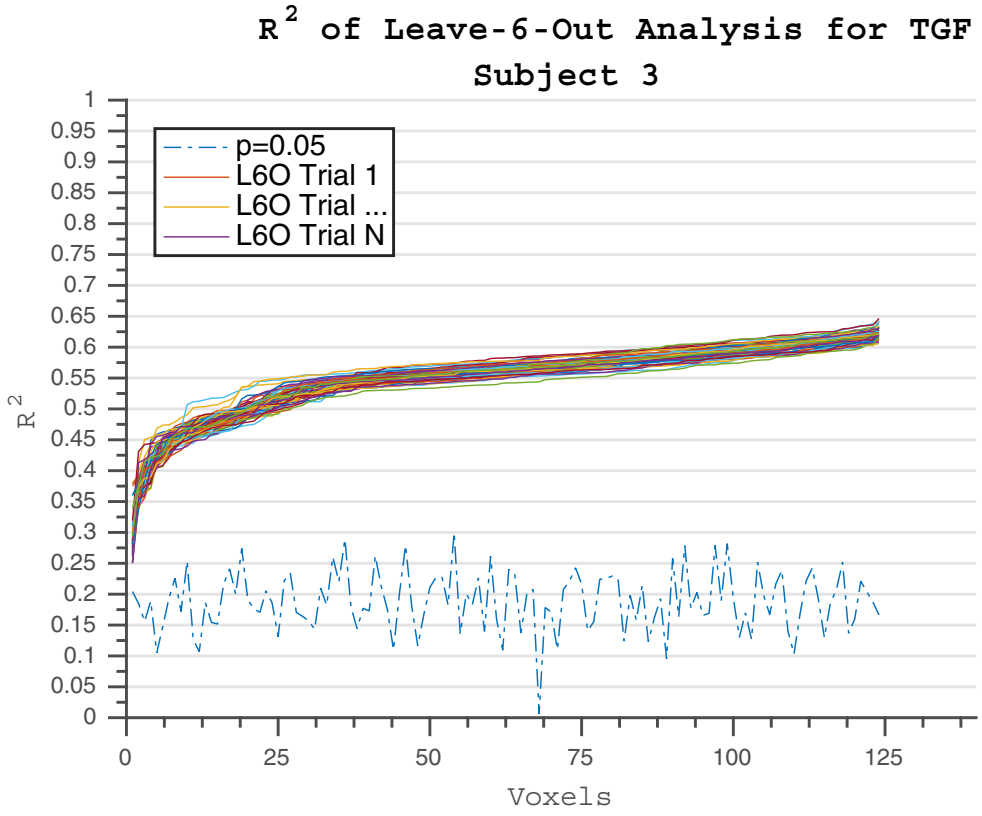
Figure 4.13: The correlation coefficient between the predicted and actual value of the stimulus set according to Leave-One-Out analyses for each voxel.



(a) The correlation coefficient between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for Subject 1.



(b) The correlation coefficient between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for Subject 1.



(c) The correlation coefficient between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for Subject 1.

Figure 4.14: The correlation coefficients between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for each subject.

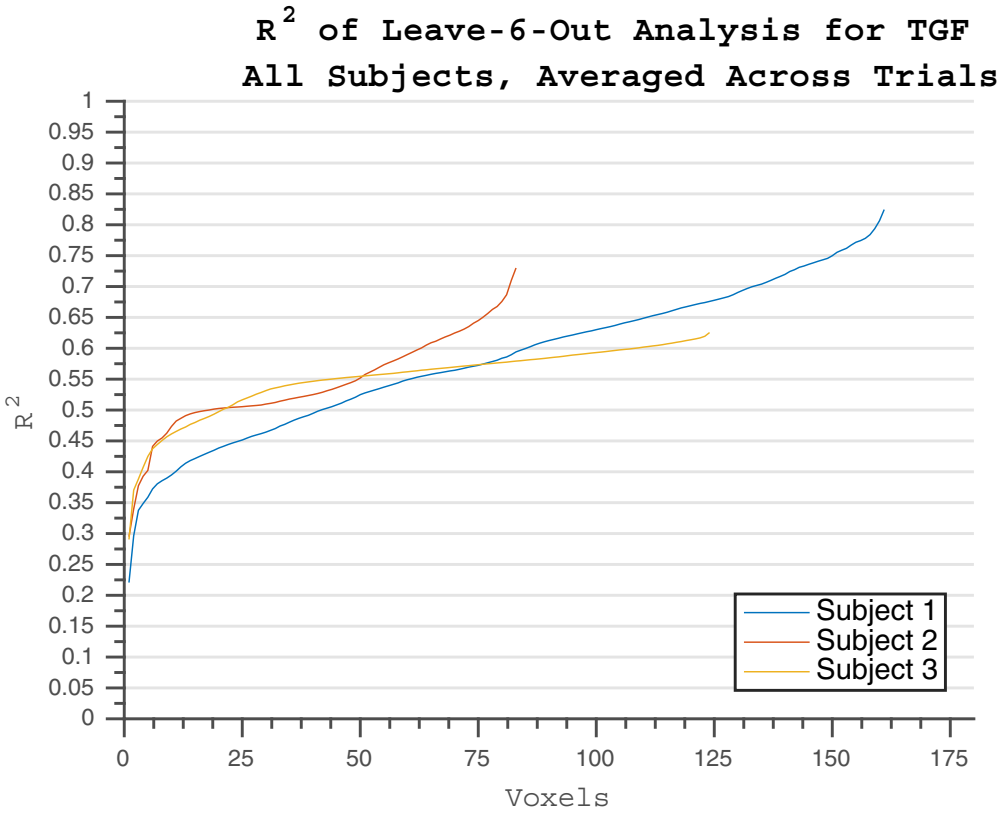


Figure 4.15: The correlation coefficient between the predicted and actual value of the stimulus set averaged across all L6O analyses for each voxel for each subject.

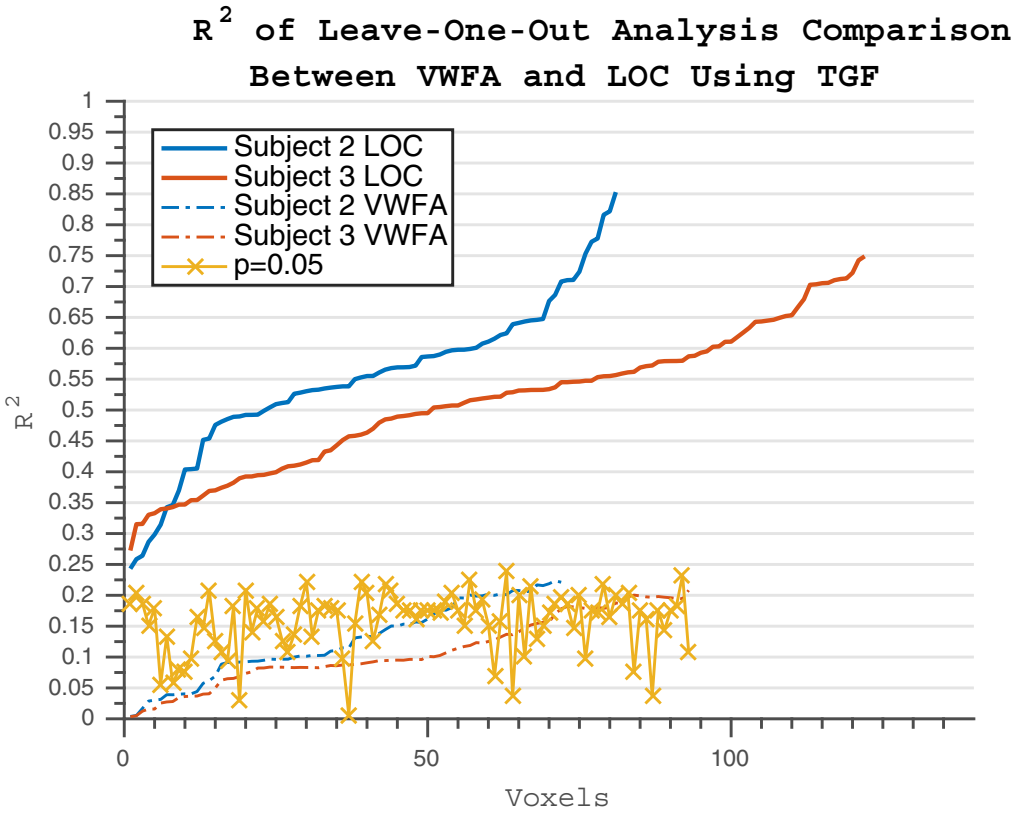


Figure 4.16: Performance of the structural parametrization on VWFA voxels compared to the LOC voxels.

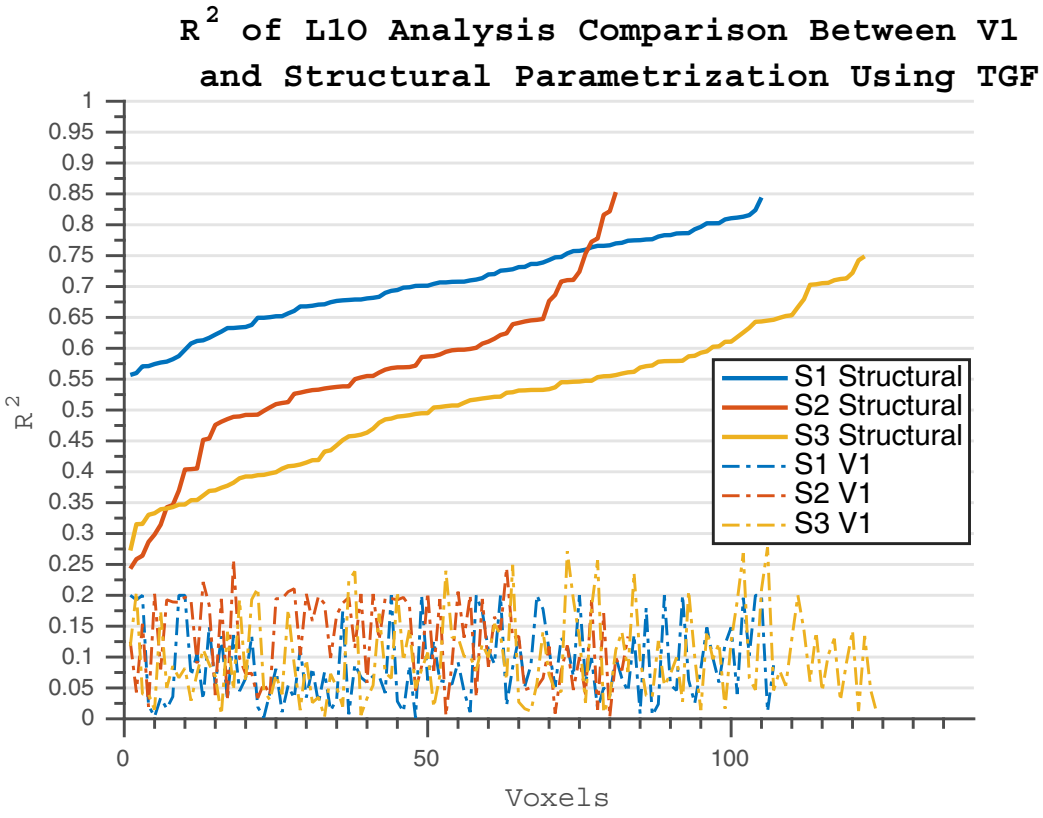


Figure 4.17: Performance of the V1-parametrization compared to the structural parametrization on LOC voxels.

4.6 Conclusions

The goal of this chapter, and ultimately of this work is to provide evidence for structural coding of object shape in the lateral occipital complex voxels. We designed a set of medial axis stimuli and an experiment structure to this end in Chapter 2. In the following chapter, we extracted voxel information to be used to this end from raw fMRI data. The specific objective of this chapter was to derive a method of parametrizing the stimuli in a way analogous to how inferotemporal cortex neurons in monkeys code for similar shapes, then validate that LOC voxels also respond to visual stimuli in a manner consistent with this parametrization using a computational model. After an initial experiment with a complex stimulus set failed, we settled on the final set of stimuli, each being composed of two to three individual medial axis fragments. We divided this parametrization into four quadrants, and created a binary vector of presences for each fragment in each quadrant. With three unique fragments representing different curvature values, used at differing angles and in each quadrant, we were able to use these parts to approximate a parametrization that spanned curvature, orientation and polar position values.

