EXTRACTING FMRI BRAIN PATTERNS SIGNIFICANTLY RELATED TO BEHAVIOR VIA INDIVIDUAL PREPROCESSING PIPELINE OPTIMIZATION

by

Robyn Spring

A thesis submitted in conformity with the requirements for the degree of Master of Science

Institute of Medical Science
University of Toronto

© Copyright by Robyn Spring (2012)
Abstract

Background: Functional magnetic resonance imaging (fMRI) can require extensive preprocessing to minimize noise and maximize signal. There is evidence suggesting that fixed-subject preprocessing pipelines, the current standard in fMRI preprocessing, are suboptimal compared to individual-subject pipelines.

Aim: We sought to test if individual-subject preprocessing pipeline optimization, compared to fixed, resulted in stronger and more reliable brain-patterns in episodic recognition.

Methodology: 27 young healthy controls were scanned via fMRI while performing forced-choice episodic recognition. Several sets of fMRI preprocessing pipelines were tested and optimized in a fixed and individual-subject manner, using methods outlined by Churchill et al. (2011).

Results: Individual-subject pipeline optimization, compared to fixed, significantly increased reproducibility, significantly increased the detection of positively and negatively activated voxels, and resulted in a brain-pattern with significant correlation to a task behavioral measure.
Conclusions: Individual-subject pipeline optimization, compared to fixed, led to stronger and more reliable brain-patterns that are significantly correlated with behavior.
Acknowledgments

First and foremost, I would like to express my sincere and utmost gratitude to my supervisor Dr. Stephen Strother. His steadfast guidance, patience and support allowed me to develop as a researcher and critical thinker. I would like to thank him for providing me the opportunity to learn from him and be a member of his laboratory. His expertise and dedication to this work has been invaluable and I could not have accomplished this thesis without him.

Thank you to my wonderful committee members, Dr. Cheryl Grady and Dr. Mary Pat McAndrews for dedicating their precious time to this project. Their guidance, support, knowledge and feedback have been invaluable to the completion of my Master’s thesis.

I am also grateful for all of the support I received from the members of the Strother laboratory. Their advice and guidance were integral to the completion of this thesis. Specifically, I would like to thank Nathan Churchill for his endless help and support. I would also like to thank Dr. Shafagh Fallah for her assistance and advice.

Finally, I would like to thank my family and friends for their unconditional love and support. Thank you for learning all about fMRI research in order to help me improve my writing and presentation skills. I am very lucky to have such a strong support network!
Table of Contents

Acknowledgments ........................................................................................................ iv

Table of Contents ......................................................................................................... v

List of Abbreviations ..................................................................................................... viii

List of Tables ................................................................................................................ xi

List of Figures ................................................................................................................. xii

1 Introduction ............................................................................................................... 1
  1.1 Functional Neuroimaging ....................................................................................... 3
  1.2 Fundamentals of Functional MRI .......................................................................... 5
  1.3 Functional MRI Stimuli ........................................................................................ 9
    1.3.1 Encoding and Recognition Tasks ................................................................... 9
    1.3.2 Neuropsychological Tests versus Functional MRI Response Modes .......... 10
    1.3.3 Functional MRI Task Design ....................................................................... 12
  1.4 Functional MRI Noise Sources ............................................................................. 14
  1.5 Functional MRI Preprocessing Techniques .......................................................... 16
    1.5.1 Physiological Noise Correction .................................................................... 16
    1.5.2 Slice Timing Correction .............................................................................. 17
    1.5.3 Motion Correction ........................................................................................ 18
    1.5.4 Spatial Smoothing ....................................................................................... 19
    1.5.5 Temporal Detrending and Motion Parameter Regression ............................ 20
    1.5.6 Between Subject Alignment ....................................................................... 21
  1.6 Fixed versus Individual Pipeline Optimization ..................................................... 21
  1.7 Episodic Memory and Neuroimaging .................................................................... 23

2 Methods ..................................................................................................................... 31
  2.1 Task Design ............................................................................................................ 31
2.2 Subject Selection .............................................................................................................. 36
2.3 Data Acquisition .............................................................................................................. 37
2.4 Preprocessing .................................................................................................................. 38
2.5 Evaluating Behavior and Outliers ................................................................................... 41
  2.5.1 Reaction Time and Accuracy .................................................................................. 41
  2.5.2 Task-Correlated Motion (TCM) ........................................................................... 42
2.6 Pipeline Optimization Procedure .................................................................................... 42
  2.6.1 Determining the Optimal Fixed Pipeline ......................................................... 44
  2.6.2 Determining the Optimal Individual Pipeline ............................................... 45
2.7 Creating a Standard Group Template ............................................................................. 46
2.8 Analysis of Optimal SPM Results ................................................................................... 48
  2.8.1 Singular Value Decomposition (SVD) ............................................................... 48
2.9 Preliminary study ............................................................................................................. 51
3 Results ................................................................................................................................ 53
  3.1 Behavioral Results ......................................................................................................... 53
    3.1.1 Behavioral Outlier Analysis ............................................................................... 53
    3.1.2 Behavioral Results with Outliers Removed .................................................... 56
3.2 Subject Motion ................................................................................................................ 56
3.3 Physiological Parameters ............................................................................................... 58
3.4 Fixed Pipeline Optimization ......................................................................................... 59
3.5 Individual Pipeline Optimization .................................................................................... 69
3.6 Fixed versus Individually Optimized Pipelines ............................................................. 78
4 Discussion ............................................................................................................................ 85
  4.1 Limitations ...................................................................................................................... 95
  4.2 Future Directions .......................................................................................................... 96
4.3 Conclusions .................................................................................................................. 100
References .......................................................................................................................... 102
Appendices .......................................................................................................................... 110
Appendix A .......................................................................................................................... 110
Appendix B .......................................................................................................................... 112
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2cCVA</td>
<td>Two-class canonical variates analysis</td>
</tr>
<tr>
<td>ACC</td>
<td>Accuracy</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike’s Information Criterion</td>
</tr>
<tr>
<td>BOLD</td>
<td>Blood oxygenation-level dependent</td>
</tr>
<tr>
<td>BSCC</td>
<td>Baycrest Stroke and Cognition Clinic</td>
</tr>
<tr>
<td>D</td>
<td>Euclidean distance</td>
</tr>
<tr>
<td>DET</td>
<td>Temporal detrending</td>
</tr>
<tr>
<td>DMN</td>
<td>Default mode network</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>fixACC</td>
<td>Fixation accuracy</td>
</tr>
<tr>
<td>fixRT</td>
<td>Fixation reaction time</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional magnetic resonance imaging</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of view</td>
</tr>
<tr>
<td>fSNR</td>
<td>Functional signal-to-noise ratio</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full-width-at-half-maximum</td>
</tr>
<tr>
<td>HERA</td>
<td>Hemispheric encoding/retrieval asymmetry model</td>
</tr>
<tr>
<td>HRF</td>
<td>Hemodynamic response function</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter-stimulus interval</td>
</tr>
<tr>
<td>K+</td>
<td>Potassium ions</td>
</tr>
<tr>
<td>MC</td>
<td>Motion correction</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetoencephalography</td>
</tr>
<tr>
<td>MMSE</td>
<td>Mini-Mental Status Examination</td>
</tr>
<tr>
<td>MNI</td>
<td>Montreal Neurological Institute</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MPE</td>
<td>Motion parameter estimate</td>
</tr>
<tr>
<td>MPR</td>
<td>Motion parameter regression</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>Na⁺</td>
<td>Sodium ions</td>
</tr>
<tr>
<td>NPAIRS</td>
<td>Nonparametric Prediction, Activation, Influence, and Reproducibility reSamping</td>
</tr>
<tr>
<td>P</td>
<td>Prediction</td>
</tr>
<tr>
<td>PC</td>
<td>Principal component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PFC</td>
<td>Prefrontal cortex</td>
</tr>
<tr>
<td>PHYCAA</td>
<td>PHYsiological correction using Canonical Autocorrelation Analysis</td>
</tr>
<tr>
<td>PMR</td>
<td>Posterior midline region</td>
</tr>
<tr>
<td>PNC</td>
<td>Physiological noise correction</td>
</tr>
<tr>
<td>R</td>
<td>Reproducibility</td>
</tr>
<tr>
<td>recACC</td>
<td>Recognition accuracy</td>
</tr>
<tr>
<td>recRT</td>
<td>Recognition reaction time</td>
</tr>
<tr>
<td>RSN</td>
<td>Retrieval success network</td>
</tr>
<tr>
<td>RT</td>
<td>Reaction time</td>
</tr>
<tr>
<td>RTdif</td>
<td>Difference between recognition and fixation reaction times</td>
</tr>
<tr>
<td>SART</td>
<td>Sustained attention to response task</td>
</tr>
<tr>
<td>SM</td>
<td>Spatial smoothing</td>
</tr>
<tr>
<td>SPM</td>
<td>Statistical parametric map</td>
</tr>
<tr>
<td>STC</td>
<td>Slice timing correction</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular value decomposition</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>TAP</td>
<td>Transfer-appropriate processing</td>
</tr>
<tr>
<td>TCM</td>
<td>Task-correlated motion</td>
</tr>
<tr>
<td>TPN</td>
<td>Task positive network</td>
</tr>
<tr>
<td>TR</td>
<td>Repetition time</td>
</tr>
<tr>
<td>VPC</td>
<td>Ventral parietal cortex</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1 Median and maximum displacement variance .......................................................... 52

Table 3.1 Fixed pipeline optimization SVD peak voxel activation ......................................... 68

Table 3.2 Individual pipeline optimization SVD peak voxel activation ................................... 77

Table 3.3 Individual subject pipelines versus fixed pipeline optimization ............................. 79
List of Figures

Figure 2.1 FMRI tablet configuration ................................................................. 33

Figure 2.2 Task design ...................................................................................... 36

Figure 2.3 108 Preprocessing pipelines applied on data ..................................... 40

Figure 2.4 Reproducibility and Prediction in NPAIRS ...................................... 45

Figure 2.5 Median rank metric ......................................................................... 46

Figure 3.1 Reaction time and accuracy plots for all subjects ........................... 54

Figure 3.2 Incorrect trials and false hits .......................................................... 55

Figure 3.3 Reaction time and accuracy plots without outliers ............................ 57

Figure 3.4 Motion correction versus maximum displacement .......................... 58

Figure 3.5 Physiological data ........................................................................... 59

Figure 3.6 Fixed pipeline optimization ............................................................. 61

Figure 3.7 Fixed pipeline optimization SVD results ......................................... 65

Figure 3.8 Number of subjects individually optimized with each preprocessing step ................................................................. 70

Figure 3.9 Z-scored SPMs with motion artifacts .............................................. 71

Figure 3.10 Individual pipeline optimization SVD results .................................. 76

Figure 3.11 Comparison of fixed and individually optimized significant voxels ................................................................. 80
Figure 3.12 Comparison of fixed and individually optimized SVD results ........................................84
1 Introduction

Functional magnetic resonance imaging (fMRI), a method of neuroimaging reflecting brain activity, is becoming an essential tool in the study of aging and stroke. It is based on the blood oxygenation-level dependent (BOLD) response, which is an indirect measure of neural activity. Functional MRI can potentially detect age-related changes in brain function that can, in turn, help validate current cognitive aging theories, as well as generate new hypotheses. It can also highlight abnormal brain function in patients, which may aid in the diagnosis and evaluation of age-related conditions such as stroke, Alzheimer’s disease and other dementias.

Although fMRI has potentially important clinical applications, there are still several limitations associated with fMRI that impede its use as a standard clinical tool. First, the inside of the fMRI scanner is very small leading to claustrophobia in some subjects, and images are extremely sensitive to movement. As a result, cognitive tasks included in fMRI experiments must often undergo significant modifications to adapt to the scanner environment. Any movement required to complete a task must be limited as motion during the fMRI experiment causes distortions in the resulting images. In addition, cognitive tasks used in fMRI experiments are often required to be much longer than clinical cognitive assessments in order to obtain a strong brain signal associated with the task. These differences make it difficult to compare clinical cognitive assessments completed outside of the MRI with similar cognitive tasks performed within the scanner.

Another limitation of clinical fMRI is that there are many sources of noise in the resulting images. The brain signal measured is small relative to these noise sources (i.e. there is a small
functional signal-to-noise ratio, fSNR), which makes it difficult to measure neural activity related to a task. There are various post-scan signal processing programs that increase fSNR, which are referred to as preprocessing tools. Most experiments apply several preprocessing tools in a ‘pipeline’ in order to reduce the amount of noise in their data. However, steps in these pipelines are often chosen without evaluating if they are needed in the specific dataset, and are applied in a fixed manner across all subjects in the dataset. This results in preprocessing pipelines that may be sub-optimal for the majority of subjects included in an experiment.

Individually optimizing preprocessing pipelines on a subject-by-subject basis could increase the overall fSNR. Unfortunately, this is not practiced as it is time consuming and there is no standardized method to assess which preprocessing pipeline is optimal per subject.

Since clinical fMRI would be most useful on an individual subject basis, it could be beneficial to have techniques that individually optimize preprocessing pipelines. This places great importance on developing techniques to optimize preprocessing pipelines on an individual subject basis and to determine the effect on the fSNR relative to standardized fixed pipelines. Thus, the focus of this thesis is to optimize preprocessing pipelines on an individual subject basis in comparison to fixed pipeline optimization.

There are two goals. The first goal is to optimize fixed preprocessing pipelines as well as individual subject preprocessing pipelines on fMRI tasks that have been altered for clinical purposes, using an optimization procedure created by a colleague in the Strother laboratory (Churchill et al., 2011). These tasks are shorter and have different response modes than traditional fMRI tasks, and are comparable to cognitive batteries administered in the clinic. The second goal is to compare fixed and individually optimized pipelines and see how they affect our resulting cognitive interpretation of the brain patterns associated with the task. This
includes measuring if individually optimized pipelines result in stronger brain patterns via the measurement of significant increases in significant positive and negative voxels, and significant relationships between brain patterns and behavior.

**The following hypothesis was specifically tested:** individual subject pipelines, compared to fixed pipelines, will result in stronger and more reliable brain patterns for shortened cognitive tasks, specifically forced-choice episodic recognition.

The rest of the chapter provides the necessary background information to understand the research that was conducted for this thesis. It provides a brief explanation of how fMRI works, followed by detailed descriptions of fMRI stimuli, noises sources, preprocessing and analysis tools, fixed versus individually optimized preprocessing pipelines, and it concludes with a section describing episodic memory. The second chapter describes the methodology, the third chapter details the results from two different preprocessing experiments and the fourth chapter explains the significance of the research conducted as well as future directions.