Following the establishment of the parametrization, we set out to show that this parametrization could be used to develop a computational model for how LOC voxels respond to visual stimuli. We tried several models, and eventually developed our own algorithm that separates the data into two groups and learns a linear regression on one group that eliminates terms from its model based on how it performs on the

other group. By using this approach, we were able to improve performance of our algorithm compared to existing methods. Since we were concerned about overfitting, we developed a series of leave-N-out cross-validation tests, and demonstrated the robustness of our model.

To further cement that this parametrization is specific to the LOC, we conducted two further tests. We applied the same parametrization using our model to voxels found within the visual word-form area and demonstrated that the behavior of voxels in this area cannot be adequately explained by such a parametrization. This strengthened our claim that this structural parametrization is particular to the behavior of the LOC. The second test was to apply parametrizations derived from models of early visual areas to the voxels found in LOC. This was done in order to show that the behavior of LOC is distinct from early visual areas, and our parametrization is more appropriate for describing LOC voxels than other approaches. The results validated our hypothesis in this case as well.

These results are significant, as they are the first results demonstrating the human LOC coding for object shapes in a parametric fashion. While prior studies have shown that LOC voxels respond to object images, our work is the first to explore how these voxels respond to object images based on characteristics of those images. These preliminary findings could lead to further exploration of the LOC and mapping of its functionality using more in-depth studies that utilize higher resolution scanners, span more subjects and wider parametrizations. Additionally, further neuronal recording

studies on monkeys can be linked to LOC experiments, bridging yet another link between the species by solidifying the homology between these areas.

Furthermore, the two-group fitting algorithm we proposed can be used as a general machine learning technique for fitting data (in our case brain voxels) that is generated by degenerate input. This is an iterative linear regression algorithm that solves for a sparse set of coefficients without requiring a parameter to balance sparsity against fit quality. We have demonstrated the viability of the algorithm on our own fMRI data and on a publicly available dataset.

Chapter 5

Reconstruction Of Stimuli Using fMRI Images

In the previous chapters, we designed and performed experiments, then analyzed the resulting data with the goal of showing evidence for structural coding of medial axis shape in the Lateral Occipital Complex. More specifically, in Chapter 4 we derived a parametrization and designed an algorithm to model the activations of LOC voxels. Using this model, we were able to describe the behavior of the subjects' brains as a result of visual object stimuli being presented as a function of the structural fragments present within the stimuli better than any other known model, and with statistical significance. To complete the picture, the next step would be to follow the parametrization in the other direction, namely using voxel activations to predict the stimuli being viewed by the subjects based on the structural fragments

they're composed of. In this chapter, we will address this problem. We will develop a reconstruction algorithm that takes into account the nature of the data and parametrization.

5.1 Overview

In previous chapters, we have designed and conducted an experiment to create a model for how voxels in the Lateral Occipital Complex respond to images being shown to the subject, based on a parametrization representing curvature, orientation and polar position. We have developed a computational model that sought to explain the behavior of voxel activity using a fragment-based parametrization of our stimuli. However, that approach only validates our hypothesis in one direction. Being able to predict brain activity from fragments validates part of our assumption. The other direction is also important, namely the ability to predict stimuli based on voxel activations. In this chapter we develop a novel computational model for reconstructing the stimuli being viewed by a subject based on the activity of their LOC voxels. It is worth noting that in this chapter we are not trying to validate the statistical model we fit onto the voxels. Our aim in this chapter is to validate our hypothesis that states the LOC codes for visual stimuli based on orientation, curvature and polar position. As such, this chapter will not focus on models developed in Chapter 4. Instead it will be using a hypothesis-driven approach to predict likely candidates for fragments

and use those to reconstruct the stimuli. The entirety of this chapter will be devoted to discussing and validating our novel contribution to reconstructing stimulus images from LOC voxel activations.

In Section 5.3, we will describe the procedure used to reconstruct images of left-out stimuli using our novel approach. Section 5.3.1 will discuss the approach used to calculate the presence probability of each fragment in a stimulus being predicted. Section 5.3.2 will describe the major contribution of this chapter, the algorithm with which candidate fragments are selected based on the ranking derived from their presence probabilities. Section 5.4 will discuss the results of the reconstruction process and their validation. Section 5.5 will discuss potential future applications of these findings. Finally, Section 5.6 will explore the implications of the novel algorithm in this chapter, its potential uses, lessons learned from its development and future work.

5.2 Related Work

In this chapter, we will be providing more results supporting our argument that the LOC codes for images based on structural fragments. As such, it is important to return some of the work visited in Section 4.2. We will address the relevant works, and contextualize them within the scope of this chapter.

5.2.1 Neurophysiology Studies On Object Structure

As explained in Chapter 2, the primary motivations for our work come from monkey studies concerned with describing how the inferotemporal cortex codes for object structure. Since deriving a similar parametrization for LOC is the primary goal of this chapter, it is important to reiterate some of the background work here. Since the inferotemporal cortex is considered the homolog of the LOC in monkeys [37], the coding we will propose for the LOC is dependent on understanding how IT neurons code for object structure. Thus, this section will be reiterating some of the discussion from Section 4.2. However, some of the works cited will be contextualized from a reconstruction perspective instead.

Studies as early as Pribram and Barry [51] and Wilson [64] have shown via ablation of inferior temporal regions in monkeys that these areas hold a visual function. Dubner and Zeki [17] have shown that the IT receives inputs from the primary visual area (V1) relayed by V2 and V4. Further studies by Dean [15], Gross et al. [20] and Mishkin [40] have shown via ablation that the IT specifically affects visual discrimination or recognition of objects. Perhaps the most important finding that establishes the groundwork for this work is by Tanaka et al. [58], demonstrating a specific coding for objects of differing structural parameters in IT.

Recent work in monkey IT has shown evidence of parametrization for specific char-

acteristics of objects based on structural components. Parametric coding of object shapes in monkeys has been shown in Pasupathy and Connor [49]. Neuron recodings have enabled the authors to reconstruct stimuli being viewed, and this work will also follow a similar design, but with fMRI voxels instead. Pasupathy et al. have fit Gaussians to the tuning responses of neurons to predict the stimulus eliciting the neural spike pattern. Since fMRI is less direct than neuronal recordings, and since our stimuli are discrete instead of continuous (due to fMRI limitations discussed in Chapter 2), their approach is inapplicable as-is to ours. However, we will imitate their method in a manner, as we will be predicting a presence probability for each of our discrete fragments given a novel stimulus. Our prediction method is inspired by their approach.

The IT has been shown to code for combinations of straight and curved shape fragments by Brincat and Connor [4]. The authors have used a model to describe the response to stimuli as parametrized by such fragments, which is the inspiration for our approach. Our work also uses straight lines and curved lines. Their work has line fragments joining at edges to create closed shapes, whereas our work has line fragments going out from the fixation center, and closed shapes will not be allowed by our reconstruction algorithm. The reasoning for this is based on Brincat et al.'s work, as they have demonstrated that continuous shapes are represented differently from discontinuous shapes, and we did not have enough experiment time to represent both shape types within our stimulus set. Closed shapes also create curvature ambiguities,

whereas medial axis lines are unambiguous in terms of their curvature. Hence the most directly relevant work to our is Hung et al. [25], who have used medial axis stimuli to demonstrate that the tuning function of IT neurons can be modeled using fragments of such stimuli. As a result, our work has uses medial axis stimuli as described in Chapter 2.

Yamane et al. [65] have developed a coding for three-dimensional object shape and spatial configurations based on IT neuron responses. Due to fMRI limitations a parametrization with the complexity of Yamane et al.'s is out of reach (see Chapter 2 for discussion of fMRI limitations), we have taken into account the spatial variations of fragments as a factor in our parametrization, specifically polar positions of fragments around the visual fixation point.

5.2.2 fMRI Studies On The Visual Cortex

5.2.2.1 Early Visual Areas

The fMRI studies in visual cortex are directly relevant to our work in this chapter, as many of them focus on reconstructing either stimuli being viewed or classifying images and predicting a class for the image being viewed. Our approach leans more towards the latter, as our stimulus set is discrete. However, our approach takes inspiration from both, and it is important to reiterate some of these studies.

Several recent fMRI studies have used Gabor wavelet based approaches to

parametrize the functionality of voxels in early visual areas. As mentioned earlier, a large amount of fMRI work is based on classifying brain activity. Unconstrained reconstruction is a more challenging task. Miyawaki et al. [42] have reconstructed binary image patches using such an approach. Looking at early visual areas and parametrizing image stimuli with Gabor wavelets, they fit a model to voxel activations to predict images associated with them. Miyawaki et al. have used letter-like stimuli similar to our medial axis shapes, however their approach is dependent on voxels within the primary visual cortex (V1), and the functionality of that area is very well-understood due to neuron recording studies in animals, and the high correlation of animal V1 to human V1. Due to LOC being a higher level visual area, such correlations are not as obvious, and voxel activity is significantly harder to record. As such, we were not able to use a parametric reconstruction. The original stimulus set we described in 2.4 was more suited towards such an approach, however we were unable to obtain enough signal from this area to achieve such a reconstruction.

More relevant to our approach is the research of Jack Gallant and collaborators. Their work, specifically Kay et al. [28], Kay and Gallant [27], Naselaris et al. [44], Naselaris et al. [45] and Vu et al. [63] has used fMRI of early visual areas to reconstruct natural images being viewed by subjects. These works have used Gabor wavelet decompositions to parametrize images, then use a bag-of-words inspired model to reconstruct images being viewed by the subject. Since their approach is classification-based, their approach was a direct inspiration for ours.

5.3 The Reconstruction Process

In this section, we will provide an overview of the procedure used for reconstructing stimuli from voxel activations. In the following sections, we will explain in detail what each of these steps entail. In Section 4.3.2, we have demonstrated that the activations caused by stimuli can be distinguished to some extent depending on which fragments they contain. The first step of our reconstruction is based on this, calculating the probability of the presence of each fragment. The second step is picking candidate fragments and the final step is verifying the results.

5.3.1 Computing Candidate Fragments

The first step of the reconstruction process involves calculating a presence probability for each fragment given a novel stimulus. The existing data is separated into two clusters per fragment, one cluster containing voxel activations of stimuli where the fragment is present, and the other containing voxel activations of stimuli where the fragment is absent. The activation value of the novel stimulus is projected onto the line between the means of both clusters for each fragment, then the probability of the fragment's presence is computed as the position of that projection on the line, as seen in Figure 5.1 Repeating this process for every fragment gives us a vector of probabilities for the novel stimulus.

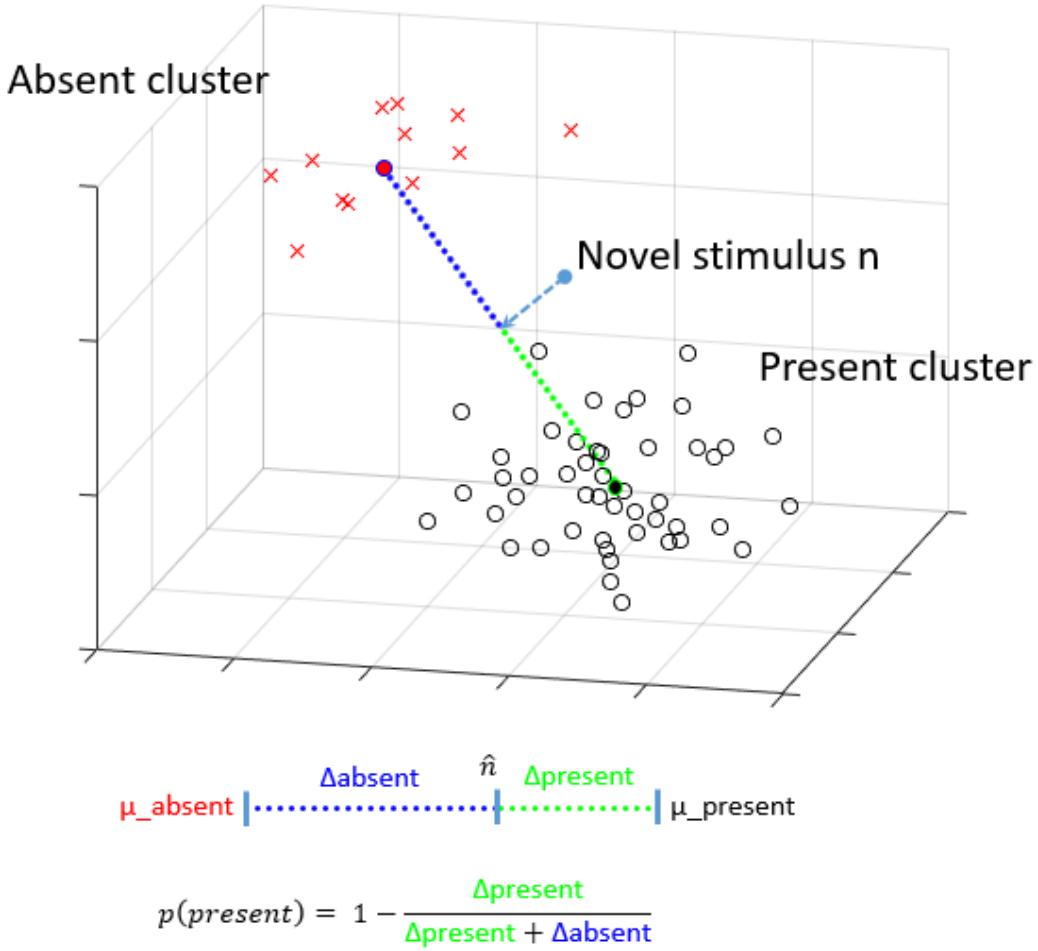


Figure 5.1: Projecting the novel stimulus onto the line between the means of the present and absent clusters.

5.3.2 Rank-Based Fragment Selection

Following the computation of the presence probability of each fragment in Section 5.3.1, we must make a decision and choose fragments from the list of probabilities. This process is not trivial. As such, we have developed a procedure for picking the correct fragments given a list of probabilities. As mentioned in Sections 2.5 and4.4, our

parametrization works on an assumption of quadrants, with stimuli being constructed of four quadrants containing 0-3 fragments each. Given that we have integrated an "empty fragment" for each quadrant and compound fragments that consist of multiple individual fragments, we can reduce that assumption to each quadrant containing exactly one fragment. This simplifies the task of picking candidates for each quadrant, as having to decide between picking a single fragment versus multiple fragments would complicate the algorithm immensely. This is yet another benefit of choosing a "compound fragments using a linear model" approach as per Section 4.4.

Once the presence probabilities are calculated, the fragments are ranked in descending order of probability. Ideally, we could pick the highest ranked fragment from each quadrant and that would give us the final shape. In practice, this doesn't work as well, due to noise other imperfections in the data. Picking the highest ranked fragment from each quadrant can lead to invalid shapes, for example shapes that have conflicting fragments, closed loops (which aren't allowed by our rules) or other such "illegal" configurations. To combat this, we iterate over every possible choice for a quadrant, going from the highest ranked to the lowest ranked.

Given the fMRI responses to the stimulus set Y, for fragment f_i, the responses can be separated into two clusters. One for responses to stimuli that contain the fragment, and one for stimuli that don't contain the fragment. The means of these clusters will be denoted as μ_{pi}, μ_{ai} where p denotes the cluster where the fragment is present, and a denotes the cluster where it is absent. The vector between these

two cluster centers is given by $M_i = \mu_{pi} - \mu_{ai}$. Given a novel response pattern $\hat{x}$, the probability of the stimulus that generated that response containing fragment i can be denoted:

$$p(f_i \| \hat{x}, Y) = \frac{M_i \cdot \mathcal{X}_i}{\|\mathcal{X}_i\|} \tag{5.1}$$

where $\mathcal{X}_i = x - \mu_{ai}$. For the sake of convenience, we will denote $p(f_i \| \hat{x}, Y) = \rho(f_i)$. As mentioned previously, our parametrization separates fragments into four quadrants. We will denote fragments belonging to these clusters as F_j where $j = 1 \cdots 4$ and $F_1 = \{f_1 \cdots f_{11}\},\ F_2 = \{f_{12} \cdots f_{22},\ F_3 = \{f_{23} \cdots f_{33},\ F_4 = \{f_{34} \cdots f_{44}\}$. Given a novel stimulus, our goal then is to solve the following integer programming problem:

$$\begin{gathered} \max_{a,b,c,d} \rho(f_a) + \rho(f_b) + \rho(f_c) + \rho(f_d) \\ \textbf{subject to } conflicts(f_a, f_b, f_c, f_d) = 0 \\ \textbf{where } f_a \in F_1,\ f_b \in F_2,\ f_c \in F_3,\ f_d \in F_4 \end{gathered} \tag{5.2}$$

where $conflicts()$ is a function that returns the number of conflicts that would occur if a stimulus was generated from the chosen fragments. This function can be expressed as sets of linear equations with a variable for each fragment. We have developed Algorithm 5 to solve the problem given in Equation 5.2.

Algorithm 5 was used to reconstruct stimuli from the corresponding activation patterns. The functions called within are: $Project\_To\_Line(N, \overrightarrow{AP})$, which projects the point N onto the line defined by the vector $\hat{AP}$. The mean-projection procedure is

Algorithm 5: Rank-based Reconstruction

Data: Voxel activation pattern Z;
Binary fragment presence vector for stimulus i: X_i;
Novel stimulus activation pattern: N;
Fragment list: F;
Result: List of fragments predicted to be present in stimulus i
$P \leftarrow 0_{44\times 1}$;
foreach *fragment* f **do**
  $A \leftarrow Z_i$ where stimulus i doesn't contain f;
  $P \leftarrow Z_i$ where stimulus i contains f;
  $\mu_a \leftarrow mean(A)$;
  $\mu_p \leftarrow mean(P)$;
  $\overrightarrow{AP} \leftarrow \mu_a - \mu_p$;
  $n \leftarrow Project\_To\_Line(N, \overrightarrow{AP})$;
  $P[f] \leftarrow (1 - \frac{n}{\mu_a + \mu_p})$;
$[P, order] \leftarrow Descending\_Sort(P)$;
$selection\_size \leftarrow 4$;
$conflicts \leftarrow True$;
while *conflicts* **do**
  $c\_list \leftarrow order[1 \ \ldots \ selection\_size]$;
  $candidates \leftarrow Enumerate\_NChooseK(c\_list, 4)$;
  for $i \leftarrow 1$ **to** $\binom{selection\_size}{4}$ **do**
    $selection \leftarrow candidates[i]$;
    $conflicts \leftarrow Check\_Conflicts(F[selection])$;
    if $\neg conflicts$ **then return** $selection$;
  $selection\_size \leftarrow selection\_size + 1$;

demonstrated in Fig. 5.1. $Descending\_Sort(P)$ sorts the elements of P in descending order and returns the sorted list along with a vector of indices describing the sorting arrangement, $Enumerate\_NChooseK(c\_list, 4)$ returns a list of all possible combinations of choosing 4 elements from c_list, and $Check\_Conflicts(P[selection])$ looks in the 4-long vector $P([selection])$ and returns a boolean indicating whether there are conflicts in constructing a stimulus from those fragments. The conflict checking was done by having a list of unallowed configurations and doing a check against them.

Once the algorithm provides the list of fragments to be used to reconstruct a stimulus, the images of those fragments are composited in order to create the reconstructed stimulus image. For examples see the next section.

5.4 Results

For testing and validation of our results, we used the same leave-six-out scheme used in Section 4.5.1.3. The difference in this section is that the stimulus parameters are the values being predicted and the fMRI activation patterns are the input into the algorithm. Also, we only used 10 sets of leave-6-out groups, so that each stimulus is only left out once. We picked leave-out groups to ensure that left-out stimuli had minimal to no overlap with left-in stimuli. Using this approach, we applied Algorithm 5 to each left-out stimulus. As mentioned in Section 5.3.1, the end result of the algorithm is, for each stimulus, a list of fragment probabilities for each quadrant.

While the reconstructed stimuli can be assessed visually, another way to assess the quality of the reconstruction is looking at the rank sum. For each quadrant, we rank the fragments according to their presence probability in descending order. Given that we already know the correct fragment for each quadrant, we can find where the correct fragment is ranked within this sorted list. For each stimulus, we assign its rank-sum to be the sum of the ranks of the correct fragments across all quadrants. This means that a perfectly reconstructed stimulus would have rank-sum of 4, and a stimulus that has the correct fragments ranked lowest would have a rank sum of 44 (given 11 fragments per quadrant, including the blank fragment). It is worth noting that the list of possible configurations is larger than 44, in fact it is 11^4. However, the rank-sum approach acts more like a distance-based metric, as all configurations that are off by one fragment are equally ranked. This would mean that predicting all quadrants incorrectly would have the same score as predicting 3 quadrants correctly but assigning a very-low-rank prediction to the fourth. The results for this analysis can be seen in Figure 5.2. As such, an alternative approach to visualizing these results can provide more information. The number of correctly predicted quadrants for each stimulus is thus displayed in Figure 5.3. As a counterpart to the rank-sum analysis, this measure does not represent how poorly mispredicted a given fragment is. Due to this, a reconstruction that predicts the correct fragment as the second-most-likely candidate would receive the same score as one that predicts the same fragment as the least-likely candidate. To get a complete understanding of these

results, one must view both metrics in tandem. In our case, most incorrect predictions weren't too far removed from the correct ranks, so these two metrics have provided similar results. We have also measured the significance of the reconstruction results against 10,000 randomized trials. As can be seen in Figure 5.4, our reconstructions where at least two quadrants are correctly predicted are also statistically significant. Alternatively, reconstructions with a rank sum of over 38 also all have statistical significance. It is worth noting that the reconstruction prefers more complex stimuli, and it performs poorly in stimuli where there are two blank quadrants. We presume this is due to the fact that it is difficult to differentiate a lack of fMRI response caused by blank quadrants from fMRI noise. In general, we have found that responses to presence are much more strongly driven than responses to absence. Figure 5.5 shows a comparison between a reconstructed stimulus set and the actual stimuli. This comparison shows the tendency of the representation to favor curved fragments and avoid blank fragments. Visual inspection reveals that even stimuli that have incorrect predictions generally preserve the overall structure of the stimulus for the most part.

An additional result to consider is the rate at which fragments are mis-predicted as each other. To this extent we created a "confusion matrix", counting how many times across the stimulus set has a particular fragment been incorrectly labeled as another given fragment. Figure 5.6 displays this matrix.

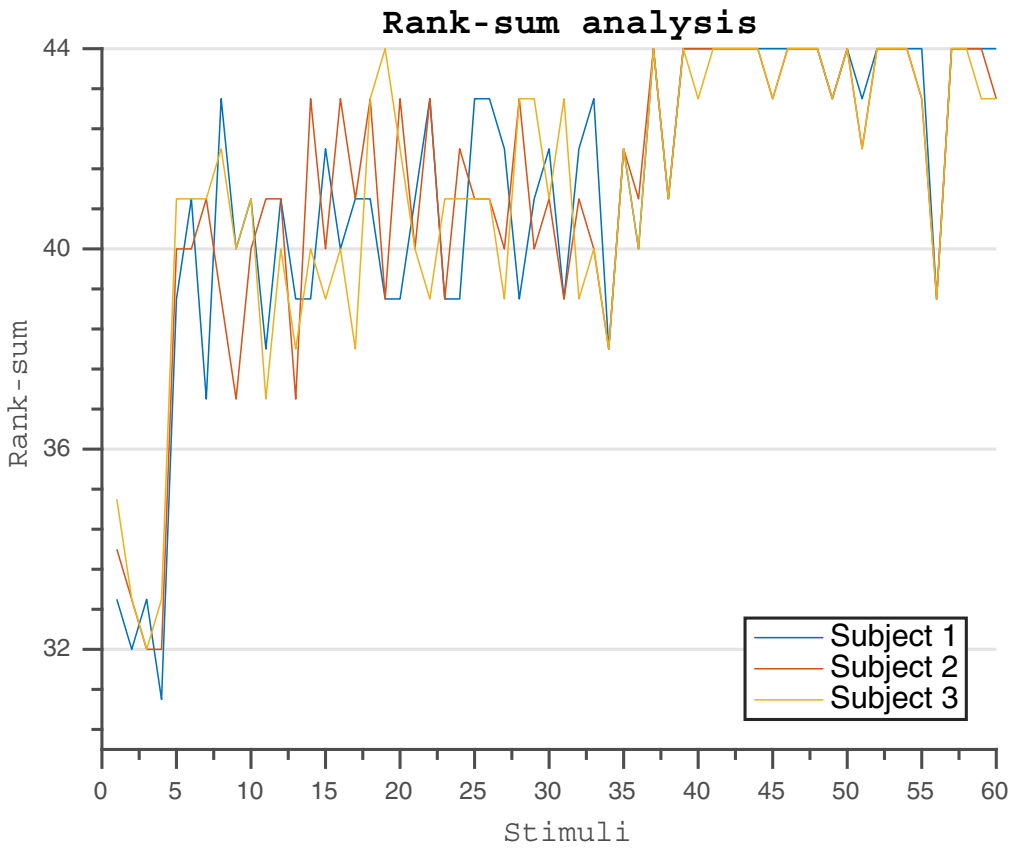


Figure 5.2: The rank sum of each stimulus for each subject. Perfect reconstruction would have a rank of 44.