### 1.1 Functional Neuroimaging

Functional neuroimaging is the process of using non-invasive imaging tools to map brain structure and function (Shibasaki, 2008). Several methods of functional neuroimaging exist including: fMRI, positron emission tomography (PET), electroencephalography (EEG), and magnetoencephalography (MEG). Each imaging method measures different parameters in the brain and thus results in different spatial and temporal resolution. EEG and MEG are characterized by high temporal resolution and poor spatial resolution, while fMRI and PET display high spatial resolution and poor temporal resolution (Poeppel & Krause, 2008).
EEG functions as follows: when neurons are active, they generate electrical currents that can be detected by electrodes placed on the scalp (Eliassen et al., 2008). MEG detects magnetic field oscillations from the electrical currents generated from active neurons. These are measured by magnetometers within the MEG array (Eliassen et al., 2008). MEG and EEG both have high temporal resolution. However, spatial resolution is compromised since MEG and EEG detect the signal on the scalp, on the order of square centimeters. Furthermore, no anatomical information is collected during a session. This makes it difficult to localize the source of neural activity, which is especially important when examining age-related conditions in the brain (Eliassen et al., 2008).

PET requires the injection of specialized radioactive tracers into a subject/patient and measures the metabolic response from the tracer’s uptake into brain cells and subsequent brain activity (Eliassen et al., 2008). PET has the ability to measure metabolic parameters relating to neural activity with relatively high spatial resolution (Poeppel & Krause, 2008). However, the use of radioactive tracers in PET exposes the patient/subject to low doses of radiation. As well, PET is characterized by poorer temporal resolution than fMRI (Eliassen et al., 2008).

Functional MRI takes similar measurements to PET without the use of radioactive tracers. FMRI measures brain activity by detecting changes in blood flow and blood oxygenation with relatively high spatial resolution. It is non-invasive and safe for subjects, with no known short-term or long-term side effects. Some of the disadvantages associated with fMRI include: the constrained scanner environment for claustrophobic subjects, poor fSNR, and complicated post-scanner processing. Although each neuroimaging method is subject to weaknesses, the accessibility and non-invasive nature of fMRI make it an attractive candidate for neuroimaging research, relative to the other methods. In addition, the spatial resolution of fMRI is higher than
the other methods and it is regularly used for clinical structural MRI scans. This largely contributes to the desirability of fMRI as a research neuroimaging method as well as a potential clinical method, overcoming the clinical drawbacks mentioned above.

There has been increasing interest in multi-modal imaging in the last few years, since there are various pros and cons to each technique. Two or more imaging techniques are acquired simultaneously, resulting in an overall increase in spatial and temporal resolution (Shibasaki, 2008). However, there is no widely accepted standard methodology regarding the integration of the various signals (Sato, Rondinoni, Sturzbecher, de Araujo, & Amaro, 2010), and more research is required before this type of imaging can be considered for clinical purposes.

1.2 Fundamentals of Functional MRI

Functional MRI is a complex neuroimaging method that detects brain activity based on the BOLD signal, and is carried out by a MRI scanner. FMRI involves acquisition of several scans over a period of time, often while a subject performs a task, and each scan is a three-dimensional image volume typically covering the whole brain. Each image volume is comprised of a matrix of ‘voxels’, a small volume of brain (or volume element) in three-dimensional space, similar to a pixel in two-dimensional space. The time-series of volumes are preprocessed and analyzed, to form statistical parametric maps (SPMs). The basic principles that form the fundamentals of fMRI are described below.

Functional MRI scans are obtained via magnetic resonance imaging (MRI). MRI uses radio-frequency signals to excite hydrogen atoms, which are abundant in the brain, and measures the energy emitted by the hydrogen atoms as they return to equilibrium, or alternatively, how long they take to return to their original energy state. Specifically, BOLD fMRI uses a physical
parameter called $T2^*$ based on transverse or spin-spin relaxation, $T2$. $T2$ is an MRI signal parameter resulting from local magnetic field variations within a tissue. When exposed to a magnetic field, hydrogen within a tissue will precess or spin around that axis until magnetization in the transverse-plane is lost as a result of local magnetic field fluctuations. $T2$ measures the decay rate at which transverse-plane magnetization is lost. $T2^*$ decays at a quicker rate than $T2$, as it measures the effect caused by inhomogeneities in the magnetic field. Since all tissues have different properties, as well as densities, each hydrogen atom will have a different $T2^*$, this is the phenomenon underlying BOLD contrast. Hydrogen atoms with larger field inhomogeneities (i.e. tissues or molecules with paramagnetic properties) will have decreased image intensity because the hydrogen atom returns to its original energy state at a faster time (Huettel, Song, & McCarthy, 2009). BOLD fMRI is an indirect measure of neural activity in the brain, as outlined below.

There are two main types of cells in the brain: neurons and glial cells. Neurons are the central processing unit of the brain, while glia serve biochemical purposes. Neurons are comprised of dendrites, the cell body (soma), axons and axon terminals. The dendrite is the receiver terminal responsible for taking in incoming information; the soma then takes this information from the dendrites and passes it to the axon hillock, which controls whether a neuron fires an action potential; the axon extends from the cell body and sends the neural impulse to the axon terminals which then transmits the neural signal to other neurons. Neurons communicate with each other via the release of neurotransmitters at the synapse, which is the space between the axon terminal of one neuron and the dendrite of another.

The following outlines the mechanism by which neurons communicate, via action potentials. At resting state, the soma of a neuron has a negative charge relative to the outside of the neuron,
which is positively charged. Ion channels maintain this charge differential by actively responding to changes in charge inside and outside of the neuron. Positively charged potassium ions (K+) outside the neuron are attracted to the negative charge within and readily cross the cell membrane, while sodium (Na+) ions leak into the neuron in small amounts, as they are unable to cross the neuron’s membrane. Eventually there is more K+ inside the cell than outside and K+ gets pushed outside until the neuron is at equilibrium. A sodium-potassium pump in the cell membrane maintains the negative charge inside the neuron by pumping Na+ outside, and K+ inside of the cell. When an action potential occurs, the neuron becomes depolarized due to Na+ channels opening, releasing an influx of positive charges inside the cell. The membrane potential is now positive relative to the outside of the neuron. The action potential propagates down the axon triggering the release of neurotransmitters, from the axon terminals into the synapse, stimulating ion channels in the dendrite of the next neuron. This response is repeated in the next neuron, thus propagating the cycle.

The metabolic resources of a neuron are depleted during this process. Neurons require oxygen and glucose for energy, which is supplied by local blood flow (Heeger & Ress, 2002). Red blood cells contain molecules that supply oxygen to cells, called hemoglobin. When there is neural activity, oxygen is consumed from local blood flow, causing a decrease in the concentration of oxygenated hemoglobin in the blood and an increase in deoxygenated hemoglobin. As a result, local blood flow is recruited to replenish the neurons’ oxygen supply, causing a large increase in oxygenated hemoglobin relative to deoxygenated hemoglobin (an oversupply of oxygenated blood) and an increase in local blood volume (D'Esposito, Deouell, & Gazzaley, 2003; Heeger & Ress, 2002). The concentration of oxygenated hemoglobin to deoxygenated hemoglobin subsequently returns to equilibrium.
Oxygenated hemoglobin and deoxygenated hemoglobin have different magnetic properties; thus detecting the changes in concentration between the two hemoglobin forms is essential for BOLD fMRI contrast. Oxygenated hemoglobin is diamagnetic, which means it is less reactive to the magnetic field (i.e. less field inhomogeneities) and therefore has a longer $T2^*$. Deoxygenated hemoglobin, on the other hand, is paramagnetic, which means that in a magnetic field it creates more inhomogeneities causing quicker transverse decay, i.e. a shorter $T2^*$ (Heeger & Ress, 2002). A neural event causes a local increase in blood flow, which decreases the concentration of deoxygenated hemoglobin in the microvasculature in the surrounding region, increasing the ratio of oxygenated hemoglobin to deoxygenated hemoglobin, thus increasing the BOLD contrast in that region (Ogawa, Lee, Kay, & Tank, 1990).

In summary, the BOLD signal is based on the hemodynamic response. When a neural event occurs there is an increased metabolic demand causing oxygen and glucose consumption, as neurons do not keep internal reserves of either. This causes the ratio of deoxygenated to oxygenated hemoglobin to increase in the surrounding vasculature, i.e. the BOLD signal decreases. Next, blood flow is recruited to the recently activated region to replenish the neuron’s supply of oxygen and glucose. This occurs at a delay of around 2 seconds after the stimulus onset (Heeger & Ress, 2002). The surge of blood flow causes an oversupply of oxygen in the active brain region, leading to an increase in the BOLD signal due to an increasing concentration of oxygenated hemoglobin, which peaks around 5 seconds after the stimulus onset. The BOLD signal subsequently decreases and returns to baseline many seconds after neural activity has finished (Huettel et al., 2009).
1.3 Functional MRI Stimuli

The basis of functional MRI is that it measures brain activity, under predefined conditions. For example, subjects often view stimuli and/or perform a particular task during a scanning session. Experiments can be carefully designed to examine a particular cognitive theory or phenomenon of interest. In addition, functional MRI has the potential to be useful in the clinic for diagnosing and evaluating diseases that affect cognitive functioning. With clinical use in mind, tasks need to be carefully designed and altered to mimic cognitive batteries used in the clinic in order to provide a similar comparison to standard neuropsychological tests used in the clinic. A key aspect of this is trying to adapt clinical neuropsychological tests to the scanner environment with minimal change. The following section describes the clinical relevance of the encoding and recognition tasks used in this study, the various methods of response available in the fMRI environment versus outside of the scanner, and the key components of task design in fMRI experiments.

1.3.1 Encoding and Recognition Tasks

Encoding and recognition are tasks commonly used in fMRI experiments to study memory storage and retrieval. Encoding is the processing and storage of stimuli for subsequent retrieval. Specifically, semantic encoding involves processing the meaning of stimuli, such as whether a presented picture represents a living or non-living object (Langley & Madden, 2000). Recognition is the process of retrieving and correctly identifying previously stored information. Both tasks, when used together, are extremely important in the study of age-related impairments, which almost always affect memory. While paper neuropsychological tests are helpful at diagnosing and tracking age-related memory disorders, fMRI has the capability of determining
the localization of the subject’s memory issues by examining neural activity. A major advantage of fMRI over clinical testing is that it has the potential to identify the brain mechanisms underlying neuropsychological disorders, the neurological and functional implications in the patient, which can lead to better modes of therapy and rehabilitation customized to each patient.

Semantic encoding, in the context of this study, is the active viewing of pictures containing living and non-living entities while naming them out loud. The task itself is important in detecting cognitive impairment, as it is a measure of language retrieval. This type of task can help differentiate healthy aging from cognitive impairment and dementia (Hachinski et al., 2006). The recognition task is a direct measure of memory and can be especially sensitive to vascular cognitive impairment (Hachinski et al., 2006).

1.3.2 Neuropsychological Tests versus Functional MRI Response Modes

Neuropsychological tests need to be adapted for use in the fMRI to accommodate the restricted scanner environment. This is mainly due to the confined space in a scanner and the fact that any movement, whether caused by speech or body motion, results in noisy and artifactual fMRI images (described below). Therefore, the mode of response in fMRI tasks is almost always altered from the original psychological test it is based upon in order to minimize motion in the scanner and the associated noise.

One disadvantage associated with altering tasks for the scanner environment is that brain activity for semantic processing has been shown to differ based on how responses are given (Jennings, McIntosh, Kapur, Tulving, & Houle, 1997). Unfortunately, few studies have
examined this effect, providing minimal information about the way brain activity changes due to response mode. However, this finding has direct implications for all fMRI experiments, which alter the way classic neuropsychological tests are administered. For example, traditional cognitive language–related tasks, such as overt semantic encoding, rely on overt verbal responses for speech production. However, there are many problems associated with speaking aloud in the scanner and covert (silent) responses are often recruited as an alternative. Unfortunately it has been shown that overt verbal responses cannot be replaced by covert responses, since subject performance cannot be assessed or measured via covert responses and they result in different patterns of activation (Barch et al., 1999; Jennings et al., 1997; Palmer et al., 2001; Shuster & Lemieux, 2005).

The recognition task is classically administered as a paper and pen test, or more recently as a tablet computer task that can be completed at a stroke patient’s bedside (i.e. Baycrest’s Stroke and Cognition Clinic, BSCC). Functional MRI studies have adapted this task to the scanner environment by using button box responses instead of a pen response to minimize movement, allowing the accurate measure of response time by a computer linked to the button box. However, similar to overt verbal encoding, this can significantly change the pattern of brain activity (Jennings et al., 1997).

An fMRI compatible tablet and stylus has been created that can be effectively implemented in fMRI experiments. This is an ideal alternative to the button box as it closely mimics the response mode used for many of the clinical neuropsychological tests while conforming to the requirements of the fMRI environment (Tam, Churchill, Strother, & Graham, 2011). The tablet has a screen and a pen, and rests across a subject’s lap. Subjects can easily be trained to use this device prior to a scanning session, without looking at their hand, in order to complete a task in
the fMRI scanner. While there are still other modifications needed with respect to adapting tasks to the fMRI environment, the tablet narrows the gap between clinical and fMRI neuropsychological testing response modes.

1.3.3 Functional MRI Task Design

There are several key components to designing a functional MRI task that attempts to examine a cognitive process of interest. Classically designed fMRI tasks include ‘task’ and ‘baseline’ conditions. The two can be contrasted against each other during analysis in an attempt to isolate the corresponding brain pattern. Task and baseline conditions can be arranged in two different ways: block and event-related designs. The amount of time within each condition, in what order they are arranged, and what design is used, all play a big role in the quality of results that come out of the analysis. There is no consensus as to what the best type of task design is, i.e. either block or event-related designs, and selection of one over the other is dependent on the goals of the study.

Block designs divide the task and baseline conditions of an experiment into alternating blocks containing repeated events of each. For example, a typical block experiment could have a subject overtly naming pictures for 30 seconds (task) and passively viewing fixation crosses for 30 seconds (baseline). Event-related designs, on the other hand, administer the task as short ‘events’ often interspaced with random timing. For example, an event-related design could have a subject overtly naming a picture for a minimum of 1 second with a variable inter-stimulus interval (ISI) averaging 2 seconds, although ISIs can be constant as well.