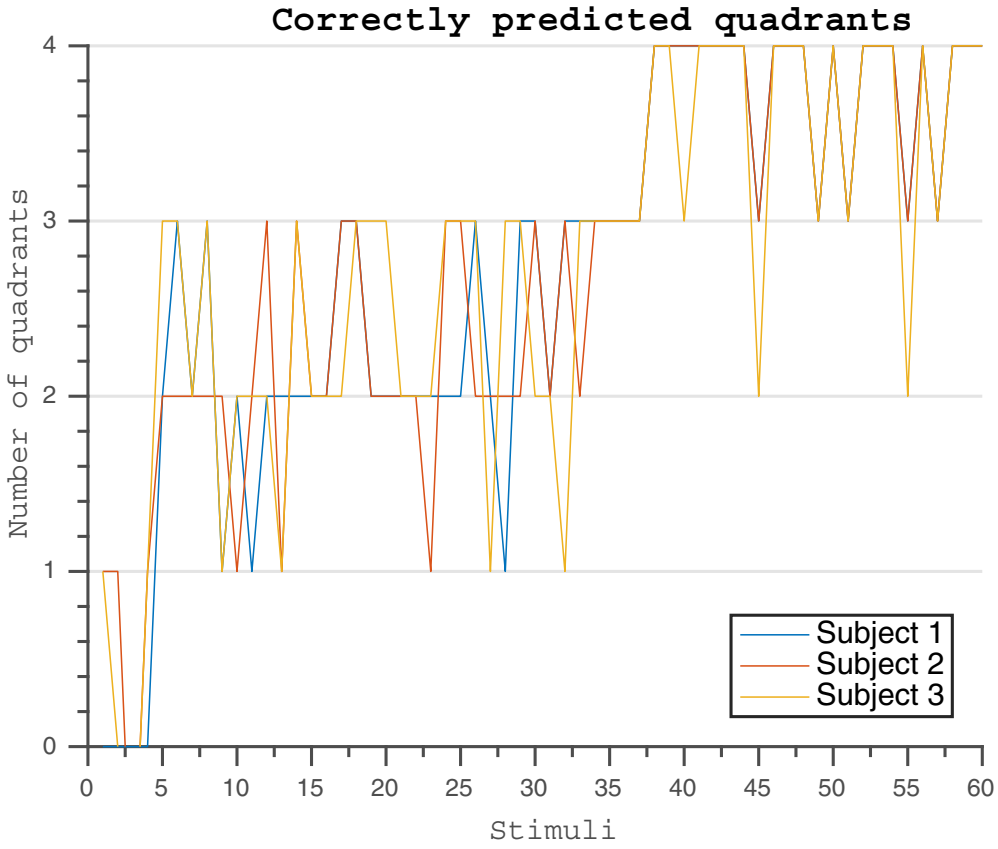


Figure 5.3: The number of correctly predicted fragments for each stimulus.

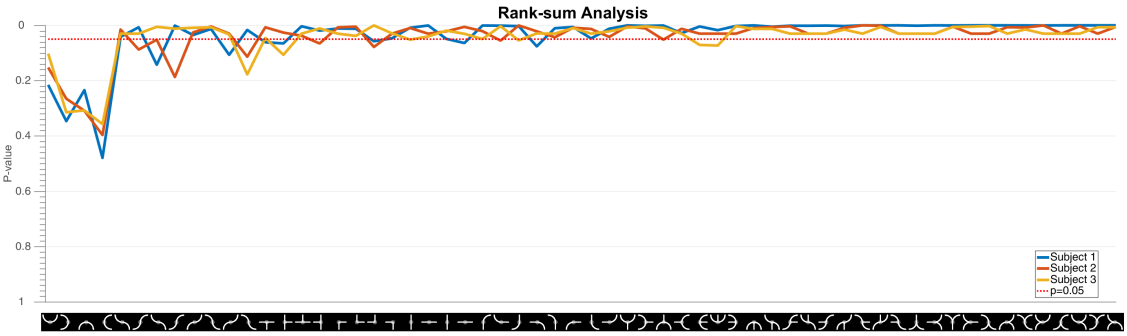


Figure 5.4: The correlation coefficients between the predicted and actual value of the stimulus set according to Leave-Six-Out analyses for each voxel for each subject.

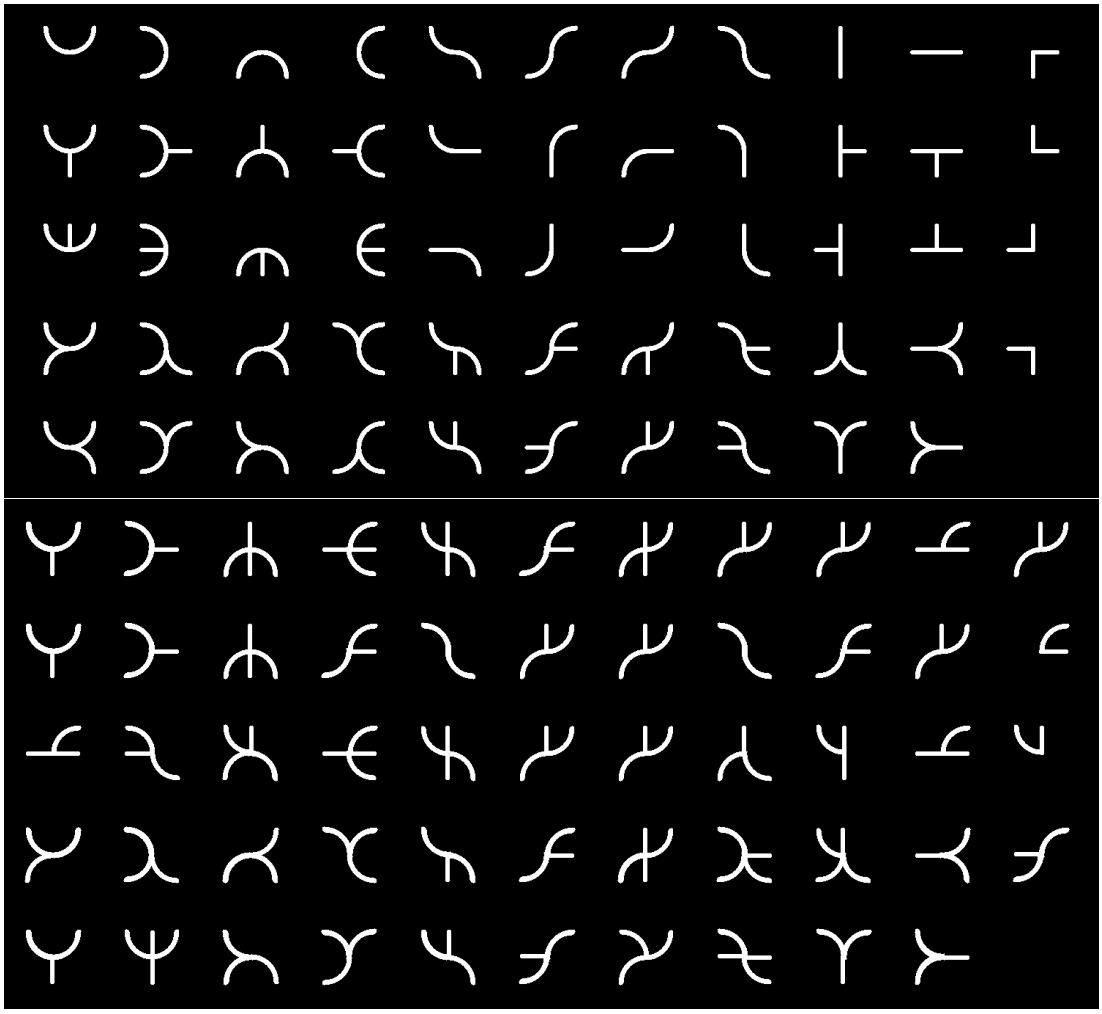


Figure 5.5: Comparison of a reconstruction set (below) to the stimulus set (above). For most fragments, even when they are incorrectly predicted, their appearance is close to the actual fragment.

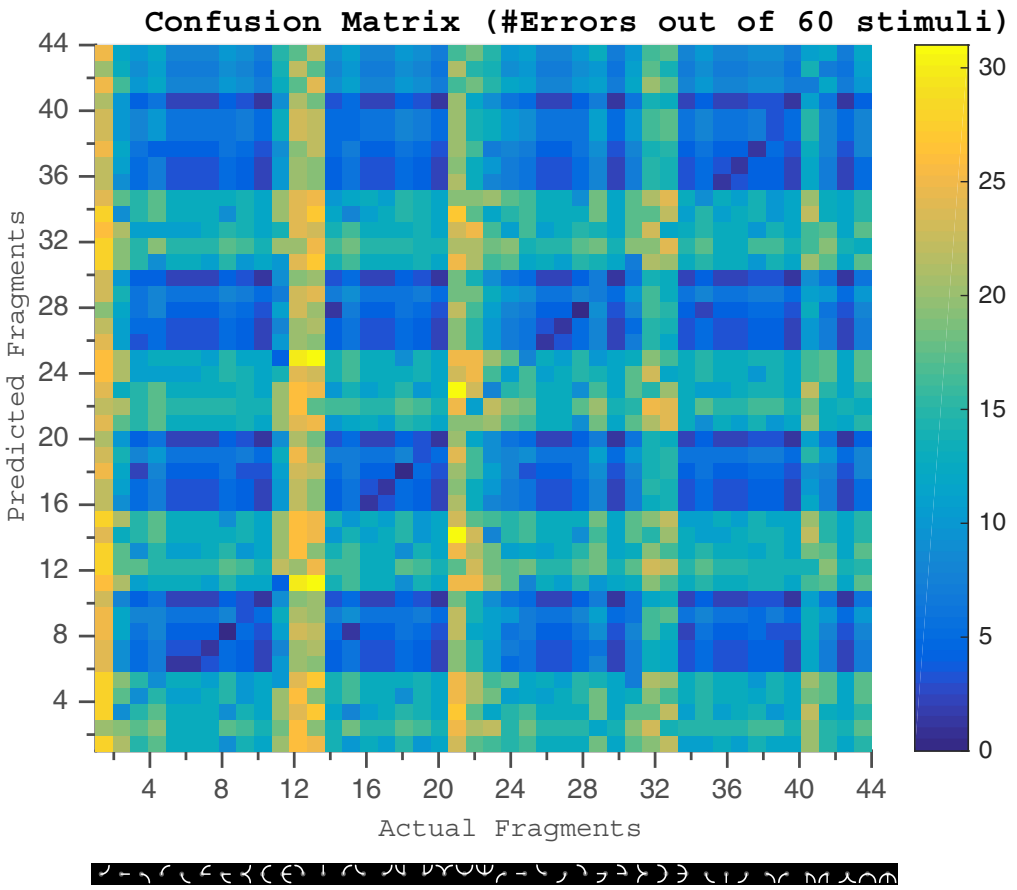
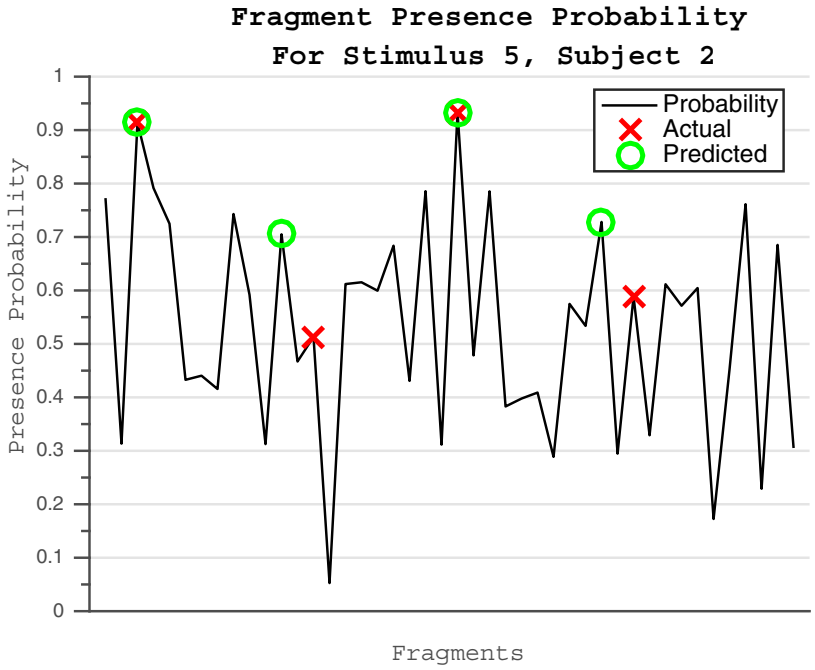
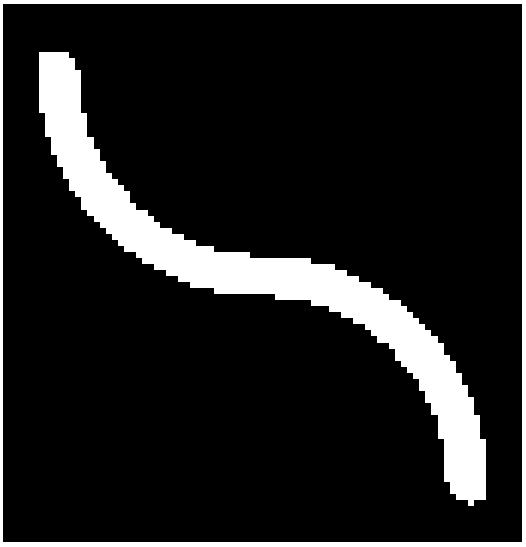