There are several advantages to using block designs, the first being higher signal detection and statistical power. This is because the blocked activity creates a larger contrast between brain
states (task and baseline); there are large differences in BOLD signal between the brain states, leading to maximal variability between the blocks, resulting in the increased detection of brain activity (Chee, Venkatraman, Westphal, & Siong, 2003; Friston, Zarahn, Josephs, Henson, & Dale, 1999; Huettel et al., 2009). In addition, block designs are less sensitive to the model chosen for the hemodynamic response function (HRF), which is used to estimate the hemodynamic response in fMRI analyses (Carter, Heckers, Nichols, Pine, & Strother, 2008). However, event-related designs have the advantage of detecting BOLD temporal dynamics better than block designs. As well, event-related designs are less sensitive to task-correlated motion and are tuned for post-hoc trial sorting (Huettel et al., 2009). Since there are advantages to both types of task designs, the decision to use one method over the other depends on the goals of the experiment. Event-related designs should be implemented in experiments examining the brain’s response to a particular stimulus occurring over a short interval. Block designs, on the other hand, are effective in experiments studying brain states and networks in response to a particular task, over a longer period of time (Huettel et al., 2009).

The amount of time assigned to task components is important with respect to studying brain activation. As mentioned previously, fMRI is characterized by low functional SNR. One way to increase fSNR is by acquiring more data, i.e. increasing the length of task or using block designs, which allows more trials per unit time than event-related designs. However, increasing the duration of the task is difficult in clinical experiments as it is hard for patients with age-related disorders to concentrate for long periods of time as well as comfortably lie in the scanner. Consequently, additional methods to increase fSNR or remove noise in fMRI scans are needed.
1.4 Functional MRI Noise Sources

There are several types of fMRI noise, which can be divided into two general categories: scanner and subject noise. Noise from the scanner is caused by variability in the MRI hardware. This can be caused by magnetic gradient instabilities or scanner drift, where the resonance frequency of protons is altered by the variations in field strength (Huettel et al., 2009).

There are many sources of subject-related noise, many of which are introduced into the data by the response mode used in the experiment. Subject noise sources include head motion, motion related to the experimental paradigm, overt speech, and respiratory and cardiac cycles. Subject-related noise is complicated, making it difficult to accurately identify and correct since the major sources of noise will vary subject by subject.

Head motion is a large source of unwanted variability in fMRI data. When fMRI volumes are acquired, it cannot be assumed that each voxel has a fixed anatomical location, i.e. that the head is fixed in position. Subject head motion is unavoidable since subjects often get restless during the scan and they are required to give responses. These movements can cause distortions in the anatomical localization of functional activations as activation in one cortical location may be shifted to different voxels in space and time (Munhall, 2001). These shifted voxels would appear as active even though they are actually inactive, which can lead to incorrect interpretations and conclusions.

In addition, there is often motion correlated with the task paradigm, i.e. task-correlated motion (TCM). This type of motion affects the resulting brain pattern and can often cause severe artifacts in the images. This is because most fMRI paradigms require responses, and as careful as the subject is to stay motionless while responding, motion corresponding to the task
paradigm/timing occurs. This also appears as false activation, especially around the edges of the brain (Johnstone et al., 2006).

Overt verbal responses in fMRI experiments also cause significant noise in fMRI results. When a subject speaks aloud, they are required to move their mouth, jaw, tongue and head, along with several muscles. These movements lead to changes in the magnetic field, which induce artifacts in the resulting images within the scanners imaging volume, or field of view (FOV), and outside the FOV (Birn, Bandettini, Cox, Jesmanowicz, & Shaker, 1998). Motion within the FOV is likely to cause data from the same voxel to be acquired at different cortical locations over the duration of the fMRI scan, similar to the effect of head motion, causing uncorrected images to have false positive or negative activations. These artifacts often occur at the edges of the brain, in the inferior regions or near the ventricles (where there is a change in tissue type); this is because small movements cause large changes in fMRI signal intensity in these regions and thus the voxel location moves to a different cortical location or tissue type (e.g. CSF to brain) (Barch et al., 1999). This type of overt-verbal motion can easily be corrected via standard motion correction techniques.

Motion outside the FOV, due to overt speech, often induces changes in the magnetic field causing the image to be warped. These types of artifacts are so severe that they cannot be corrected using standard preprocessing methods (Birn et al., 1998). Specifically, these artifacts arise during speech from head or vocal tract movement, and volume changes in the nasal cavity, sinus cavities or pharynx. These changes are what cause distortions in the magnetic field which induce the artifacts (Barch et al., 1999; Birn et al., 1998; Munhall, 2001).
Another source of unwanted variability in fMRI data is physiological noise. This comes from the respiratory and cardiac cycles. One major problem is that these cycles negatively influence the BOLD signal. Respiratory cycles can cause up to 10% fluctuations in the BOLD signal and the cardiac cycle is even worse with the potential to cause up to 40% fluctuations in the BOLD signal from baseline (Birn, Diamond, Smith, & Bandettini, 2006; Dagli, Ingeholm, & Haxby, 1999). Uncorrected physiological noise can overpower the BOLD signal and introduce artifacts into fMRI results when uncorrected (S. C. Strother, 2006).

1.5 Functional MRI Preprocessing Techniques

Noise and artifacts in fMRI results are currently an unavoidable part of the acquisition process. Preprocessing pipelines and techniques have been created to correct for unwanted variability in fMRI datasets. However, there is no gold standard in preprocessing pipelines. There are several versions of each preprocessing step and several ways each step can be applied. As well, preprocessing steps can interact with each other and sometimes introduce more noise into the dataset. When needed, a preprocessing technique is beneficial to a dataset, but when it is not needed it can be detrimental to the fMRI results. Commonly used preprocessing steps are: physiological noise correction, slice timing correction, motion correction, smoothing, detrending, motion parameter regression, and between subject alignment. The following section describes the preprocessing steps above and in what scenarios these steps can be beneficial to the data.

1.5.1 Physiological Noise Correction

As mentioned above, cardiac and respiratory cycles cause a significant amount of noise in fMRI images (Glover, Li, & Ress, 2000). The most effective way to remove this type of noise is by
applying physiological noise correction (PNC). PNC usually requires the collection of respiratory and cardiac cycles during the scanning session, although many studies do not obtain this information even though it is an essential part of the preprocessing technique. Analysis of Functional Neuroimage’s (AFNI) (Cox, 1996) program called RETROICOR models physiological noise per voxel using the respiratory and cardiac cycles recorded by the scanner, and then subtracts this signal from the time-course of each voxel (Glover et al., 2000). Another method called PHYCAA (PHYsiological correction using Canonical Autocorrelation Analysis), which was created by a lab member, has been shown to significantly improve model reliability in block and event-related designs, compared to RETROICOR and no PNC. PHYCAA measures physiological effects within the data and removes them (Churchill et al., 2012). A major advantage of this method is that it is data-driven and does not require cardiac and respiratory cycle measurements from the scanner. However, this method was not implemented in the data for this thesis as it was not available and published until after the preprocessing pipelines were applied.

1.5.2 Slice Timing Correction

Functional MRI volumes are usually collected every 2-3 seconds (i.e. during the repetition time, TR). Within this time frame several slices are collected, the last one collected 2-3 seconds after the first, unless interleaved slice acquisition is used where slices are collected in alternating order (e.g. even then odd slices), preventing cross-slice excitation where the preceding excitation pulse pre-excites the adjacent slice (Huettel et al., 2009). However, interleaved slice acquisition still results in timing discrepancies, as the last two slices will be collected 1-1.5 seconds apart. Since task timing is essential to fMRI analysis, these delays may create issues in the interpretation of results because the time series will be inaccurate. Slice timing correction
(STC) was created to eliminate this problem by interpolating between volumes in adjacent TRs in the same slice and voxel. This method spatially aligns across time in order to predict the signal in all slices had all the slices been obtained at the exact same time (S. C. Strother, 2006). In addition, STC has been shown to be unnecessary in many block design studies (S. C. Strother, 2006). AFNI’s 3dTshift is a useful tool for STC where the preferred type of interpolation can be chosen.

It has been recommended that slice timing correction come after RETROICOR if the study employs interleaved slice acquisition as performing slice timing correction first alters the physiological timing impeding effective PNC (Jones, Bandettini, & Birn, 2008).

1.5.3 Motion Correction

As previously mentioned, head motion during fMRI acquisition causes artifacts in the resulting images. Motion correction (MC) uses one of a variety of algorithms that have been developed to register all volumes in a time series to one volume within that series. The reference volume is a consistent volume selected across each time series and is usually selected arbitrarily by the researcher, i.e. the same point in each time series is chosen for all subjects. The reason that one point in the time series is used across all subjects is to have consistent processing across all subjects. As well, there is no research showing that the specific volume chosen as the reference volume impacts the motion correction results. However, arbitrary volume selection can often result in suboptimal motion correction leading to incorrect conclusions about the amount of motion in the data, since subject motion is measured by the distance, in mm, a volume’s voxels have to move in order to align to the reference volume, e.g. the maximum displacement of the voxel that moves the most during realignment. Maximum displacement is calculated from the
six motion parameter estimates measured across all directions of the volume (right/left, anterior/posterior, inferior/superior, roll, pitch, and yaw).

The MC process corrects for motion with rigid-body transformations, which are calculated and applied to each volume relative to a reference scan. Various cost-functions exist to calculate optimal transformations on the volumes (S. C. Strother, 2006). AFNI’s 3dvolreg program is commonly used for motion correction. It calculates the maximum displacement across the six measured motion parameters (in mm) across all time points. This is an indicator of how much overall motion occurred throughout the duration of a particular fMRI task run. Head motion greater than 1-2mm is often considered to be detrimental to the fMRI images and volumes of motion greater than the set threshold should be discarded (S. C. Strother, 2006).

1.5.4 Spatial Smoothing

This preprocessing method has the potential to increase fSNR by removing unwanted variance in volumes in return for losing some spatial resolution. Functional MRI volumes are smoothed by recalculating the value of any voxel in the volume as a weighted sum of neighboring voxels. There are a variety of filters or kernels available to chose the weights. However, while spatial smoothing increases fSNR, it ‘blurs’ the resulting volumes, which means some spatial resolution is lost. Nevertheless, the benefits of smoothing far outweigh the loss in resolution (Parrish, Gitelman, LaBar, & Mesulam, 2000; S. Strother et al., 2004). A Gaussian kernel 1-2 voxels in full-width-at-half-maximum (FWHM) is commonly used as it has been shown to be the optimal size because it increases signal detection power (S. C. Strother, 2006).
1.5.5 Temporal Detrending and Motion Parameter Regression

Local changes in blood flow or spontaneous neuronal activity can cause low-frequency fluctuations in the fMRI signal (Tanabe, Miller, Tregellas, Freedman, & Meyer, 2002). These can be difficult to remove. One way of removing these low frequency trends is by regressing them out (Kay, David, Prenger, Hansen, & Gallant, 2008). Modeling a voxel’s time series by fitting and regressing Legendre polynomials, along with low frequency trends fitted with the experimental paradigm, has been shown to be effective at improving fSNR in fMRI data (Kay et al., 2008). Detrending works by modeling a voxel’s time series and then by regressing low frequency trends out of the modeled signal. AFNI’s 3dDetrend can be used to perform this preprocessing technique.

Motion parameter regression (MPR) is implemented in the detrending process as a separate option. It is effective for removing head motion correlated temporally with the fMRI signal but trades off with a reduction in fSNR. This method uses motion parameter estimates (MPEs) from the motion correction step as additional regressors. There are conflicting results in the literature about MPR. While some studies have found MPR to reduce activation strength in block design studies (Johnstone et al., 2006; Ollinger et al., 2009), others have found it caused a reduction in unwanted variability within fMRI datasets (Evans, Todd, Taylor, & Strother, 2010; Lund, Norgaard, Rostrup, Rowe, & Paulson, 2005). Churchill et al. (2011) found MPR was detrimental in a low-motion data set, but also found MPR in conjunction with PNC benefited subjects with increased head motion (Churchill et al., 2011). The current importance of MPR and when it should be used remains unclear.
1.5.6 Between Subject Alignment

Every single brain is different; they vary in size, scaling and even shape. When subjects are analyzed as a group, individual brain differences must be accounted for so each voxel represents the same space across subjects. Between-subject alignment is a preprocessing tool that registers all subject’s fMRI volumes to the same template. While standardized templates exist (e.g. Montreal Neurological Institute (MNI) and Talairach), creating group specific templates for each study has the potential to reduce the amount each subject’s SPM is distorted in order to get into a common space (S. C. Strother, 2006). FSL’s (fMRI analysis package by the FMRIB group, www.fmrib.ox.ac.uk/fsl/) FLIRT can be used to complete the alignment process as well as to create group specific templates. FLIRT can be used to calculate and apply transformations to the SPMs relative to the template in two stages: 1) a transformation from the subject’s functional volume to its own structural image and 2) a transformation from the structural image to the template (Jenkinson, Bannister, Brady, & Smith, 2002). These two transformations are combined and applied to the functional images.

1.6 Fixed versus Individual Pipeline Optimization

Functional MRI preprocessing pipelines can be fixed across subjects or individually optimized. As preprocessing techniques are increasingly examined and developed, a growing number of studies have found conflicting results with respect to the effectiveness of these tools. Some studies have found that preprocessing techniques interact with each other (Churchill, Abdi, & Strother, 2010), which impacts the results. Furthermore, various studies have found that the effect of preprocessing varies depending on the dataset (Churchill et al., 2011; S. C. Strother, 2006). The reason there are differing results regarding preprocessing steps is that each study
uses datasets comprised of differing tasks and subjects. Since it has been shown preprocessing steps vary depending on the dataset, consistent results should not be expected.

Instances of interactions of preprocessing techniques affecting the results have been observed. For instance, a study by a fellow lab member, Churchill et al. (Churchill et al., 2010), studied the effects of ordering of PNC, STC and MC. The order of PNC and STC was shown to have a significant effect that was subject dependent. Subjects with increased variability in their cardiac rates were optimized by STC coming before PNC (Churchill et al., 2010). Furthermore, detrending has been shown to interact with spatial smoothing in addition to polynomial order (Kay et al., 2008; Shaw et al., 2003; Tanabe et al., 2002). These findings illustrate the importance of evaluating the order in which preprocessing steps are applied to the data, and the specific steps to include per subject due to between subject variability.

Certain preprocessing steps have been found to significantly reduce noise in some datasets and create more noise in others. For instance, motion correction, which causes significant reductions in motion-related noise, can also introduce severe artifacts into the data (B. Kim, Boes, Bland, Chenevert, & Meyer, 1999) perhaps due to the specific algorithms chosen which can cause activation bias in the data (Orchard, Greif, Golub, Bjornson, & Atkins, 2003). In other words, preprocessing steps should not be applied to datasets as a precaution; they should only be implemented in datasets when needed (S. C. Strother, 2006).

Shaw et al. (2003) demonstrated that individually optimized preprocessing pipelines can lead to increased fSNR, by determining which preprocessing tools result in increased prediction and reproducibility (two measures which reflect fSNR) (Shaw et al., 2003). Results from Churchill et al. (2011) indicate that individually optimized subject pipelines lead to significantly higher
reproducibility, reflecting increased fSNR, than fixed preprocessing pipelines indicating that there are subject-dependent effects in preprocessing choices. This is important because significant brain patterns may be revealed through the use of individual subject optimization that are otherwise not revealed via fixed preprocessing pipelines (Churchill et al., 2011).

The above section supports the use of individually optimized preprocessing pipelines. Many of the methods have been shown to significantly correct for unwanted variability in some datasets, and introduce more noise in other datasets. Thus, if subjects’ pipelines are optimized on an individual basis they can maximize the overall fSNR while reducing the introduction of extra error by applying unnecessary preprocessing steps, resulting in strong, reliable brain patterns. This is important because it supports clinical fMRI, which would ideally require individually optimized pipelines.

1.7 Episodic Memory and Neuroimaging

The brain is a dynamic, variable and complex organ, differing across individuals. While fMRI can tell us what regions and possible networks are activated during cognitive processes, it is the relationship between these regions and cognitive behavior that tells us what the activations may mean and why there are individual differences. Two behavioral measures are often obtained with recognition: reaction time and accuracy. Testing for a relationship between these two measures with networks and regions of activation per subject can indicate why there are certain patterns of activation, as well as why there are individual differences in activation and behavior. The following section describes declarative memory (specifically episodic memory), its associated brain regions and relation to behavior.
Semantic memory is defined as the conscious recollection of general knowledge and factual and conceptual information, while episodic memory is the conscious recollection of events that were experienced (Burianova & Grady, 2007; Tulving, 1987). Both types of memory have commonalities and interact with each other. For example, episodic memories can access semantic memories. There are two different models that describe how they interact: 1) multiple memory systems that interact with shared executive and attentional processes but are anatomically independent; or 2) a common anatomical memory network (Burianova & Grady, 2007; L. Nyberg, Forkstam, Petersson, Cabeza, & Ingvar, 2002; Rajah & McIntosh, 2005; Tulving, 1987). Semantic and episodic memory, along with autobiographical memory, the conscious recollection of a personal event, makes up declarative memory (Burianova & Grady, 2007).

Episodic memory, in an experimental context, can be evaluated by learning and retrieving lists of stimuli (Grady, Springer, Hongwanishkul, McIntosh, & Winocur, 2006). It is most often studied by using tasks that require a subject to recall or recognize information presented within the study (Buckner, Koutstaal, Schacter, Wagner, & Rosen, 1998). These stimuli often are encoded in different conditions, such as semantic and perceptual conditions. In addition, there are two processes that comprise episodic memory during recognition: recollection and familiarity (Yonelinas, 2001). Recollection has been found to require the retrieval of qualitative information about previously studied events. Familiarity, on the other hand, has been found to require accepting studied items as being familiar (Yonelinas, 2001). While some episodic memory studies evaluate these two parameters separately, others do not with the expectation that changes in neural responses during the recognition task can reflect either recollection or familiarity (L. Nyberg et al., 2000). “Remember/know” paradigms are often used in studies of
episodic retrieval in order to distinguish between recollection and familiarity, where remembered items reflect recollection and known items reflect familiarity (H. Kim, 2010; Spaniol et al., 2009).

The neural correlates of successful encoding and recognition are related. Rugg et al. (2008) combined two frameworks into a single model to describe the relationship between encoding and recognition: transfer-appropriate processing (TAP) and cortical reinstatement hypothesis (Morris, Bransford, & Franks, 1977; Norman & O'Reilly, 2003). The model summarizes the relationship between encoding and recognition as follows (Rugg et al., 2008):

1. Cortical regions are activated during the encoding of items
2. The hippocampus encodes and stores this pattern of activity
3. Original pattern of activity is partially reinstated if successful retrieval occurs during presentation of the retrieval cue, which feeds forward to the hippocampus.
4. Due to overlap between the successful retrieval brain pattern and the stored pattern of activity, the hippocampal representation is reactivated leading to full reinstatement of activity at the cortical level

Left prefrontal regions work with posterior regions while forming memory representations during encoding (L. Nyberg et al., 2000). The right anterior and dorsolateral prefrontal cortex (PFC), and medial temporal regions appear to be brain regions implicated in episodic memory retrieval (Rajah & McIntosh, 2005). Buckner et al. (1998) found activations in the right anterior PFC, the bilateral anterior insular regions and a left dorsal prefrontal region (Buckner et al., 1998). Burianova et al (2007) found that episodic retrieval specifically activated the right middle frontal gyrus, dorsomedial PFC, left precuneus, and superior parietal lobule in comparison to semantic and autobiographical memory (Burianova & Grady, 2007). Most
studies have found that the right frontal cortex regions are preferentially involved in episodic retrieval processes while the left frontal cortex regions are preferentially involved in episodic encoding; this formed the hemispheric encoding/retrieval asymmetry (HERA) model (L. Nyberg, Cabeza, & Tulving, 1996; Tulving, Kapur, Craik, Moscovitch, & Houle, 1994).

While the model by Rugg et al. (2008) postulates that the hippocampus is active during retrieval (as well as other models, e.g. Moscovitch & Nadel, 1998), the evidence surrounding this phenomenon is contentious (Ranganath, 2010; Rugg et al., 2008). The disparity in results over hippocampal activations could be due to differences in recollection versus familiarity. Neuroimaging studies have shown that activity in the hippocampus and parahippocampal cortex increases during encoding and recognition of recollected items, i.e. high levels of retrieval (Ranganath, 2010). Many studies have found hippocampus activations during conditions with higher levels of retrieval or recollection (L. Nyberg, McIntosh, Houle, Nilsson, & Tulving, 1996; Rugg, Fletcher, Frith, Frackowiak, & Dolan, 1997; Schacter, Alpert, Savage, Rauch, & Albert, 1996). However, activity in these regions does not increase or decrease based on item familiarity. On the other hand, activity in the perirhinal cortex, the anterior part of the parahippocampal gyrus, has been found to relate to item familiarity, not item recollection (Davachi, 2006; Ranganath, 2010). Thus, studies that employ recognition without adding a component to evaluate recollection and familiarity may have conflicting results.

The roles of the perirhinal cortex, parahippocampal cortex and hippocampus have been found to correspond to each other, while differentially contributing to recollection and familiarity processes. The perirhinal cortex is thought to contribute to familiarity processes via item memory. The parahippocampal cortex is thought to contribute to context memory, it receives input from areas involved with processing visual objects (‘what’) and spatial information
(‘where), i.e. both familiarity and recollection processes. Finally, the hippocampus is thought to be involved with recollection processes by binding item and context associations. (Eichenbaum, Yonelinas, & Ranganath, 2007; Ranganath, 2010; Spaniol et al., 2009).

Some of the regions mentioned form the ‘task positive network’ (TPN), a network of regions activated while performing a task, i.e. regions involved with cognitive control (Fox et al., 2005; Grady et al., 2010). This includes the inferior frontal gyrus, dorsolateral PFC, sensorimotor areas and some inferior parietal regions (Grady et al., 2010). The TPN is thought to be anti-correlated with default mode network (DMN), a network of regions active during internal cognitive processes. The DMN includes the posterior cingulate and medial parietal regions, inferior parietal lobe, medial prefrontal cortex, superior frontal gyrus, anterior portions of inferior temporal cortex, medial temporal cortex and medial cerebellum (Grady et al., 2010; Raichle et al., 2001; Shulman et al., 1997; Toro, Fox, & Paus, 2008). In addition, it has been found that the DMN is comprised of several smaller sub-networks (Andrews-Hanna, Reidler, Sepulcre, Poulin, & Buckner, 2010). The DMN and TPN arise in many fMRI studies, when task components are contrasted with a low-level baseline (such as fixation), causing neural networks to switch ‘on’ and ‘off” during the task, i.e. relative signal increases or activations, and reductions often referred to as “deactivations”. Since the task component is contrasted against the baseline component, the resulting contrast components usually show activations in TPN regions and deactivations (or anti-correlation) in the DMN regions.

Relating behavioral data during a scanning session to such SPMs (e.g., task-baseline contrast components) has provided greater meaning to the interpretation of brain patterns. Left medial temporal activity has been found to relate to retrieval success in studies of episodic memory (L. Nyberg et al., 1996). Increased activity in superior temporal and medial parietal regions has
been found to be associated with memory performance (Heckers et al., 1998; Kapur et al., 1995; L. Nyberg et al., 1995; Rugg et al., 1997). Right dorsolateral PFC activity as well as left medial temporal activity has been found to relate to the presentation of old versus new stimuli (Duzel et al., 1999; Tulving et al., 1994). In addition, the inferior parietal cortex (supramarginal gyrus and angular gyrus) (VPC) and posterior midline region (posterior cingulate cortex and precuneus) (PMR) have increased activity during retrieval success compared to retrieval failure, as well as encoding failure compared to encoding success (Daselaar et al., 2009). The retrieval success network (RSN) has been coined and includes the following regions: medial PFC, posterior parietal cortex and posterior midline region, which comprises the precuneus, posterior cingulate and retrosplenial cortex (Buckner, Andrews-Hanna, & Schacter, 2008).

Other studies have examined reaction time and episodic recognition, specifically familiarity versus recollection. While there is conflicting literature on this topic, many studies have reported recollection, i.e. remember versus know judgments, is performed faster than familiarity (Gimbel & Brewer, 2011; Spaniol et al., 2009; Wixted, 2009). On the other hand, other studies have found that these two processes had similar reaction distributions when confidence was equated (Rotello & Zeng, 2008). In addition, reaction time has been shown to influence the DMN in a remember-know paradigm: as expected the DMN was suppressed during the task, with increased suppression during slower remember and know trials. However, while the hippocampus showed greater activity during slower recollection trials, reaction time had no influence on hippocampal activity during familiarity trials (Gimbel & Brewer, 2011).

The above studies utilized behavioral data collected in the scanner and related it to the pattern of activation produced after fMRI data analysis. Evaluating inter-individual differences, especially in single subject analyses, can provide valuable explanations as to why there are certain patterns
of activation or behavior, as well as support cognitive theories. In addition, many functional imaging studies employ subtraction paradigms in their analysis in order to compare the neuronal differences between two conditions. Few studies have examined how reaction time relates to regional differences in episodic memory via fMRI (Gimbel & Brewer, 2011), and none have examined how taking the difference in reaction time between conditions affects the results (e.g., (McNab et al., 2008)). Since a subtraction analysis measures the difference between two conditions, a behavioral variable that measures the change in behavior between the two conditions could potentially lead to increases in the strength of brain and behavior relationships.

The results of studies examining episodic recognition greatly depend on the task paradigm. As previously mentioned, response mode (Jennings, McIntosh, Kapur, Tulving, & Houle, 1997), task length and task design can influence results. Another factor influencing results is task difficulty. Memory strength is related to regions of activation. For example, high levels of retrieval are associated with recollection and activity in the hippocampus (Ranganath, 2010). Memory tasks without high levels of retrieval might show patterns related to familiarity, i.e. if there is an absence of item and context associations (Ranganath, 2010). The recognition task implemented in this study, forced-choice episodic recognition of pictures of objects, did not require high levels of retrieval, i.e. item and context associations were not needed to successfully complete the task. Thus, it was hypothesized that there would be little or no hippocampal activity in the recognition results, and the pattern would primarily reflect familiarity-based memory processes. In addition, due to the task design, it was also hypothesized that the recognition results would produce regions related to visual processing and object recognition, such as the fusiform gyrus (Garoff, Slotnick, & Schacter, 2005; Simons, Koutstaal, Prince, Wagner, & Schacter, 2003).
The above sections provide the necessary background to understand the subsequent chapters and results. The descriptions of fMRI, fMRI task design and preprocessing help understand the methodology described and implemented in the experiment (Chapter 2). The descriptions of fMRI noise sources, and memory setup the framework to understand how and why the results are presented and interpreted in Chapters 3 and 4.
2 Methods

The following section describes the experiment and procedures implemented in this study. First there is a description and explanation of the tasks used in the experiment, followed by the subject selection procedure. After that the entire fMRI session is described: before, during and after the fMRI scan. Finally, the extensive preprocessing and analysis procedures are described including full pipeline optimization and data analysis. The last section of the chapter describes a preliminary motion correction optimization study, the results of which are discussed in the future directions section.

2.1 Task Design

Tasks for this study were chosen to fulfill two goals: the task’s ability to clinically evaluate stroke and the task’s ability to be adapted to the scanner environment. In the end, four tasks from BSCC were chosen: overt semantic encoding (as introduced in the introduction), trail making task, episodic recognition (as described in the introduction) and a sustained attention to response task (SART). All tasks had to be adapted from the clinical environment to the scanner environment with minimal changes so that each task served as a similar cognitive probe inside and outside of the scanner. In order to evaluate within session reliability and between session reliability, each task was administered twice (with different trials) within each fMRI session and each subject was scanned twice approximately 3 months apart. This thesis focuses on a subset
of the recognition task data only. However, the preliminary motion correction optimization study was applied on both the encoding and recognition tasks.

The stimuli used for both encoding and recognition were black line drawings of simple objects displayed on a white background taken from the Snodgrass 260-picture collection (Snodgrass & Vanderwart, 1980). Figures were chosen based on a name agreement of >95%, and the final set was chosen based on image agreement, familiarity and visual complexity measures.

Overt verbal responses during encoding were recorded with a Fibre Optic MRI-II noise-canceling microphone (http://www.optoacoustics.com/). Recognition responses were recorded using an fMRI compatible ‘tablet’ device with a touch screen and a pen, see Fig. 2.1A, similar to the device used in the BSCC. This device rested over the subject’s stomach with the pen placed in the subject’s right hand (Tam et al., 2011), see Fig. 2.1B. Subjects could monitor their performance in real time on the projection screen visible within the scanner. The tablet is useful for adapting common cognitive tests requiring pen and paper from outside the scanner with minimal change (Tam et al., 2011).