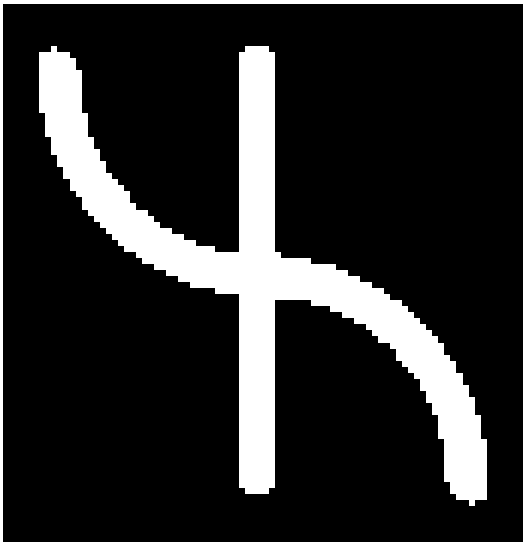
Figure 5.6: The confusion matrix displaying how many times a given fragment has been mis-predicted as another fragment within the 60 stimuli.



(a) The presence probability plot for stimulus 5. The predicted fragments are highlighted with green circles, and the actual fragments are highlighted with red crosses.

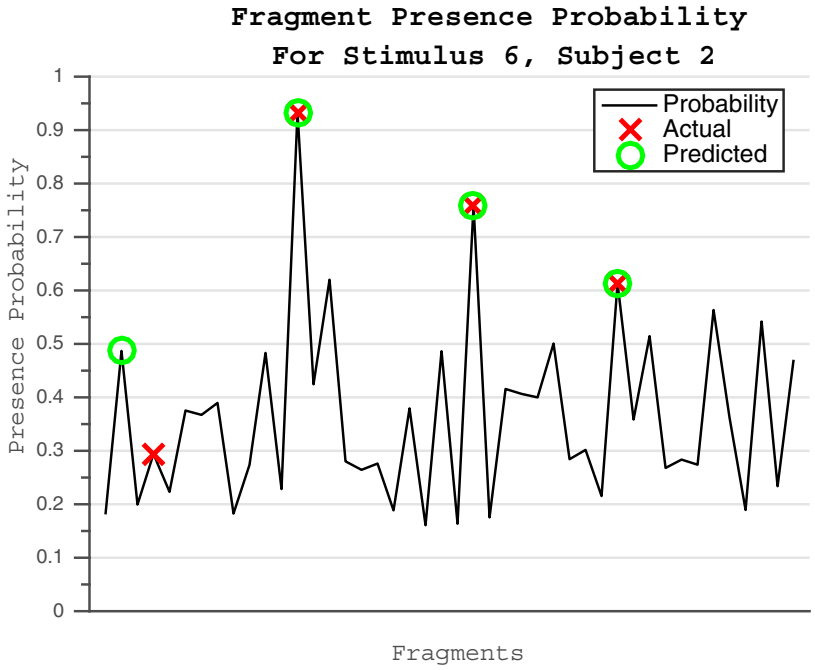


(b) Actual image of stimulus 5.

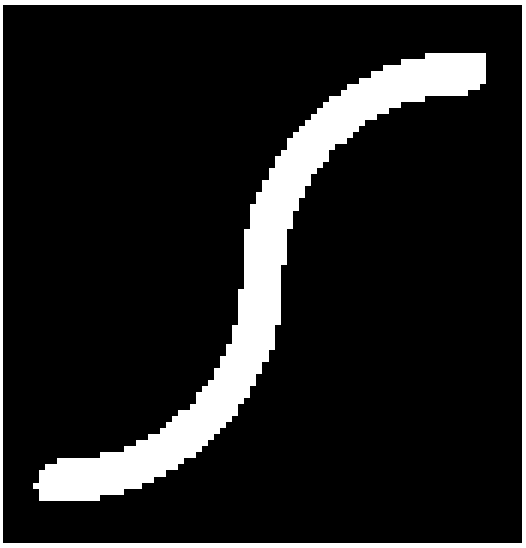


(c) Reconstructed image of stimulus 5.

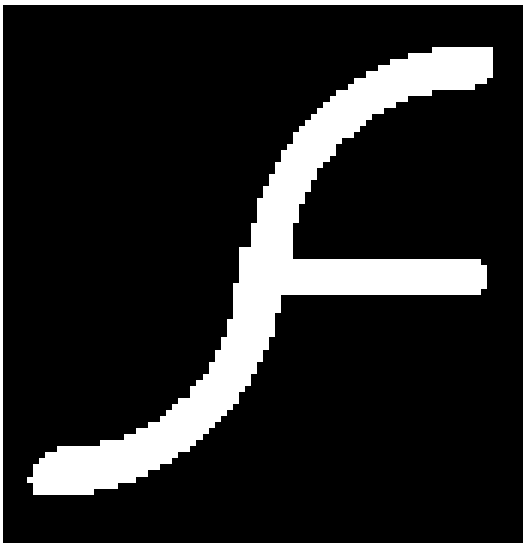
Figure 5.7: Demonstrating the reconstruction of stimulus 5. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. This reconstruction has an error in the second and fourth quadrants, which correspond to the quadrants facing 90 and 270 degrees, respectively. In both cases, instead of predicting an empty fragment, the algorithm predicted a straight line. In general, the algorithm seems to not favor predicting empty fragments.



(a) The presence probability plot for stimulus 6. The predicted fragments are highlighted with green circles, and the actual fragments are highlighted with red crosses.

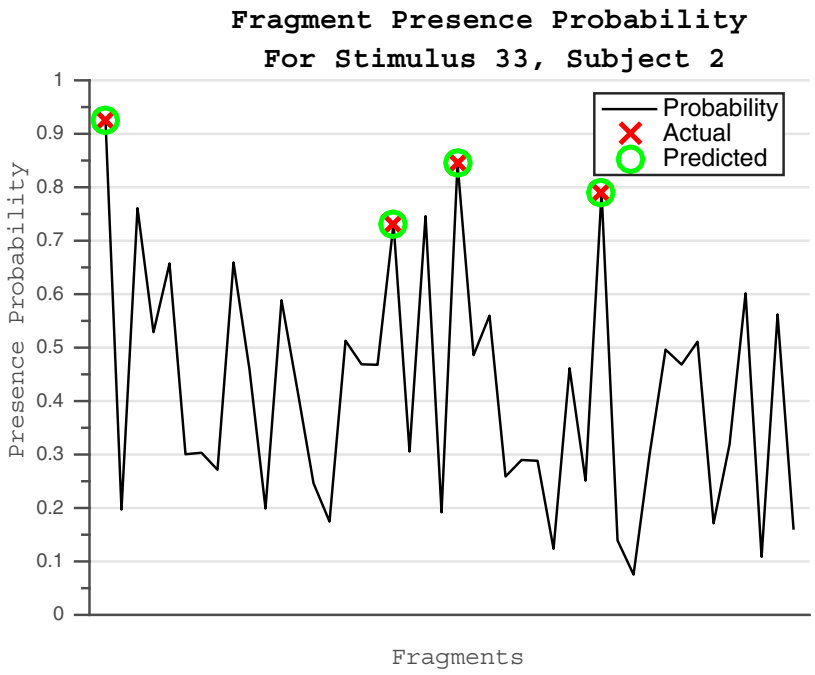


(b) Actual image of stimulus 6.

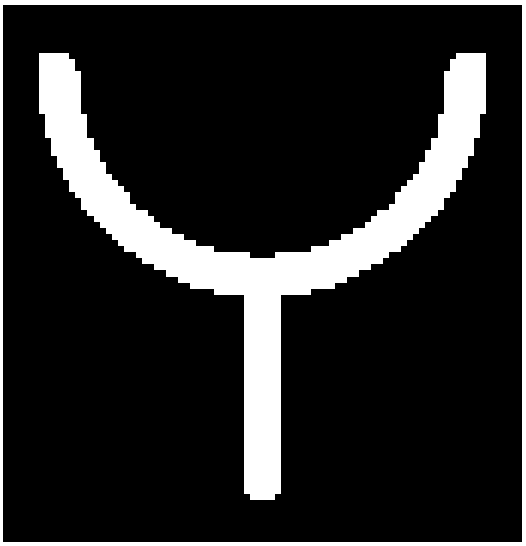


(c) Reconstructed image of stimulus 6.

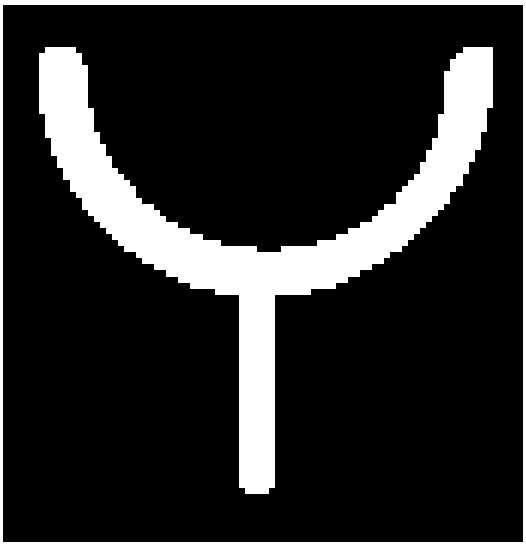
Figure 5.8: Demonstrating the reconstruction of stimulus 6. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. This reconstruction has an error in the first quadrant, facing 0 degrees. Again, the algorithm prefers a straight line over an empty fragment. However, in this case it is able to correctly predict and empty fragment in the third quadrant, facing 180 degrees.