Subjects were trained prior to the fMRI session to acclimatize them to the enclosed scanner environment and teach them how to perform each task. Each subject underwent a thirty-minute simulator session, which included tablet training, inside and outside of the simulator. As well, detailed instructions were presented to the subject. This allowed the subject to read the instructions at their own pace and ask questions. A shortened version of each task was administered during the simulator session. When needed, subjects were given the opportunity to practice the tasks a second time in the simulator. This was important to reduce subjects’ anxiety regarding fMRI scanning prior to the scanning session.
During encoding, the subject was shown a picture that they were required to name aloud (overtly). This task had a block design with task and baseline durations of 10 seconds because it was found to reduce motion from overt speech artifacts with optimal BOLD signal detection (Birn, Cox, & Bandettini, 2004), Fig. 2.2(a). Each image was presented visually on a screen every 2.5 seconds and the subject overtly named each object.

Baseline blocks were administered in two different ways: 10 subjects passively viewed the letter X or O that randomly appeared in the center of the screen every two seconds, and 17 subjects
actively named these letters during the baseline component. This was done because of conflicting literature on tasks with overt speech: while Birn et al. (2004) recommend having the baseline component without speech, Barch et al. (1999) found that overt verbal responses were feasible for fMRI studies if the two conditions being compared both used overt speech (Barch et al., 1999; Birn et al., 2004). There were 4 events or images in each block of 10 seconds, with a total of 8 task and 8 baseline blocks per encoding run. A set of instructions was shown at the beginning of the task for 24 seconds, with the whole task duration (i.e., one task run) totaling 184 seconds. This task was adapted from the covert-encoding task used by Grady et al. (2006).

The first encoding run was followed by a 3 minute Trails A-B task and then the first recognition run. The recognition task had 4 task and 4 baseline blocks. Each block lasted 24 seconds and was comprised of 8 trials lasting 3 seconds each. Task blocks consisted of a series of forced-choice recognition trials where the subject was shown three Snodgrass line drawings: one was seen during encoding, one was a semantic foil and one was a perceptual foil. The subject used the tablet’s pen to select the picture they had previously seen (no speech required). The baseline component was set up the same as the task component, with different images: two scrambled images and a fixation cross. The subject had to select the fixation cross on the screen using the tablet’s pen. The location of the target image and fixation cross, i.e. the top, middle or bottom of the screen, varied across trials. This is an adaptation of the recognition task used by Grady et al. (2006), Fig. 2.2(b). A set of instructions was shown at the beginning for 24 seconds, with the whole run taking 216 seconds to complete.
A Encoding

TASK

Say 'cat' out loud

BASELINE

x OR o

Paradigm 1

Say 'x' or 'o' out loud

x OR o

Paradigm 2

Do NOT say 'x' or 'o' out loud

4 task images (10 seconds)

4 baseline images (10 seconds)

INSTRUCTIONS (24 sec)

... x8 (160 sec)
2.2 Subject Selection

Subjects were recruited from the Rotman Research Institute subject pools. They were initially contacted by telephone and a health questionnaire was administered to determine if they were eligible for the study. Subjects were required to be right-handed and between the ages of 18-34. Subjects provided their medical history via self-report. Exclusion criteria included history of
psychiatric disease, significant head injury, alcohol or drug abuse, diabetes, stroke, use of medications that could compromise brain function or hemodynamics, and cardiovascular disease.

2.3 Data Acquisition

If subjects were eligible to participate, an fMRI scanning session was scheduled at the Rotman Research Institute. Prior to the scanning session, subjects signed a consent form approved by the Baycrest Research Ethics Board, completed a detailed medical history form, and also completed paper tests, including the Mini-Mental Status Examination (MMSE) (Folstein, Folstein, & McHugh, 1975) and the Edinburgh Handedness Inventory (Oldfield, 1971).

Twenty-seven young, healthy volunteers (eleven male), aged 20-33 (24.8±3), participated in the study. All subjects were confirmed to be right handed using the Edinburgh Handedness Inventory. Subjects were screened for cognitive and neurological deficits using the MMSE, which had a group mean of 29.7± 0.6, out of 30; with the scores ranging from 28-30. Prior to the simulator session, visual acuity was assessed via Snellen charts, and subjects wore fMRI-compatible prescription glasses to correct for visual acuity when needed (SafeVision LLC. -6 to +6 diopters available in 0.5 increments).

Magnetic resonance imaging was conducted at a 3.0 Tesla magnetic field strength on a research-dedicated system (Magnetom Trio, Siemens, Canada) using the standard 12-channel “matrix” head phased-array coil.

Structural imaging was conducted using three-dimensional T1-weighted imaging (oblique-axial 3D MPRAGE, 2.63/2,000/1,100 ms TE/TR/TI, 9˚ FA, 256x192 matrix, 160 slices per volume,
voxel dimensions 1x1x1 mm$^3$). Functional MRI was conducted using BOLD fMRI (2D gradient-echo EPI, 30/2,000 ms TE/TR, 70° FA, 64x64 matrix, 30 slices per volume, voxel dimensions 3.125x3.125x5 mm$^3$). Presentation software (Neurobehavioral Systems, Inc.) was used for stimulus delivery and behavioral response recording. Auditory instructions were administered via the Avotec Silent-Scan Audio System (SS-3100, Avotec, Florida).

Subjects were compensated for their participation in the study after the fMRI session was completed. They were excluded from subsequent analysis if they had abnormal behavioral results and extreme TCM; the criteria for each are presented in the results section, along with the final dataset.

## 2.4 Preprocessing

The fMRI data were extensively preprocessed after fMRI acquisition and reconstruction in order to examine the results of the experiment. The following section describes preprocessing for the first run of recognition, in order to gain insight into the impact of individually preprocessing subject’s fMRI data. This was chosen as the experiment to be completed for this thesis because the recognition task is less complex to process than encoding (i.e. no corrections for speech were necessary).

For all subjects, the data were preprocessed with AFNI and FSL, open-source fMRI processing software ([Cox, 1996], [http://www.fmrib.ox.ac.uk/fsl/](http://www.fmrib.ox.ac.uk/fsl/)). 108 possible pipelines were tested for each subject, and were implemented and evaluated in order to determine the impact of certain steps alone and in conjunction with others, using the techniques described by Churchill et al. (2011). Pipelines were evaluated with and without the following preprocessing techniques: PNC, STC, MC and MPR. All pipelines always had at least one of the following steps: PNC,
STC or MC. The following steps were fixed across all pipelines: spatial smoothing (SM) and temporal detrending (DET). However, the polynomial order of detrending was not fixed and orders 0 (i.e., remove the mean) to a 5\textsuperscript{th} order polynomial were applied to each possible pipeline. STC, PNC, MC, spatial smoothing, detrending and MPR were all applied using AFNI, see Fig. 2.3.

Since the ordering of preprocessing steps impacts the final results, pipelines with both PNC and STC, which were the first steps applied if included in a pipeline, were applied to the data with PNC preceding STC, and STC preceding PNC (Churchill et al., 2010). MC followed STC and PNC (if they were applied) using the standard method, where all volumes within a subject’s run were registered to the 20\textsuperscript{th} scan/volume. MPEs are saved for MPR during this preprocessing step. After these three steps, there are 9 different pipelines, see Fig. 2.3.

Following MC, spatial smoothing was applied to all datasets with a 6.0 mm in-plane Gaussian kernel. Temporal detrending with Legendre polynomials of orders 0-5 was also applied to all datasets in order to test the effect of detrending order. Since the 9 pipelines each underwent detrending with six different Legendre polynomial orders, the number of pipelines after this step increases to 54. Finally, pipelines were preprocessed with and without MPR, which created the 108 different pipelines, see Fig. 2.3. The MPE time courses underwent principal component analysis (PCA). The two components with the largest variance were used as the MPR regressors in order to avoid over-fitting the MPEs to the fMRI time courses (Woods, Grafton, Holmes, Cherry, & Mazziotta, 1998). The two components accounted for 94.4±4.8\% variance across subjects.
Figure 2.3 108 Preprocessing pipelines applied on data

The figure illustrates the 108 pipelines applied to the data and the order each step was applied. Step A shows the first three pipelines, each with slice timing correction (STC), physiological noise correction (PNC) or motion correction (MC). Step B shows how these three steps were combined into 6 different pipelines with two or three different combinations of these steps. 9 pipelines come from steps A and B, represented by numbers 1-9 on the boxes. Pipelines from Steps A and B were spatially smoothed (SM) in step C. These 9 pipelines from Step C underwent detrending (DTR), each with orders 0 to 5, or detrending AND motion parameter regression (MPR), each with orders 0 to 5 in Step D. This resulted in 108 pipelines: 9 (with fixed smoothing, C) x 2 (DTR or DTR + MPR) x 6 (orders of detrending).

Three-dimensional masks were then created for each subject from the functional data with Brain Extraction Tool (BET), part of the FSL open-source software package (Smith et al., 2004). The brainstem was manually edited out of the functional masks using MRIcon
Each subject’s structural T1 image was used as an anatomical reference point during this process.

2.5 Evaluating Behavior and Outliers

It is important to assess subject behavior during the task by examining the two behavioral measures, reaction time and accuracy. In addition, abnormalities can be evaluated in certain parameters from the fMRI session as they may reflect that the subject was not properly performing the task, or moving too much during the duration of the task. This could be seen in task behavioral measures or in motion correlated with the task paradigm.

2.5.1 Reaction Time and Accuracy

The reaction time (RT) and accuracies (ACC) were examined by trial per subject for each condition, i.e. baseline and task RT and ACC were examined separately.

Occasionally, subjects had missing or fast RTs, which likely meant the response pen was pressed too quickly to record a true RT (<100ms). In this case the trial was recorded as incorrect, decreasing the subject’s overall ACC. These trials were defined as ‘false hits’ and were dropped from the behavioral analysis. RT and ACC were calculated as averages per condition, excluding these ‘false hit trials’, in milliseconds and fractions, respectively. The number of false hits was calculated as well as the number of incorrect recognition scores, per subject per task condition. Subjects with too many incorrect fixation trials and/or false hits in fixation and recognition were determined to be outliers, with the threshold conservatively set to greater than 2 standard deviations from the mean across subjects (equivalent to 5 or more trials).
2.5.2 Task-Correlated Motion (TCM)

TCM was calculated by correlating the first principal component of the motion parameter estimates with the task paradigm convolved with the HRF. The HRF was estimated using AFNI's standard gamma-variate basis function, SPMG1. A threshold for removing subjects with excessive TCM was created by taking the mean of all TCM measures across subjects, and removing subjects with TCM greater than 3 standard deviations from the mean. The TCM results are reported in Results section 3.2.

2.6 Pipeline Optimization Procedure

The preprocessed images from the 108 pipelines were then put into single-subject, single-pipeline analyses in order to produce metrics (described below) that summarize the effectiveness of each pipeline. The SPMs produced in this step were used for further analyses, described in Methods section 2.8.

For each individual subject, a two-class, multivariate canonical variates analysis (2cCVA) was performed on each of the 108 pipelines per subject using the NPAIRS (Nonparametric Prediction, Activation, Influence, and Reproducibility reSampling) framework (S. C. Strother et al., 2002). One transition scan was dropped at the beginning of each block, due to the delay in the hemodynamic response. The two ‘classes’ were defined as ‘task’ and ‘baseline’. NPAIRS uses a method called split-half resampling, where the scans from one half of a run are compared to its other half. The splits are independent of each other; one is the ‘training’ set and the other is the ‘test’ set. The procedure described above is applied to each split to generate two spatially z-scored SPMs.
The analysis started with an initial PCA on the entire task epoch per subject to denoise and reduce the dimensionality of the data (S. C. Strother et al., 2002). Only the first 30% of the PC components were kept after ordering them from largest to smallest variance. After that, the reduced data was split into two groups: the first split contained the first 44-scan epoch in the task (the first two recognition and two fixation blocks with one transition scan discarded from each block), and the second split contained the second 44-scan epoch in the task (the last two recognition and two fixation blocks with one transition scan discarded from each block). A second PCA was performed independently on each split group.

The 2c-CVA was computed on the second PCA of each split-half group. The 2cCVA produced a spatially z-scored SPM for each of the two splits. The two independent z-scored SPMs were combined into a single reproducible z-scored SPM with the largest difference in activity between the task and baseline classes by plotting each pair of z-scores for each voxel on a scatter plot, projecting these points onto the signal (major) axis, and normalizing by the standard deviation of the noise (minor) axis (S. C. Strother et al., 2002).

NPAIRS calculates two quantitative statistics that reflect the consistency of the data across splits without having to set an SPM detection threshold, see Fig. 2.4. The first is reproducibility, which is a measure of similarity between two SPMs. Reproducibility (R) in NPAIRS is obtained by measuring the correlation between all pairs of unthresholded brain voxel values between the z-scored SPMs of the two split groups. Prediction (P) is calculated as the median rate of using the first split’s 2cCVA model to correctly predict the class of each scan in the second split with posterior Bayesian probability, and vice versa. P ranges from 0.5 to 1 and R ranges from 0 to 1, with perfect P and R both at 1.
This analysis was performed on each individual subject and individual pipeline, for 2-9 principal components (PCs) for each of the 2nd split-half PCAs. The PC number can significantly affect model (R,P). The optimal PC number for each analysis was chosen based on which set produced the maximal (R,P) values. This was measured via Euclidean distance (D), which uses the R and P values per PC number to quantify how close performance was to having the perfect model (R=1, P=1), with a smaller D being closer to perfect (R,P). The PC number for each analysis was thus chosen based on minimizing D.

2.6.1 Determining the Optimal Fixed Pipeline

An optimal fixed pipeline is the “best” fixed pipeline on average across all subjects. Following individual-subject NPAIRS analysis, the R and P values were obtained and used for determining the optimal fixed pipeline. The optimal fixed pipeline was chosen using the procedure outlined by Churchill et al. (2011), see Fig. 2.5:

1. Each subject’s pipeline performance was assessed via D, quantifying how close performance was to having the perfect model (R=1, P=1), see Fig. 2.5A.

2. Pipelines per subject were ranked by arranging the D’s in a matrix and assigning the smallest D (closest to perfect) as the highest rank, which formed a matrix of ranks ranging from 108 to 1 per subject, see Fig. 2.5B.

3. The median of each subject’s ranking matrix was calculated for each pipeline across subjects. This was the ‘median rank metric’, see Fig. 2.5C.

4. Significance of the optimal fixed pipelines were tested using a Friedman test.

This procedure identified the optimal fixed pipeline if it existed, and those pipelines that could not be statistically distinguished from it, across all subjects (Churchill et al., 2011).
2.6.2 Determining the Optimal Individual Pipeline

Pipelines that are optimal for each individual subject usually have increased R and/or P relative to most of the other 108 pipelines, and the preprocessing steps included generally differ from the fixed pipeline (Churchill et al., 2011). The optimal individual pipeline was chosen by calculating D (mentioned above) per pipeline per subject and selecting the pipeline per subject that produced the lowest D-value.
Figure 2.5 Median rank metric

The procedure for identifying the optimal fixed pipeline across subjects in a 5 pipeline example. Euclidean distance, D, is obtained per pipeline to measure the pipeline closest to perfect R and P (A), ranked from least optimal (lower score) to most optimal (higher score) (B) for all subjects, and then median rank is obtained for each pipeline across all subjects, producing a median-rank profile (C). Adapted from Churchill et al. (2011) (Churchill et al., 2011). Copyright © 2011, John Wiley and Sons. The adaptation is by permission of the copyright holder.

2.7 Creating a Standard Group Template

The pipeline optimization procedure was performed on images in “subject space”, which means when comparing the SPMs (from single-subject analysis) between subjects, the voxels from one subject will not necessarily match up to the same anatomical location in another subject. All subjects’ SPMs were transformed into a “common space” by creating a group specific template and then by transforming all SPMs to this template.

The group template was created through an iterative process using FSL (Guimond, Meunier, & Thirion, 2000):

1. All subjects’ T1 images were transformed to the MNI152 template.
2. The masked, transformed T1 volumes were averaged.

3. The original T1 images were transformed to this first group-average template.

4. The masked, transformed T1 volumes were averaged, creating a second-group average template.

5. The original T1 images were transformed to this second group-average template.

6. The masked, transformed T1 volumes were averaged, creating a third (and final) study-specific template with high-quality inter-subject alignment.

Subsequently, SPMs from the NPAIRS analysis were transformed to this group template, once again using FSL.

In addition, a group mask was created for analysis (section 2.8) from each subject’s preprocessed images. This was done by:

1. Taking the mean of the functional volume to produce a three-dimensional image per subject, using 3dTstat in AFNI.

2. Multiplying the mean functional image by the functional mask used in the NPAIRS analysis in order to produce a segmented mean functional image. This was done using fslmaths in FSL.

3. Combining matrices from the registration of: each subject’s segmented mean functional image to their structural T1, and each subject’s structural T1 to the final subject template (using flirt in FSL).

4. Applying the combined matrix to the final registration of each subject’s segmented mean functional image to the final subject template (using flirt in FSL).

5. Averaging the aligned, segmented mean functional images across subjects using fslmaths.
6. Converting the average image into the final binary group mask using BET in FSL.

2.8 Analysis of Optimal SPM Results

Optimal z-scored SPMs, which were transformed to the group template, for fixed and individually optimized pipelines were saved for further analysis. Fixed and individually optimized pipelines were analyzed separately and then the results were compared to each other. Two-tailed, paired Wilcoxon tests, $\alpha=0.05$, were used to compare the R and P values for fixed and individually optimized pipelines.

2.8.1 Singular Value Decomposition (SVD)

Singular value decomposition (SVD) was performed on the subject’s SPMs to identify common patterns of activation across subjects’ SPMs. SVD, similar to PCA (without mean SPM pattern subtraction), was performed on the transformed, z-scored SPMs across subjects for fixed and individually optimized SPMs. SVD identifies activation across subjects and separates these patterns into components. These components are ranked by the amount of variance they account for across subjects, with the first component accounting for the most variance and each component after that accounting for successively decreasing variance.

When the SVD is performed on the data, it produces three separate pieces of information (per component): singular values (which can be converted to percent variance per component), spatial eigenimages (similar to SPMs) and subject weights or brain scores per component. The subject weights per component represent how each subject relates to the spatial eigenimage for that component. For example, subjects with a strong positive weight in the first component
strongly express the positive pattern of activation in the first component’s eigenimage and
negatively express the negative pattern in the first component’s eigenimage.

This analysis was carried out in Matlab 7.4. The group mask created from the group template
was applied to all SPMs prior to the SVD to ensure voxels being compared across subjects were
consistent.

When interpreting SVD results, it is important to identify which components are significant to
minimize false positive results. Significant components will contain meaningful activation
patterns and insignificant components will contain random/null activation. Unfortunately, it is
difficult to assess which components are meaningful and which contain mostly noise.
Insignificant components closer to the meaningful components may contain random activations,
even though the majority of the component is noise. In order to determine which components in
the SVD were significant and how to threshold the data, an SVD was performed on a simulated
dataset comprised of ‘noise’. This was done to model what an SVD would look like if the data
comprised pure noise in order to distinguish significant components from insignificant
components in the SVD on the fMRI data. Insignificant components in the real data should
have similar distributions when compared to components in the simulated data SVD. A
description of the simulation is in Appendix A.

In order to determine which components were significant in the real data, the intensity values
were examined in the fixed and individually optimized SVD eigenimages, across all
components. This was done by plotting the range of intensity values for each component, fitting
a t-location scale distribution (in Matlab,

http://www.mathworks.com/help/toolbox/stats/brn2ivz-145.html) to each component, and
comparing the parameters of each curve across components. A detailed description of this process is described in Appendix B. Using the method described in Appendix B, reasonable evidence was found to conclude that components 1 and 2 contained significant patterns for the fixed pipeline optimization, and components 1, 2, and 3 contained significant patterns for the individually optimized pipeline.

In order to create a meaningful statistical threshold for each component, a null distribution was created from the null components in the SVD results. Null component selection is described in Appendix B. The three curve parameters from the fitted t location-scale distribution (mu, sigma, and nu) were then averaged across these designated null components, creating a null distribution. The inverse t-location-scale distribution probability was then taken at 2.5% and 97.5%, the lower and upper bounds, respectively, for a 95% threshold. This was done for both sets of pipelines, creating 95% thresholded SVD eigenimages for meaningful components.

After the eigenimages were thresholded, minimum cluster size was determined by running Monte Carlo simulations (3dClustSim, AFNI), resulting in a minimum 10-voxel cluster. There was a five-percent chance a cluster survived due to chance. Peak clusters greater than the set minimum, and above the calculated threshold, were voxels connected via nearest-neighbor (cluster, FSL). Peak voxel coordinates were transformed into MNI space and the AAL atlas was used to identify active regions (Tzourio-Mazoyer et al., 2002).

Following thresholding, each subject’s weight within each component from the SVD was regressed against the various behavioral measures in order to evaluate if the network activation in the SVD eigenimages were related to behavior. These measures were fixation accuracy (fixACC), recognition accuracy (recACC), fixation RT (fixRT), recognition RT (recRT), and the
difference between recognition and fixation RT (RTdif). Correlation and goodness of fit, $R^2$, were used to assess the strength of the relationship between the two variables. Cook’s Distance, a method used in regression to check how influential each point is, was used to determine if any subjects were outliers in the regressions (Cook & Weisberg, 1982).

Akaike’s Information Criterion (AIC) was used to determine whether fixed or individually optimized subject weights resulted in a better regression model with behavior. AIC is a model selection technique often used as a measure of goodness of fit for the statistical model of interest. An optimal model will minimize AIC (Akaike, 1969).

2.9 Preliminary study

This preliminary study only examined motion correction, in both encoding and recognition runs; the fMRI images were not analyzed or put into preprocessing pipelines. These results are included to demonstrate that significant decreases in motion can be achieved in overt speech and motor tasks by adjusting a commonly implemented preprocessing technique. The standard motion correction technique (all scans aligned to the same target scan, #20 within its own run) was adjusted in order to reduce the impact of correcting within-run head motion. This was done by aligning all scans within a run to the scan with minimum, mean displacement. Impact was measured by taking the median of the post-motion correction maximum displacement values across time, i.e. the median maximum displacement. The median was used, instead of the mean, because it was less influenced by outliers within the time series, which are not reflective of the total amount of motion occurring during a run, see Table 2.1. This was tested by comparing the variances of median maximum displacement across subjects versus the variance of maximum displacement across subjects (i.e. the variances displayed in Table 2.1). The variances of
median maximum displacement were significantly less than the variances of maximum displacement (p<0.01, Paired Wilcoxon).

**Table 2.1 Median and maximum displacement variance**

Variance of maximum and median maximum displacement for standard and altered motion correction across subjects for both runs of encoding and recognition

<table>
<thead>
<tr>
<th>Task</th>
<th>Standard MC</th>
<th></th>
<th>Altered MC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum Displacement</td>
<td>Median Displacement</td>
<td>Maximum Displacement</td>
<td>Median Displacement</td>
</tr>
<tr>
<td>Encoding 1</td>
<td>0.667</td>
<td>0.118</td>
<td>0.707</td>
<td>0.023</td>
</tr>
<tr>
<td>Encoding 2</td>
<td>0.722</td>
<td>0.074</td>
<td>0.427</td>
<td>0.046</td>
</tr>
<tr>
<td>Recognition 1</td>
<td>0.268</td>
<td>0.088</td>
<td>0.235</td>
<td>0.037</td>
</tr>
<tr>
<td>Recognition 2</td>
<td>1.350</td>
<td>0.119</td>
<td>0.771</td>
<td>0.085</td>
</tr>
</tbody>
</table>

One-tailed paired Wilcoxon tests were used to compare the median maximum displacement values from the altered motion correction technique to the standard one (α =0.05). The results are presented and discussed in the future directions, section 4.2.
3 Results

3.1 Behavioral Results

RT and ACC were examined across subjects for fixation and recognition task blocks. Mean fixRT was 1114.3±267.3 seconds and mean recRT was 1593.2±270.5 seconds, see Fig. 3.1A. Mean fixACC was 0.958±0.089, and mean recACC 0.934±0.099, see Fig. 3.1B. Fig. 3.1 displays some clear outliers in fixation and recognition components. These outliers had to be further assessed prior to completing the rest of the analysis.

3.1.1 Behavioral Outlier Analysis

Trials were broken down into two parameters subdivided by fixation and recognition components: the number of trials classified as incorrect and as false hits, as defined in section 2.5.1, see Fig. 3.2 A and B respectively. Subjects with too many incorrect trials and/or false hits were determined to be outliers; the criteria are described in section 2.5.1. Outliers are the colored points in Fig. 3.2. In total, there were four subjects removed from analysis due to too many incorrect trials or false hits.
Figure 3.1 Reaction time and accuracy plots for all subjects

Reaction time (A) and accuracy (B) plots for all subjects subdivided by fixation and recognition blocks.
Figure 3.2 Incorrect trials and false hits

Plots for the number of incorrect trials (A) and false hits (B) for fixation and recognition task components. Colored points are outliers 1-4. If points are the same color on plot A and B, they are the same subject.
3.1.2 Behavioral Results with Outliers Removed

Behavioral results were re-examined with the four outliers removed, see Fig 3.3. Mean fixRT was 1102.5±256.8 seconds and mean recRT was 1605.2±194.5 seconds, see Fig. 3.3A. RecRT is significantly longer than fixRT (p<0.05, two-tailed paired T-Test), subjects are consistently slower as expected in recognition versus fixation trials. In addition, subjects that had an increased speed during fixation progressively slowed down during recognition trials relative to subjects with slower fixRT, as displayed in Fig. 3.3A (R^2 = 0.79). Mean fixACC was 0.984±0.024, and mean recACC 0.964±0.028, see Fig. 3.3B. FixACC scores are significantly higher than RecACC scores (p<0.05, two-tailed paired T-Test). However, recognition scores are high, indicating that subjects did not find the recognition task too difficult.

3.2 Subject Motion

Maximum displacement and task-correlated motion were evaluated, see Fig. 3.4. Mean maximum displacement across all subjects (without behavioral outliers) was 0.970±0.517mm. Only one subject had motion greater than 2mm, which was the same subject who had significant task-correlated motion, greater that 0.3. This subject had to be dropped from the dataset as there was too much motion during the fMRI session and there was reasonable evidence to conclude that the motion would influence the fMRI results. Aside from this one subject, motion appears to have minimal impact on the dataset and is generally less than half a voxel width in maximum displacement.
Figure 3.3 Reaction time and accuracy plots without outliers

Reaction time (A) and accuracy (B) plots for all subjects without behavioral outliers, subdivided by recognition and fixation blocks. The line of identity is displayed on 3.3(A) to demonstrate the trend in the recognition data.
Motion assessment plot displaying the maximum voxel displacement (mm) measured with AFNI motion correction for recognition.

### 3.3 Physiological Parameters

Subject’s physiological measures were assessed for potential physiological outliers, see Fig. 3.5. There are no outliers, as the physiological measures per subject were all less than 3 standard deviations from the mean. Within-subject standard deviation is much larger for cardiac rate compared to respiration rate. Mean cardiac rate was 1.231±0.218 Hz and mean respiration rate was 0.379±0.047 Hz. Since there were no physiological outliers, the final dataset included 20
subjects; 4 subjects were behavioral outliers, 2 subjects had missing physiological measures and 1 subject had TCM.

**Figure 3.5 Physiological data**

Mean and standard deviation of cardiac and respiration rates per subject (in Hz).

### 3.4 Fixed Pipeline Optimization

Using the median rank metric, it was determined that motion correction with third order detrending (DET3) is the optimal fixed pipeline for all subjects ($p < 0.001$, Friedman test), see Fig. 3.6. Median rank represents the degree of pipeline optimization, with a higher median rank indicating a pipeline is more optimal across all subjects compared to other pipelines. There are
five significant pipelines, three of which have MC with smoothing and various orders of detrending (DET0, DET3, DET5), as well as two with STC + MC + DET3 and PNC+MC+DET3
Figure 3.6 Fixed pipeline optimization

Comparison of fixed pipeline optimization in recognition data (task versus baseline) using the median rank metric based on NPAIRS reproducibility and prediction, represented by the black line. Combinations of the following steps were tested with 108 pipelines in total: motion correction (M), slice timing correction (T), physiological noise correction (P), spatial smoothing, detrending (DET), and motion parameter regression (MPR). M, P and T are applied in the same order as the pipeline label, with smoothing and MPR last. The highest median rank indicates the most optimal pipeline. The dotted lines represent the upper and lower distribution quartiles. Pipelines that fall between the two red lines are not significantly different from the optimal pipeline (circled) according to the Nemenyi critical difference test (p = 0.05).