(a) The presence probability plot for stimulus 33. The predicted fragments are highlighted with green circles, and the actual fragments are highlighted with red crosses.

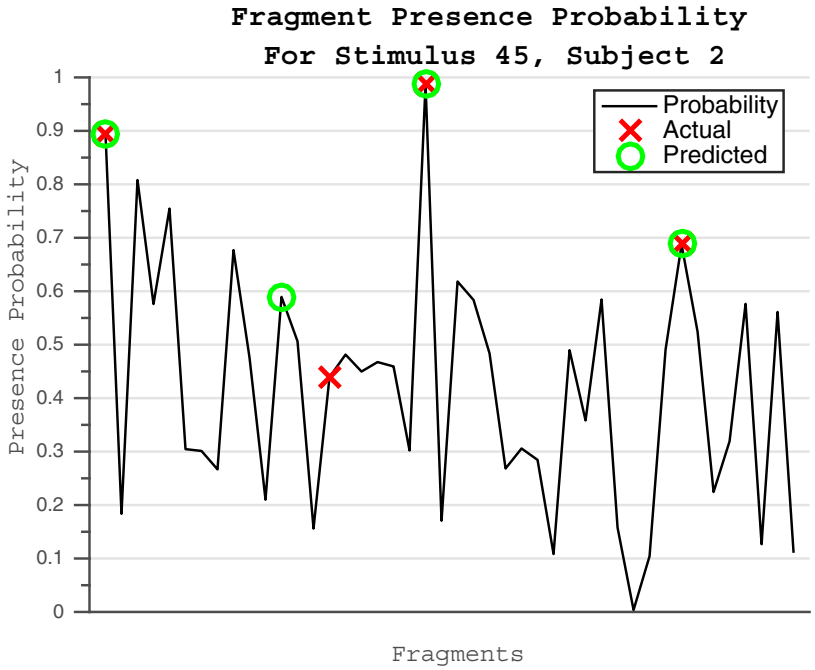


(b) Actual image of stimulus 33.

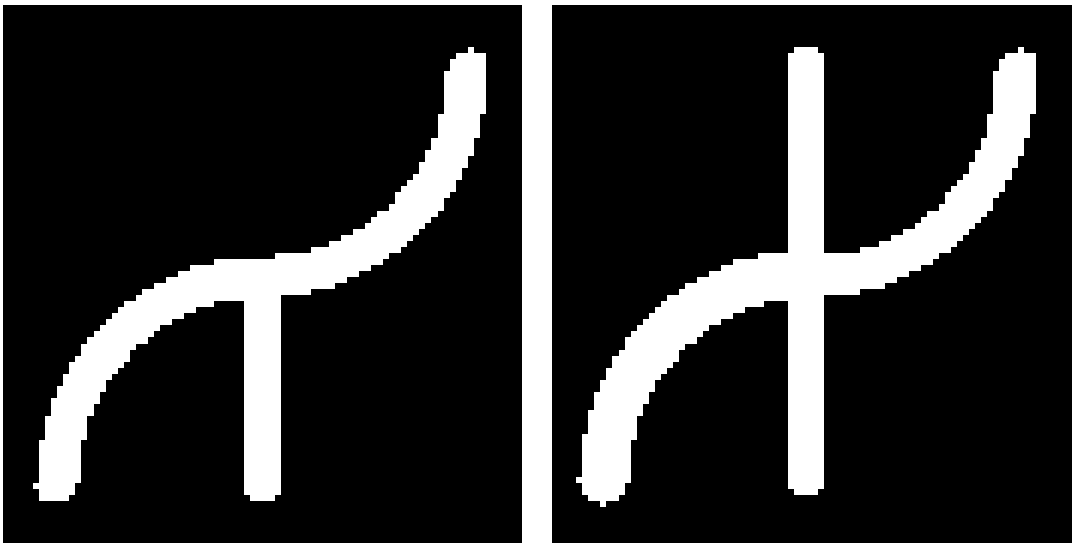


(c) Reconstructed image of stimulus 33.

Figure 5.9: Demonstrating the reconstruction of stimulus 33. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. This stimulus is reconstructed accurately. The algorithm does better with three-fragment stimuli than it does with two-fragment stimuli, as can be seen in Figure 5.4.



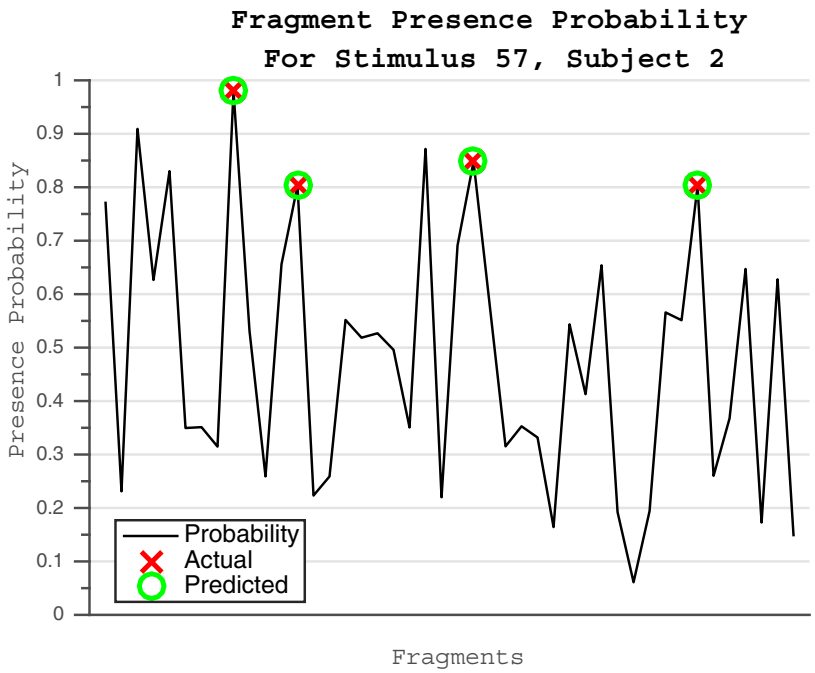
(a) The presence probability plot for stimulus 45. The predicted fragments are highlighted with green circles, and the actual fragments are highlighted with red crosses.



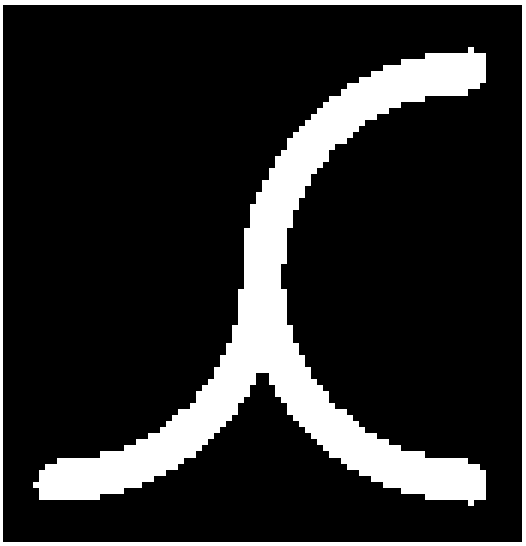
(b) Actual image of stimulus 45.

(c) Reconstructed image of stimulus 45.

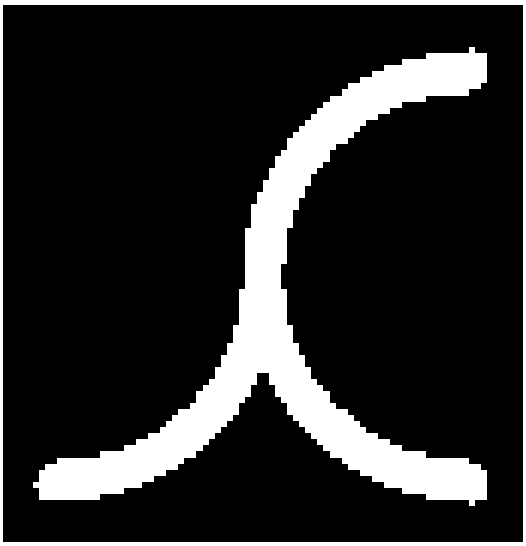
Figure 5.10: Demonstrating the reconstruction of stimulus 45. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. Similar to Figure 5.7a, the algorithm predicts a straight line in place of an empty fragment.



(a) The presence probability plot for stimulus 57. The predicted fragments are highlighted with green circles, and the actual fragments are highlighted with red crosses.



(b) Actual image of stimulus 57.



(c) Reconstructed image of stimulus 57.

Figure 5.11: Demonstrating the reconstruction of stimulus 57. The fragment presence predictions by Algorithm 5 are displayed, and the actual stimulus is shown alongside the reconstructed version. This stimulus is reconstructed accurately. The algorithm does better with three-fragment stimuli than it does with two-fragment stimuli, as can be seen in Figure 5.4.

5.5 Practical And Future Applications

Biological understanding of vision has long driven computer vision algorithms. One of the most commonly used feature detectors, SIFT, by Lowe [36] is based on how the primary visual cortex models images. There is a plethora of biologically-driven computer vision algorithms [21]. Understanding how the human brain represents images is key to developing algorithms that are better able to replicate human behavior. As such, a method of reconstructing objects based on brain images would open up new avenues for image processing. This has long been a goal of neuroscience.

Given more data, it could be possible to build a more complete model of how objects are constructed in the visual cortex. Neuron recording studies trace the signal caused by visual stimuli throughout several brain areas to see how the representation of images evolves. With a more robust parametrization, we could create a link between the spectral coding of images in early visual areas and the structural coding. This could provide a great leap in developing naturally-driven computer vision algorithms. There is work, for example by Zhu and Mumford [67] that tries to build a complete model of describing images, starting from local features, moving to structural descriptors, then to semantic constructs. This is analogous to how the human brain works, and developing a model based on brain imaging that fills in the gaps of this research would let vision algorithms represent images similarly to how humans do so.

This could also lead to, with the improvement of scanning technology, "mind

reading" - recreating one's thoughts, or at least the visual imagery they're thinking of, from their brain images. Work has shown that these visuals code not only for images being viewed but also being mentally imagined [59]. These technologies could help disabled patients communicate with vastly improved effectiveness. It can also be used for general communication, robot operation, and improved diagnostics of neurological conditions. These developments could also potentially be used against one's will, possibly extracting images they are thinking about without their consent. While these applications seem far away, there is already work in using fMRI for lie detection, and many legal questions facing its admissibility in court [14, 30, 38]. As technology develops, we will be forced to answer these questions. As a more specific example directly relevant to this work, an vastly improved version of the medial axis reconstruction approach could possibly be used to extract blueprints or diagrams from the mind of an unwilling participant. As such, to quote Naselaris et al. [44], "We believe that researchers in this field should begin to develop ethical guidelines for the application of brain-reading technology."

5.6 Conclusions

As per Chapter 4, the goal of this chapter is to provide evidence for structural coding of object shape in the lateral occipital complex voxels using fMRI. The previous chapter was focused on modeling the activity of voxels found in the LOC, using our

proposed medial axis stimuli (discussed in Chapter 2). This chapter focused on the same problem from the opposite direction, namely using voxel activation values to predict parameters of stimuli being viewed by the subject. Towards this end, we have developed an algorithm based on the fragment-driven parametrization proposed in Chapter 4. Given a novel stimulus, the algorithm calculates the presence probability of each fragment in the novel stimulus based on comparing the voxel activation values of stimuli containing the fragments to those who don't. Ranking these probabilities and post-processing the ranks, we obtain a prediction for each left-out stimulus. For 15 out of 60 stimuli, we have been able to reconstruct the stimuli perfectly, and for 25 out of 60 stimuli (averaged across subjects), we have been able to reconstruct the stimuli with a single error. We have then evaluated the quality of our results, and demonstrated their statistical significance against randomized trials. Even when the reconstructions had an error or two, they still performed better than significance. Only with stimuli that contained two fragments did the algorithm perform poorly. We attribute this to the unwillingness of the algorithm to assign empty fragments to quadrants, as it would be denoting a lack of activity, which is not well represented in our model.