(Fig. 3.6). These pipelines lie within the Nemenyi critical difference boundary (p=0.05), and thus are not significantly different from the optimal pipeline. R and P for the optimal pipeline, MC+smoothing+DET3, are 0.655±0.139 and 0.865±0.081, respectively.
103/108 pipelines fall below the critical difference boundary, i.e. they are significantly worse than the optimal set of pipelines. Switching the order of PNC and STC has no significant effect on the group pipeline optimization as all the pipelines with both steps included underperform. The addition of MPR lowers the median rank of most pipelines.

There are two significant components in the fixed pipeline SVD, determined by the component selection procedure described in Appendix B. The subject weights within each component are plotted in Fig. 3.7A. Two subjects appear to have outlying brain scores in component 2, see Fig. 3.7A. These were not further investigated since fixed preprocessing pipelines do not necessarily result in optimal SPMs in each individual subject, i.e. suboptimal preprocessing in some subjects could mean the inclusion of artifacts.

Thresholded eigenimages from component 1 are displayed in Fig. 3.7C, and significant clusters are listed in Table 3.1. Not all activated regions are listed in the table since regions with peak voxel activations are reported and some clusters span several regions. Regions in component 1 represent common patterns of activation and negative activation across all subjects, since all subject brain scores in component 1 have positive magnitude. These activations are positively related to all subjects since subjects in component 1 have positive subject weights. The first, and largest component 1 positive loading cluster, in Table 3.1, has its peak voxel activation in the R fusiform gyrus. However, this cluster spans several regions on the L and R sides including: L fusiform gyrus, L and R cerebellum, L and R inferior, middle and superior occipital gyri, extending into the L and R inferior temporal gyrus. The second largest positive cluster in component 1 is in the L superior parietal lobule, and extends into the L inferior parietal lobule. The next largest positive cluster in component 1 peaks in the L inferior frontal
gyrus (operculum) and extends into the L precentral gyrus. There are two more clusters in the R angular gyrus and R precuneus.
Figure 3.7 Fixed pipeline optimization SVD results

SVD fixed pipeline analysis results. A scatter plot of subject weights for significant components 1 and 2 are displayed in (A). Percent variance for each component is in brackets next to the axis labels. The significant regression of subject weight with behavior is plotted against the reaction time difference between fixation and recognition for component 2 (B). Points are colored in plots (A) and (B) to track two subgroups in the individual optimization results, described in subsequent sections; see Fig. 3.10. Correlation (r), the significance of correlation (p) and goodness of fit ($R^2$) is displayed on the plot. Axial slices for component 1 (C) and component 2 (D) are displayed in MNI coordinate space.
Negative loadings are negatively associated to all subjects since they all have positive weights in component 1, see Fig. 3.7C. The largest cluster in the negative loading for component 1 has peak activation in the L and R medial orbito-frontal cortex, extending into the L and R superior frontal gyrus. There are two other significant clusters in the component 1 negative loading: R inferior parietal lobule extending into the R supramarginal gyrus and R superior parietal lobule, and the L and R middle cingulate cortex.

Thresholded eigenimages from component 2 are displayed in Fig. 3.7D and peak voxels from significant clusters are listed in Table 3.1. These patterns are much more variable, appearing to represent individual variability in activation, and are strongly associated with two subjects, see Fig. 3.7A. The largest cluster has peak activation in the R precuneus and extends into the L precuneus as well as the L and R superior parietal lobule. Other activated regions in this loading include: R lingual gyrus, L and R cerebellum, L supramarginal gyrus, L and R medial orbito-frontal cortex, L superior frontal gyrus, L and R middle frontal gyrus, L and R inferior frontal gyrus (orbital part), L inferior frontal gyrus (triangular and opercular parts), L and R middle cingulate cortex extending to the posterior cingulate cortex, L superior temporal gyrus, L amygdala, R parahippocampal gyrus, R middle temporal gyrus, and R inferior temporal gyrus. There is also activation in the L and R hippocampus, although it is not a significant cluster. However, there appear to be noise artifacts in the brainstem, at the edges of the brain, and in the ventricles that appear as significant activations in the component 2 eigenimages.

There was only one significant cluster in the negative loading, in the R fusiform gyrus, which is positively associated to subjects with negative subject weights.
The regression of subject brain scores with each of the behavioral measures resulted in only one relationship of interest ($R^2>0.15$): component 2 with RTdif ($R^2=0.196$), see Fig. 3.7B. The correlation between the two variables resulted in a p-value that was just significant at the 0.05 level ($r=0.443, p = 0.05$). The relationship appeared to be heavily influenced by the two potential outliers mentioned above. However, neither point was determined to be an outlier according to Cook’s distance, i.e. the deletion of either point does not significantly alter the regression relationship. Subjects with a lower RTdif tend to be positively associated with the positive loading from component 2 and subjects with a higher RTdif tend to be negatively associated with the positive loading from component 2. The negative loading in component 2 tends to be positively associated with subjects with higher RTdif and negatively associated with subjects with lower RTdif. However, the two potential outlier subjects, who also have low RTdif, appear to drive component 2 with the strongest positive subject weights.
Table 3.1 Fixed pipeline optimization SVD peak voxel activation

Regions with peak voxel activation in nearest neighbor clusters from the SVD analysis on fixed pipeline optimization. Regions are ordered by cluster size (significant cluster size is greater than 10 voxels). Coordinates are given in MNI space.

<table>
<thead>
<tr>
<th>Anatomical Location</th>
<th>Coordinates (MNI)</th>
<th>Cluster Size (voxels)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Component 1 (45.07%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive loadings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Fusiform gyrus</td>
<td>38 -73 -17</td>
<td>1835</td>
</tr>
<tr>
<td>L Superior parietal lobule</td>
<td>-28 -63 58</td>
<td>80</td>
</tr>
<tr>
<td>L Inferior frontal gyrus (opercular part)</td>
<td>-40 9 28</td>
<td>54</td>
</tr>
<tr>
<td>R Angular gyrus</td>
<td>26 -63 48</td>
<td>17</td>
</tr>
<tr>
<td>R Precuneus</td>
<td>7 -76 53</td>
<td>13</td>
</tr>
<tr>
<td>Negative loadings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR Medial orbito-frontal cortex</td>
<td>-3 56 -7</td>
<td>129</td>
</tr>
<tr>
<td>R Inferior parietal lobule</td>
<td>57 -41 48</td>
<td>66</td>
</tr>
<tr>
<td>LR Middle cingulate cortex</td>
<td>1 -32 48</td>
<td>14</td>
</tr>
<tr>
<td><strong>Component 2 (6.82%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive loadings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Precuneus</td>
<td>7 -60 68</td>
<td>476</td>
</tr>
<tr>
<td>R Cerebellum</td>
<td>41 -79 -22</td>
<td>410</td>
</tr>
<tr>
<td>R Lingual gyrus</td>
<td>10 -54 8</td>
<td>177</td>
</tr>
<tr>
<td>L Cerebellum</td>
<td>-46 -76 -22</td>
<td>145</td>
</tr>
<tr>
<td>LR Medial orbito-frontal cortex</td>
<td>-3 56 -12</td>
<td>63</td>
</tr>
<tr>
<td>L Supramarginal gyrus</td>
<td>-53 -35 58</td>
<td>62</td>
</tr>
<tr>
<td>R Parahippocampal gyrus</td>
<td>16 -4 -17</td>
<td>34</td>
</tr>
<tr>
<td>L Middle frontal gyrus</td>
<td>-37 34 43</td>
<td>28</td>
</tr>
<tr>
<td>R Middle temporal gyrus</td>
<td>29 15 -32</td>
<td>23</td>
</tr>
<tr>
<td>LR Middle cingulate cortex</td>
<td>-6 -4 43</td>
<td>22</td>
</tr>
<tr>
<td>L Thalamus</td>
<td>-3 -16 18</td>
<td>21</td>
</tr>
<tr>
<td>L Inferior frontal gyrus (opercular part)</td>
<td>-53 15 33</td>
<td>20</td>
</tr>
<tr>
<td>L Inferior frontal gyrus (triangular part)</td>
<td>-37 34 18</td>
<td>18</td>
</tr>
<tr>
<td>L Superior temporal gyrus</td>
<td>-65 -16 13</td>
<td>15</td>
</tr>
<tr>
<td>L Middle frontal gyrus (orbital part)</td>
<td>-46 49 8</td>
<td>15</td>
</tr>
<tr>
<td>L Superior frontal gyrus</td>
<td>-24 62 13</td>
<td>14</td>
</tr>
<tr>
<td>L Amygdala</td>
<td>-18 -4 -17</td>
<td>14</td>
</tr>
<tr>
<td>R Inferior temporal gyrus</td>
<td>54 -23 -27</td>
<td>12</td>
</tr>
<tr>
<td>R Middle frontal gyrus</td>
<td>29 31 48</td>
<td>11</td>
</tr>
<tr>
<td>R Inferior frontal gyrus (orbital part)</td>
<td>44 43 -7</td>
<td>11</td>
</tr>
<tr>
<td>Negative loadings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Fusiform gyrus</td>
<td>32 -76 -12</td>
<td>20</td>
</tr>
</tbody>
</table>
3.5 Individual Pipeline Optimization

The fraction of subjects, out of 20, with each preprocessing step included in its individually optimized pipeline is summarized in Fig. 3.8. There are a wide variety of preprocessing steps included in the subject pipelines. 13/20 subjects have MC included in their pipelines, 13/20 subjects have STC, and 8/20 subjects have PNC. In addition, 7/20 subjects required MPR, and DET0 was the most commonly used detrending order, used in 7/20 subjects. 6/20 subjects have both STC and PNC in their pipelines and 5/6 are optimized with STC preceding PNC. Mean R and P for each subject’s optimal pipeline are 0.744±0.070 and 0.869±0.091, respectively. No subjects were individually optimized with the same pipeline as the fixed optimal pipeline (MC + smoothing + DET3).

The 7/20 subjects optimized with MPR did not have significantly higher maximum displacement compared to the other 13/20 subjects (p>0.05, two-tailed T-Test).

The initial SVD on the individually optimized data resulted in three components of interest. Two outliers were discovered in the subject weights for components 1 and 2. Examination of their NPAIRS SPMs revealed motion artifacts, see Fig. 3.9A and B, and both of their optimal pipelines did not contain motion correction. Motion correction was then added into both of their pipelines, which successfully removed the obvious visual artifacts, and the SVD on the individually optimized data was repeated with the refined pipelines in these two subjects.
The results of the refined SVD are presented in Fig. 3.10. The subject weights of the first two components are in Fig. 3.10A. One subject appears to have an outlying brain score in component 2, see Fig. 3.10A. As subject weights in component 1 increase, two subgroups appear to differentially grow stronger in component 2: one in the positive direction (white points) and one in the negative direction (blue points). The regression of subject brain scores with each of the behavioral measures resulted in only one relationship of interest ($R^2>0.15$): component 2 with RTdif ($R^2=0.30$), see Fig. 3.10B. The subgroups mentioned above are plotted in the same colors in Fig. 3.10B. The correlation between the two variables is significant ($r=0.548$, $p < 0.05$).

According to Cook’s distance, the one potential outlier in component 2, mentioned above, does
not significantly influence the relationship, i.e. removing this point does not significantly alter the regression relationship.

**Figure 3.9 Z-scored SPMs with motion artifacts**

Z-scored SPMs with motion artifacts from two individual subjects, A and B, identified as outliers in the individual pipeline optimization. Sagittal slice numbers are displayed in MNI coordinate space.

Thresholded eigenimages from component 1 are displayed in Fig. 3.10C, and significant clusters are listed in Table 3.2. Table 3.2 lists the peak voxel activations from significant clusters, not every active region. Regions in component 1 represent common patterns of activation and negative activation across all subjects, since all subject brain scores in component 1 have positive magnitude. The positive loadings are positively associated with all subject weights since they are positive. The first, and largest component 1 positive loading cluster, in Table 3.2,
has its peak voxel activation in the R fusiform gyrus. However, this cluster spans several regions on the R side including: R cerebellum, R inferior, middle and superior occipital gyri, extending into the R inferior temporal gyrus, as well as the R angular gyrus. The second largest positive cluster in component 1 is in the L inferior occipital gyrus, extending into the L middle and superior occipital gyri, L cerebellum, L fusiform gyrus, and L inferior temporal gyrus. The next positive cluster of activation peaks in the L superior parietal lobule and extends into the L inferior parietal lobule. The next positive cluster in component 1 peaks in the L precentral gyrus and extends into the L inferior frontal gyrus (opercular part). There are two more clusters in the L and R precuneus.

Negative loadings in component 1, are negatively associated to all subjects since they all have positive weights in component 1, see Fig. 3.10C. The largest cluster in the negative loading for component 1 has peak activation in the L and R medial orbito-frontal cortex, which extends into the superior frontal gyrus. The final cluster in the negative loading for component 1 is located in the R inferior parietal lobule.

Thresholded eigenimages from component 2 are displayed in Fig. 3.10D, and significant clusters are listed in Table 3.2. Component 2 eigenimages are much more variable. Positive loadings show activation in the R angular gyrus, which extends into the R middle and superior occipital gyri, as well as the R inferior parietal lobule, R superior parietal lobule, and R precuneus. Other activated regions in the positive loading include: L and R cerebellum, L Middle frontal gyrus, L thalamus, L supramarginal gyrus, L and R middle cingulate cortex extending to the posterior cingulate cortex, R Middle occipital gyrus, R inferior frontal gyrus (opercular part), R supplementary motor area, R parahippocampal gyrus, and R lingual gyrus. There is also activation in the L parahippocampal gyrus, even though it is not a significant cluster.
These positive loadings are associated with subjects that have lower RTdifs and negatively associated to subjects with increased RTdifs, see Fig. 3.10B. Subjects in the positive subgroup increasingly express this pattern as they increasingly express the positive component 1 pattern.

Regions activated in the negative loading of component 2 include: R middle occipital gyrus which extends into the R fusiform gyrus, R inferior occipital gyrus as well as the R inferior temporal gyrus. Other negative activations include: R rectus gyrus, R superior frontal gyrus (orbital part), and the L middle occipital gyrus. Subjects with increased RTdifs, i.e. negative subject weights, are positively associated with this negative loading and subjects with lower RTdifs are negatively associated with this negative loading in component 2, see Fig. 3.10B. However, there was noise in the brainstem and ventricles appearing as significant activation in the component 2 eigenimages. Subjects in this negative subgroup increasingly express this pattern as they increasingly express the positive component 1 pattern.