Similarly to Chapter 4, these results are of significance, as they yet again show that modeling the behavior of human LOC using a structural parametrization of image stimuli is possible. In addition, this parametrization is viable enough to reconstruct the stimuli being viewed by the subjects, in some cases perfectly or with minor errors.

Similarly to the previous chapter, these preliminary findings could lead to further exploration of the LOC and mapping of its functionality. Moreover, these findings can lead to a greater understanding of how the brain parametrizes visual stimuli, leading to building a model of how each step in the visual system breaks down images. There are several studies done in the primary visual area (V1) or other early visual areas that reconstruct stimuli being viewed by subjects using fMRI, and there is a body of work demonstrating that some of these areas code for images using Gabor wavelets. Building upon these models with a model of the next level in visual processing can also lead to more biologically-driven computer vision algorithms. Many popular computer vision algorithms are based on understanding of visual areas (SIFT, for example [36]). Building upon the understanding of these algorithms by using neurologically-driven approaches can provide not just information about the human brain, but also better image processing algorithms that mimic human vision more accurately.

Chapter 6

Conclusions

The overarching goal of this thesis is to demonstrate that the lateral occipital complex in humans codes for images of objects in a structural fashion. More specifically, it is to demonstrate a structural parametrization of medial axis stimuli and build a computational model describing fMRI images of the LOC in response to stimuli derived from this parametrization. In Chapter 1 we specified four key challenges that needed to be addressed to achieve this goal. We now recap each of these challenges, and discuss our contributions and findings for addressing those challenges:

1. *How can we create computational methods based on neuroscience principles that are capable of preparing stimuli and executing an fMRI experiment to obtain data about very fine features from a complex brain area?* We approached this problem by designing and testing an experiment paradigm, and using an existing imaging technique in a new context to acquire high-resolution fMRI images.

Our experiment design allows us to show over a thousand stimuli to our subjects over a two hour period, while keeping them attending to the task at hand. Experiments that last this long are quite uncommon, so we performed an initial study to judge viability, and improved on our findings afterwards. This experiment paradigm can be used for any cognitive task based visual fMRI experiment. Our high resolution imaging technique can be used to image the visual cortex in more detail than any previous study done on 3 Tesla scanners, which are the industry standard. We also designed an algorithm to generate medial axis stimuli based on neurophysiology work in monkeys. We then took these principles and developed a software framework to facilitate these experiments. We designed this framework with flexibility and performance in mind, for usage in future experiments.

2. *How can we use conventional registration methods to pre-process narrow-field-of-view fMRI images and identify certain brain areas?* Since we used an unconventional imaging technique, steps that would have been simple in normal experiments were non-trivial for us. We needed to register our narrow images to full-brain structural images. We addressed this issue by utilizing meta-data acquired from the scanner to bootstrap the registration process. This step is essential for the usage of our imaging paradigm. Additionally, we implemented functional localizers to identify the lateral occipital complex and visual word-form area. For the former, we implemented an existing method with a novel

image data set, and for the latter we adapted a French-based localizer using an English text corpus. We then verified the correctness of both of these localizers.

3. *Can we develop a computational model that can describe the activity of LOC voxels informed by a structural parametrization of medial axis stimuli?* To answer this question, we developed a method of parametrizing our stimuli based on their structural fragments. We then tried several computational modeling approaches to find a fit between the voxel activations and parameters. Eventually we developed our own fitting method with an algorithm inspired by neuron recording work on monkeys, then demonstrated that the parametrization is valid using this algorithm. We verified our results using extensive cross validation. We then demonstrated that this parametrization is specific to the LOC by showing that it doesn't work on the visual word-form area. In addition, we tried a popular model that's used for other visual areas and showed that it doesn't work in the LOC. These findings show the first evidence for medial axis structure coding in LOC, and demonstrate that such a parametrization can be used to describe the activity of voxels in this area. We also validated the performance of our algorithm with an independent data set.

4. *Can we develop an algorithm that uses the parametrization of medial axis stimuli to reconstruct images being viewed by a subject from their LOC voxel activity?* To answer this question, we developed a reconstruction algorithm that predicts

the probability of a fragment being present in a novel stimulus based on our parametrization. We then used this algorithm and cross-validation to reconstruct left-out stimuli. We were able to achieve moderate success in doing so. We then demonstrated the significance of these results and explored further avenues for this work.

6.1 Discussion of Contributions

To put the work in this thesis into context, we will briefly describe the current state of neuroscience, specifically the understanding of the human brain. Early visual areas are well-understood both in terms of functionality and how they implement that functionality. The V1 is known to code for basic image properties like spatial frequencies, orientation and color. Many neuronal recording experiments in animals have mapped out the V1 in terms of how it works, and these studies are easily extendable to humans because of the similarities between human V1 and the V1 of several animals. many fMRI studies have shown specifically how the V1 responds to images using Gabor wavelet decompositions, correlating the animal study similarities. There have even been studies that reconstruct images being viewed by subjects by imaging their V1. The higher visual areas are less understood. Intermediate areas like V2 and V3 are still similar enough to animals, and their functionality is simple enough that several fMRI studies have used voxels from those areas to categorically reconstruct

images being viewed by subjects. While these aren't exact reconstructions of images, they employ a bag-of-words approach to composite images based on class labels. Higher areas area less understood, though. The lateral occipital complex, which receives inputs from the aforementioned areas, had not been parametrized before this work.

Work in monkeys has demonstrated that the inferotemporal cortex, which is considered to be the homolog of LOC, has shown that it codes for objects in a structural manner. Human fMRI studies also show that it responds to intact object images and doesn't respond to scrambled versions of the same images. It's also known not to be a semantic area. However, no study before has been able to describe how it exactly codes for shapes. Using knowledge gained from monkey studies, we've developed a preliminary parametrization that models how the LOC codes for medial axis shapes. This is the first human study to demonstrate such results, and thus it lays the groundwork for understanding how the next step in the human visual system processes stimuli. This work unlocks the possibility for future studies to explore in depth the workings of the LOC, and is the first work to propose and demonstrate the validity of a parametrization of the activity of LOC voxels. In addition to being able to describe voxel activity, our parametrization can be used to go in the other direction and reconstruct stimuli being viewed by subjects from their fMRI images. While the approach isn't perfect, it is the first proof of concept that it is possible to reconstruct images using information from LOC voxels. This is an unprecedented

result that opens up avenues for future studies to explore this area.

In addition to these scientific contributions, we have developed several methods to facilitate these contributions. We have developed an experiment paradigm to perform visual fMRI experiments. We've created a software framework for conducting fMRI experiments, driven by the needs of specific neuroscience goals. We've created a stimulus set, its parametrization, and an algorithm for generating stimuli based on this parametrization. These stimuli have been used for demonstrating the functionality of the LOC. We have also developed an imaging paradigm to acquire high resolution fMRI images. We have developed registration methods to use these images in the same manner as normal fMRI images. We have implemented an LOC localizer using a publicly available image data set. We have also developed an English VWFA localizer based on a French study. These tools, experiment paradigms and software principles will be invaluable for future fMRI studies, especially ones looking to replicate the approach we've demonstrated in this work and build on it. We have also developed a fitting method for explaining LOC voxel activity using our stimulus parametrization. This fitting method can be used for highly structured data all being generated by a rigid process, and it is designed to prevent overfitting. Finally, we have developed a reconstruction algorithm to predict the stimuli being viewed by subjects based on their LOC voxel activity.

6.2 Limitations

Now we will briefly discuss the limitations of our current work. Our approach was still greatly limited by the amount of data we could acquire. Until scanner technology improves, we are still bound by the resolution and others artifacts of the scanner. As a result, our stimulus set is limited, and our parametrization is rather simplistic. Exploring a greater range of fragments possibilities was our original goal, but we were unable to acquire good enough data to do so. 7 Tesla scanners are available at some facilities, however they give subjects physical discomfort in long experiments due to the intense magnetic field, which is why we avoided them, given that our experiments are extremely long. In the future, as scanners get less intrusive and more accurate, better results will be achievable.

Additionally, the fitting approach we use is not a very efficient algorithm. Its complexity is $O(C^4N)$ where C is the number of fragments and N is the number of stimuli. This is because least squares takes $O(C^2N)$ time, and at each step we compute a least squares for the possibility of removing each fragment, and we do this for at most C iterations. This is not an efficient algorithm, but the iterative approach of the algorithm is key to solving the overfitting problem. Combined with the fact that we ran tens of thousands of randomized trials to compute the statistical significance of our result, this approach was extremely time-consuming, even when run in parallel. The algorithm was implemented in MATLAB for convenience, and an implementation in a more efficient language would help in the long run. Deriving

a solver for the algorithm that is based on optimization theory instead of an iterative solver could help as well, but given the non-convex nature of the problem, the solver could be problematic as well.

The reconstruction algorithm is limited by certain biases in the data. A fuzzy reconstruction algorithm that takes into account these biases and uses information from multiple fragments to predict a probability for each fragment could be implemented. However, given that our data is currently limited to 60 unique stimuli and that we want to validate our algorithm with cross-validation, this approach leads to data being too fragmented to develop a robust model from. With better scanning, we could obtain more data and implement such an algorithm.

It would have been possible to ask subjects to come in for multiple session to record more data from them, but this approach is known to be extremely unreliable for fMRI experiments and most of the time it amplifies noise instead of providing more signal. For simpler visual areas this would have been feasible, but for the LOC we were barely able to get any signal at all. To be able to confound the signal from the noise with multiple experiments, we would need better scanning technology.

6.3 Future Work

We have successfully shown preliminary evidence of structural coding in the lateral occipital complex. We have also developed an experiment paradigm and modeling

and reconstruction algorithms to demonstrate the validity of this paradigm. However, these results are preliminary, as they are not 100% reliable as predictors, and they are limited in scope.

Unfortunately, in terms of experimentation, there isn't much that can be done until scanner technology improves. We might have been too ambitious with the parametrization of even our latest stimulus set, and a future study could further reduce the complexity of their stimulus set in order to demonstrate more robust results from simpler data. Now that there is evidence that the LOC structurally codes for object shapes, future studies can try stimuli that are not medial axis-based for different parametrizations. The monkey IT papers could provide insight into where future LOC studies can go. Globular shapes or 3D shapes could be another avenue to explore. The reconstruction algorithm can be further improved as well, given a different parametrization scheme. Our stimulus generation algorithm can easily be improved with the introduction of new stimulus design principles.

The imaging paradigm we've developed can be used to reiterate existing visual cortex studies with higher resolutions to achieve improved results. The works that have parametrized and reconstructed stimuli from the V1 can make use of this scanning technique to obtain more accurate results. Our experiment paradigms can be extended to similar studies, and our framework could be developed into a full-fledged cross-platform neuroscience/cognitive science experimentation suite.

The two-group fitting algorithm we've developed can be used as a sparse fitting

algorithm that doesn't depend on a λ parameter to balance sparseness against fitting, and is more robust to changes in the data due to the two-group manner taking into account data that isn't being included by the linear model. As the sparseness of the model is driven by the quality of the fit instead of an external parameter, this approach can be used in a more data-driven manner.

The reconstruction approach we've developed demonstrates that the human brain represents visual stimuli as a combination of parts based on curvature, orientation and polar position. This representation can be used to develop an object recognition algorithm that describes images based on structural components that make objects up. There is already work done in this area by Zhu and Mumford [67]. These results provide further justification towards building a semantic model based on objects composed of structural parts. In the long term, understanding of LOC and other brain areas will lead to developing a computational model of the entire human brain.

Bibliography

[1] A. Amedi, G. Jacobson, T. Hendler, R. Malach, and E. Zohary, "Convergence of visual and tactile shape processing in the human lateral occipital complex," *Cerebral Cortex*, vol. 12, no. 11, pp. 1202–1212, 2002.

[2] F. G. Ashby, *Statistical analysis of fMRI data*. MIT press, 2011.

[3] R. B. Bendel and A. A. Afifi, "Comparison of stopping rules in forward stepwise regression," *Journal of the American Statistical Association*, vol. 72, no. 357, pp. 46–53, 1977.

[4] S. L. Brincat and C. E. Connor, "Underlying principles of visual shape selectivity in posterior inferotemporal cortex," *Nature neuroscience*, vol. 7, no. 8, pp. 880–886, 2004.

[5] R. B. Buxton, K. Uludağ, D. J. Dubowitz, and T. T. Liu, "Modeling the hemodynamic response to brain activation," *Neuroimage*, vol. 23, pp. S220–S233, 2004.

BIBLIOGRAPHY

[6] K. Buza, "Feedback prediction for blogs," in *Data analysis, machine learning and knowledge discovery*. Springer, 2014, pp. 145–152.

[7] J. W. Carland and J. C. Carland, "A model of potential entrepreneurship: Profiles and educational implications," *Journal of Small Business Strategy*, vol. 8, no. 1, pp. 1–14, 2015.

[8] L. Cohen and S. Dehaene, "Specialization within the ventral stream: the case for the visual word form area," *Neuroimage*, vol. 22, no. 1, pp. 466–476, 2004.

[9] L. Cohen, S. Lehéricy, F. Chochon, C. Lemer, S. Rivaud, and S. Dehaene, "Language-specific tuning of visual cortex? functional properties of the visual word form area," *Brain*, vol. 125, no. 5, pp. 1054–1069, 2002.

[10] H. J. Cordell and D. G. Clayton, "A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to hla in type 1 diabetes," *The American Journal of Human Genetics*, vol. 70, no. 1, pp. 124–141, 2002.

[11] R. W. Cox, "Afni: software for analysis and visualization of functional magnetic resonance neuroimages," *Computers and Biomedical research*, vol. 29, no. 3, pp. 162–173, 1996.

[12] ——, "Afni: what a long strange trip it's been," *Neuroimage*, vol. 62, no. 2, pp. 743–747, 2012.

[13] R. W. Cox and J. S. Hyde, "Software tools for analysis and visualization of fmri data," *NMR in Biomedicine*, vol. 10, no. 45, pp. 171–178, 1997.

[14] C. Davatzikos, K. Ruparel, Y. Fan, D. Shen, M. Acharyya, J. Loughead, R. Gur, and D. D. Langleben, "Classifying spatial patterns of brain activity with machine learning methods: application to lie detection," *Neuroimage*, vol. 28, no. 3, pp. 663–668, 2005.

[15] P. Dean, "Effects of inferotemporal lesions on the behavior of monkeys." *Psychological bulletin*, vol. 83, no. 1, p. 41, 1976.

[16] S. Dehaene and L. Cohen, "The unique role of the visual word form area in reading," *Trends in cognitive sciences*, vol. 15, no. 6, pp. 254–262, 2011.

[17] R. Dubner and S. Zeki, "Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey," *Brain research*, vol. 35, no. 2, pp. 528–532, 1971.

[18] K. Grill-Spector, T. Kushnir, S. Edelman, G. Avidan, Y. Itzchak, and R. Malach, "Differential processing of objects under various viewing conditions in the human lateral occipital complex," *Neuron*, vol. 24, no. 1, pp. 187–203, 1999.

[19] K. Grill-Spector, Z. Kourtzi, and N. Kanwisher, "The lateral occipital complex and its role in object recognition," *Vision research*, vol. 41, no. 10, pp. 1409–1422, 2001.

[20] C. G. Gross, C. Rocha-Miranda, and D. Bender, "Visual properties of neurons in inferotemporal cortex of the macaque." *Journal of neurophysiology*, vol. 35, no. 1, pp. 96–111, 1972.

[21] J. Herault, G. Cristobal, L. Perrinet, and M. S. Keil, *Biologically Inspired Computer Vision: Fundamentals and Applications*. John Wiley & Sons, 2015.

[22] A. E. Hillis, M. Newhart, J. Heidler, P. Barker, E. Herskovits, and M. Degaonkar, "The roles of the visual word form area in reading," *Neuroimage*, vol. 24, no. 2, pp. 548–559, 2005.

[23] A. Holkner, "Pyglet," http://www.pyglet.org/, 2007–2016.

[24] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional magnetic resonance imaging*. Sinauer Associates Sunderland, 2004, vol. 1.

[25] C.-C. Hung, E. T. Carlson, and C. E. Connor, "Medial axis shape coding in macaque inferotemporal cortex," *Neuron*, vol. 74, no. 6, pp. 1099–1113, 2012.

[26] R. Jennrich and P. Sampson, "Application of stepwise regression to non-linear estimation," *Technometrics*, vol. 10, no. 1, pp. 63–72, 1968.

[27] K. N. Kay and J. L. Gallant, "I can see what you see," *Nature neuroscience*, vol. 12, no. 3, pp. 245–245, 2009.

[28] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, "Identifying natural

images from human brain activity," *Nature*, vol. 452, no. 7185, pp. 352–355, 2008.

[29] W. K. Kirchner, "Age differences in short-term retention of rapidly changing information." *Journal of experimental psychology*, vol. 55, no. 4, p. 352, 1958.

[30] L. Kittay, "Admissibility of fmri lie detection-the cultural bias against mind reading devices," *Brook. L. Rev.*, vol. 72, p. 1351, 2006.

[31] Z. Kourtzi and N. Kanwisher, "Cortical regions involved in perceiving object shape," *The Journal of Neuroscience*, vol. 20, no. 9, pp. 3310–3318, 2000.

[32] ——, "Representation of perceived object shape by the human lateral occipital complex," *Science*, vol. 293, no. 5534, pp. 1506–1509, 2001.

[33] Z. Kourtzi, M. Erb, W. Grodd, and H. H. Bülthoff, "Representation of the perceived 3-d object shape in the human lateral occipital complex," *Cerebral Cortex*, vol. 13, no. 9, pp. 911–920, 2003.

[34] M. Kronbichler, F. Hutzler, H. Wimmer, A. Mair, W. Staffen, and G. Ladurner, "The visual word form area and the frequency with which words are encountered: evidence from a parametric fmri study," *Neuroimage*, vol. 21, no. 3, pp. 946–953, 2004.

[35] X. Liao, Q. Li, X. Yang, W. Zhang, and W. Li, "Multiobjective optimization for

crash safety design of vehicles using stepwise regression model," *Structural and Multidisciplinary Optimization*, vol. 35, no. 6, pp. 561–569, 2008.

[36] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[37] R. Malach, J. Reppas, R. Benson, K. Kwong, H. Jiang, W. Kennedy, P. Ledden, T. Brady, B. Rosen, and R. Tootell, "Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex," *Proceedings of the National Academy of Sciences*, vol. 92, no. 18, pp. 8135–8139, 1995.

[38] D. P. McCabe, A. D. Castel, and M. G. Rhodes, "The influence of fmri lie detection evidence on juror decision-making," *Behavioral sciences & the law*, vol. 29, no. 4, pp. 566–577, 2011.

[39] S. Michael, B. Richard, and R. George, "Cognitive neuroscience, the biology of the mind," 1998.

[40] M. Mishkin, "A memory system in the monkey," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 298, no. 1089, pp. 85–95, 1982.

[41] T. M. Mitchell, S. V. Shinkareva, A. Carlson, K.-M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, "Predicting human brain activity associated with the meanings of nouns," *science*, vol. 320, no. 5880, pp. 1191–1195, 2008.

[42] Y. Miyawaki, H. Uchida, O. Yamashita, M.-a. Sato, Y. Morito, H. C. Tanabe, N. Sadato, and Y. Kamitani, "Visual image reconstruction from human brain activity using a combination of multiscale local image decoders," *Neuron*, vol. 60, no. 5, pp. 915–929, 2008.

[43] M. M. Monti, "Statistical analysis of fmri time-series: a critical review of the glm approach," *Frontiers in human neuroscience*, 2011.

[44] T. Naselaris, R. J. Prenger, K. N. Kay, M. Oliver, and J. L. Gallant, "Bayesian reconstruction of natural images from human brain activity," *Neuron*, vol. 63, no. 6, pp. 902–915, 2009.

[45] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant, "Encoding and decoding in fmri," *Neuroimage*, vol. 56, no. 2, pp. 400–410, 2011.

[46] S. A. Nene, S. K. Nayar, and H. Murase. Columbia object image library (coil-100). [Online]. Available: http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php

[47] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fmri data," *Trends in cognitive sciences*, vol. 10, no. 9, pp. 424–430, 2006.

[48] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set:

A strategy employed by v1?" *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[49] A. Pasupathy and C. E. Connor, "Population coding of shape in area v4," *Nature neuroscience*, vol. 5, no. 12, pp. 1332–1338, 2002.

[50] R. A. Poldrack, P. C. Fletcher, R. N. Henson, K. J. Worsley, M. Brett, and T. E. Nichols, "Guidelines for reporting an fmri study," *Neuroimage*, vol. 40, no. 2, pp. 409–414, 2008.

[51] H. B. Pribram and J. Barry, "Further behavioral analysis of parieto temporo-preoccipital cortex," *J. Neurophysiol.*, vol. 19, pp. 99–106, 1956.

[52] C. J. Price and J. T. Devlin, "The myth of the visual word form area," *Neuroimage*, vol. 19, no. 3, pp. 473–481, 2003.

[53] Z. S. Saad, D. R. Glen, G. Chen, M. S. Beauchamp, R. Desai, and R. W. Cox, "A new method for improving functional-to-structural mri alignment using local pearson correlation," *Neuroimage*, vol. 44, no. 3, pp. 839–848, 2009.

[54] K. D. Singh, A. Smith, and M. Greenlee, "Spatiotemporal frequency and direction sensitivities of human visual areas measured using fmri," *Neuroimage*, vol. 12, no. 5, pp. 550–564, 2000.

[55] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[56] S. M. Smith, P. M. Matthews, and P. Jezzard, *Functional MRI: an introduction to methods*. Oxford university press, 2001.

[57] M. Szwed, S. Dehaene, A. Kleinschmidt, E. Eger, R. Valabregue, A. Amadon, and L. Cohen, "Specialization for written words over objects in the visual cortex," *Neuroimage*, vol. 56, no. 1, pp. 330–344, 2011.

[58] K. Tanaka, H.-a. Saito, Y. Fukada, and M. Moriya, "Coding visual images of objects in the inferotemporal cortex of the macaque monkey," *Journal of neurophysiology*, vol. 66, no. 1, pp. 170–189, 1991.

[59] B. Thirion, E. Duchesnay, E. Hubbard, J. Dubois, J.-B. Poline, D. Lebihan, and S. Dehaene, "Inverse retinotopy: inferring the visual content of images from brain activation patterns," *Neuroimage*, vol. 33, no. 4, pp. 1104–1116, 2006.

[60] B. Thompson, "Stepwise regression and stepwise discriminant analysis need not apply." *Report, Annual Meeting of the American Educational Research Association*, 1995.

[61] M. J. Tovée, *An introduction to the visual system*. Cambridge University Press, 1996.

[62] B. Y. University. Word frequency data corpus of contemporary american english. [Online]. Available: http://www.wordfrequency.info/

[63] V. Q. Vu, P. Ravikumar, T. Naselaris, K. N. Kay, J. L. Gallant, and B. Yu,

"Encoding and decoding v1 fmri responses to natural images with sparse non-parametric models," *The annals of applied statistics*, vol. 5, no. 2B, p. 1159, 2011.

[64] M. Wilson, "Effects of circumscribed cortical lesions upon somesthetic and visual discrimination in the monkey." *Journal of Comparative and Physiological Psychology*, vol. 50, no. 6, p. 630, 1957.

[65] 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*, vol. 11, no. 11, pp. 1352–1360, 2008.

[66] M. A. Yassa, S. M. Stark, A. Bakker, M. S. Albert, M. Gallagher, and C. E. Stark, "High-resolution structural and functional mri of hippocampal ca3 and dentate gyrus in patients with amnestic mild cognitive impairment," *Neuroimage*, vol. 51, no. 3, pp. 1242–1252, 2010.

[67] S.-C. Zhu and D. Mumford, *A stochastic grammar of images*. Now Publishers Inc, 2007.

Vita

Haluk Tokgozoglu received the Bachelors of Engineering in Computer Science and Engineering from Bilkent University in 2009, and the Masters of Science in Computer Science from Johns Hopkins University in 2012. He enrolled in the Computer Science Ph.D. program at Johns Hopkins University in 2010. His research focuses on Machine Learning, Computer Vision and Visual Neuroscience.