Component 3 accounted for 5.1% variance. However, it was not further considered because there were no relationships of interest between its subject weights and behavior measures, and the eigenimage appears confounded by noise.
A

Component 2 (5.77 %)

Component 1 (43.98 %)

B

\[ R^2 = 0.30 \]
\[ r = 0.55 \]
\[ p < 0.05 \]
Figure 3.10 Individual pipeline optimization SVD results

SVD on refined individually optimized analysis results. A scatter plot of subject weights for significant components 1 and 2 is displayed in (A). Percent variance for each component is in brackets next to the axis labels. The significant regression of subject weight with behavior is plotted against the reaction time difference between fixation and recognition for component 2 (B). Points in plots (A) and (B) are colored to display two subject weight subgroups in component 2 that increasingly grow positive (white) and negative (blue) as their component 1 subject weights increase. Correlation (r), the significance of correlation (p) and goodness of fit ($R^2$) is displayed on the plot. Axial and coronal activations for component 1 (C) and component 2 (D) are displayed with MNI slice coordinates.
Table 3.2 Individual pipeline optimization SVD peak voxel activation

Regions of activation with peak voxel activation in nearest neighbor clusters from the SVD analysis on individually optimized pipelines. Regions are ordered by cluster size (significant cluster size is greater than 10 voxels). Coordinates are given in MNI space.

<table>
<thead>
<tr>
<th>Component 1 (43.98%)</th>
<th>Anatomical Location</th>
<th>Coordinates (MNI)</th>
<th>Cluster Size (voxels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive loadings</td>
<td>R Fusiform gyrus</td>
<td>38 -73 -17</td>
<td>1006</td>
</tr>
<tr>
<td></td>
<td>L Inferior occipital gyrus</td>
<td>-46 -73 -17</td>
<td>799</td>
</tr>
<tr>
<td></td>
<td>L Superior parietal lobule</td>
<td>-28 -63 58</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>L Precentral gyrus</td>
<td>-43 9 33</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>L Precuneus</td>
<td>-6 -76 53</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>R Precuneus</td>
<td>7 -76 53</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>LR Medial orbito-frontal cortex</td>
<td>-3 59 -7</td>
<td>145</td>
</tr>
<tr>
<td>Negative loadings</td>
<td>R Inferior parietal lobule</td>
<td>48 -48 58</td>
<td>38</td>
</tr>
<tr>
<td>Component 2 (5.77%)</td>
<td>Positive loadings</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R Angular gyrus</td>
<td>35 -69 53</td>
<td>902</td>
</tr>
<tr>
<td></td>
<td>R Cerebellum</td>
<td>41 -79 -22</td>
<td>392</td>
</tr>
<tr>
<td></td>
<td>L Cerebellum</td>
<td>-34 -88 -22</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>R Inferior frontal gyrus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(opercular part)</td>
<td>48 18 38</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>L Middle frontal gyrus</td>
<td>-40 34 38</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>L Thalamus</td>
<td>-3 -16 18</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>R Lingual gyrus</td>
<td>13 -63 -7</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>R Supplementary motor area</td>
<td>1 -4 68</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>L Supramarginal gyrus</td>
<td>-65 -19 18</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>LR Middle cingulate cortex</td>
<td>-9 -4 43</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>R Parahippocampal gyrus</td>
<td>13 -32 -17</td>
<td>11</td>
</tr>
<tr>
<td>Negative loadings</td>
<td>R Middle occipital gyrus</td>
<td>38 -88 8</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>R Rectus gyrus</td>
<td>4 21 -17</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>R Superior frontal gyrus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(orbital part)</td>
<td>7 59 -22</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>L Middle occipital gyrus</td>
<td>-24 -98 13</td>
<td>13</td>
</tr>
</tbody>
</table>
3.6  Fixed versus Individually Optimized Pipelines

The final individually optimized dataset included the two subjects with MC added into their pipelines. Individually optimized pipelines have significantly higher R than fixed pipelines ($p<0.0001$, two-tailed paired Wilcoxon). There are no significant changes in P ($p>0.05$, two-tailed paired Wilcoxon). Differences in R and P as well as D are given in Table 3.3. Mean change in R is 0.088±0.093 and mean change in P, although insignificant, is 0.004±0.054. All subjects show an increase in R when pipelines are individually optimized compared to the optimal fixed pipelines. However, only 12/20 subjects have an increase in P when individually optimizing pipelines compared to fixed pipeline optimization. D represents the ‘distance’ R and P are from perfect (1,1), and thus a smaller D means the subject’s R and P are closer to perfect. Mean D for fixed pipelines is 0.374±0.151, and 0.296±0.089 for individually optimized pipelines. The mean difference between fixed and individually optimized pipelines is -0.078±0.089. Individually optimizing pipelines significantly decreases D compared to fixed subject pipelines. ($p<0.0001$, two-tailed paired Wilcoxon) and all 20 subjects have a decrease in D for individual pipeline optimization.

Individually optimized pipelines consistently result in significant increases in significant positive voxels, exceeding a z-score of 3 ($p<0.0001$, two-tailed paired Wilcoxon), see Fig. 3.11(A), and significant negative voxels, with a z-score less than -3 ($p<0.0001$, two-tailed paired Wilcoxon), see Fig. 3.11(B), than fixed pipelines. All points are on or above the line of identity, demonstrating that each subject has an increase in detected significant positive and negative voxels due to individual pipeline optimization, when compared to fixed pipeline optimization. In addition, the advantages of individual optimization become greater for subjects...
with weaker patterns that produce relatively small numbers of significant positive and negative voxels with $Z > 3.0$ and $<-3.0$, respectively.

**Table 3.3 Individual subject pipelines versus fixed pipeline optimization**

Differences in reproducibility (R), prediction (P), and distance (D) between individually optimized subject pipelines and the optimal fixed pipeline (MC+smoothing+DET3)

<table>
<thead>
<tr>
<th>Metric changes</th>
<th>ΔR</th>
<th>ΔP</th>
<th>ΔD*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.065</td>
<td>-0.075</td>
<td>-0.035</td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>-0.028</td>
<td>-0.015</td>
<td></td>
</tr>
<tr>
<td>0.148</td>
<td>0.045</td>
<td>-0.154</td>
<td></td>
</tr>
<tr>
<td>0.024</td>
<td>0.018</td>
<td>-0.027</td>
<td></td>
</tr>
<tr>
<td>0.076</td>
<td>-0.018</td>
<td>-0.064</td>
<td></td>
</tr>
<tr>
<td>0.073</td>
<td>0.015</td>
<td>-0.073</td>
<td></td>
</tr>
<tr>
<td>0.126</td>
<td>0.038</td>
<td>-0.132</td>
<td></td>
</tr>
<tr>
<td>0.015</td>
<td>0.007</td>
<td>-0.016</td>
<td></td>
</tr>
<tr>
<td>0.077</td>
<td>0.035</td>
<td>-0.084</td>
<td></td>
</tr>
<tr>
<td>0.014</td>
<td>0.022</td>
<td>-0.018</td>
<td></td>
</tr>
<tr>
<td>0.072</td>
<td>-0.074</td>
<td>-0.004</td>
<td></td>
</tr>
<tr>
<td>0.081</td>
<td>0.063</td>
<td>-0.102</td>
<td></td>
</tr>
<tr>
<td>0.023</td>
<td>-0.034</td>
<td>-0.009</td>
<td></td>
</tr>
<tr>
<td>0.023</td>
<td>0.038</td>
<td>-0.035</td>
<td></td>
</tr>
<tr>
<td>0.387</td>
<td>0.075</td>
<td>-0.380</td>
<td></td>
</tr>
<tr>
<td>0.123</td>
<td>0.031</td>
<td>-0.110</td>
<td></td>
</tr>
<tr>
<td>0.047</td>
<td>-0.046</td>
<td>-0.007</td>
<td></td>
</tr>
<tr>
<td>0.011</td>
<td>-0.018</td>
<td>-0.007</td>
<td></td>
</tr>
<tr>
<td>0.085</td>
<td>0.098</td>
<td>-0.109</td>
<td></td>
</tr>
<tr>
<td>0.272</td>
<td>-0.112</td>
<td>-0.173</td>
<td></td>
</tr>
</tbody>
</table>

* Fixed pipeline values are subtracted from individually optimized pipeline values. A negative D means the D-value decreased when pipelines are individually optimized, and R and P are closer to perfect.

Subjects in component 1 account for significantly more variance in the individually optimized SVD compared to the fixed pipeline SVD ($p < 0.05$, two-tailed paired Wilcoxon), see Fig. 3.12A. The amount of variance each subject accounts for increases in every subject in component 1 of the individually optimized SVD compared to the fixed pipeline SVD, indicated by the blue lines on the plot.
The amount of variance accounted for by subjects in component 2 of the individually optimized SVD is significantly higher than the fixed pipeline SVD (p<0.05, two-tailed paired Wilcoxon), see Fig. 3.12B. Subjects in component 2 account for less variance than in component 1, and 14/20 subjects account for more variance in component 2 of the individually optimized SVD compared to the fixed SVD, indicated by the blue lines on the plot.

Figure 3.11 Comparison of fixed and individually optimized significant voxels

The number of significant positive voxels with a z-score greater than 3 (A) and significant negative voxels, with a z-score of less than -3 (B) in single-subject SPMs from the NPAIRS analysis. The black line in the center is the line of identity and p-values are from two-tailed paired Wilcoxon tests (α=0.05).

Component 1 eigenimage results for fixed and individually optimized pipelines have similar positive and negative loadings, see Fig. 3.12(C). There are no differing regions of activation between the two components, although listed regions in Tables 3.1 and 3.2 differ because the voxels peak in different regions, while the clusters span the same regions.
Component 2 eigenimage results for fixed and individually optimized pipelines have common positive and negative activations in the regions shown in Fig. 3.12(D). The positive loading is similar, for both fixed and individually optimized pipelines, except in the temporal lobes. While Tables 3.1 and 3.2 both show peak activation in the parahippocampal gyrus, there is no overlap activation in that region. The two preprocessing methods have common positive temporal activation in the L and R lingual gyrus, R inferior temporal gyrus, and the R fusiform gyrus. Fixed preprocessing pipelines have unique positive activation in the L and R hippocampus, and in the L and R parahippocampus anterior to the parahippocampal activations in the individually optimized pipeline. Individually optimized pipelines have unique activations in the L and R parahippocampus posterior to the fixed pipeline activation, L inferior temporal gyrus, and L lingual gyrus.

Fixed and individually optimized pipelines have different regions isolated in the negative loading of component 2. There are no unique fixed pipeline activations compared to individually optimized pipelines. Unique individually optimized pipelines, compared to fixed pipeline optimization, are in the following regions: L and R rectus, R inferior temporal gyrus, and R inferior occipital gyrus.

Using AIC, it was determined that the regression between RTdif and component 2 subject weights for individually optimized pipelines was the best model, out of all possible combinations of the four component subject weights (components 1 and 2 for fixed and individually optimized pipelines) regressed with RTdif (Akaike’s AIC, 188.09).
Figure 3.12 Comparison of fixed and individually optimized SVD results

Similarities and differences in fixed and individually optimized SVD results. The variance each subject accounts for in the SVDs is plotted to compare fixed and individually optimized results for components 1 (A) and 2 (B). A red line indicates a subject accounts for less variance, and a blue line indicates a subject accounts for more variance, in individually optimized pipelines compared to fixed pipelines. P values are from paired Wilcoxon two-tailed paired Wilcoxon tests. Overlap of significant voxels in the SVD eigenimages between fixed and individually optimized pipelines is plotted for components 1 (C) and 2 (D). Red means an overlap in positive activation and blue means overlap in negative activation.
4 Discussion

The main goals of this study were to optimize fixed and individual-subject preprocessing pipelines on shortened tasks and compare how these influenced the resulting brain patterns. This study found that individual pipeline optimization, in comparison to fixed pipeline optimization, significantly increases reproducibility and subsequent model performance by decreasing D, significantly increases the detection of significant positive and negative voxels, and boosts the correlation between brain patterns and behavior.

Individually optimizing subject pipelines significantly increases reproducibility (which is correlated with fSNR) and decreases D compared to fixed pipeline optimization, replicating the results found in another data set by Churchill et al., 2011. The high (R,P) values found in this data are impressive considering that most studies use longer task durations compared to the tasks in this study which were under three minutes. This supports the view that preprocessing techniques may provide substantial boosts in fSNR that can compensate for scanning time per task.

The fixed preprocessing pipeline with the highest median rank includes motion correction, spatial smoothing and third-order detrending. However, the optimal fixed pipeline is not the optimal pipeline for any individual subject. In addition, there are a wide variety of steps included in the individually optimized pipelines. This supports the conflicting literature on the effectiveness of many preprocessing steps, providing reason to believe certain preprocessing steps can be detrimental to the fSNR of some data sets while causing significant improvements to the fSNR of different data. MPR, for example, decreases the median rank of most fixed
pipelines. As previously mentioned, MPR has been found to reduce activation strength in some studies (Johnstone et al., 2006) and reduce unwanted variability in other studies (Evans et al., 2010). Specifically, Evans et al (2010) found MPR to increase group reproducibility, while Churchill et al. (2011) found MPR causes a reduction in median rank in the fixed pipelines of an entire subject group, and causes an increased median rank for higher motion subjects when subjects are split into high and low motion subgroups (Churchill et al., 2011; Evans et al., 2010). Since the dataset in this study has relatively low motion overall, subjects who were individually optimized with MPR did not have significantly higher motion. The low motion seen in the recognition data can be partly attributed to the thorough simulator session preceding the fMRI session.

PNC also decreases median rank in most fixed pipelines, even though 8/20 subjects include it in their optimal individual pipeline. This could be due to the type of PNC used. PHYCAA has been found to be a better method for reducing physiological noise than the RETROICOR approach used here (Churchill et al., 2012) and perhaps the use of this method would have led to a further boost in (R,P). In addition, any fixed pipeline that included both PNC and STC was suboptimal suggesting that they are interacting in ways that have yet to be fully investigated and understood. 6/20 pipelines included both PNC and STC, 5 of which were optimized with STC preceding PNC. It has been shown that the variance of cardiac rate is related to the order of STC, PNC and MC (Churchill et al., 2010). The standard deviation of cardiac rate was variable across subjects, relative to the standard deviation of respiration rate. This might be why more subjects were optimized with STC prior to PNC. Further examination of this effect in our data might provide insight into the trends seen in the pipeline optimization results.
The D or (R,P), used to select optimal individual subject pipelines, can occasionally select a noisy pipeline as optimal if there is an artifact with a strong signal, i.e., artifacts can be highly spatially reproducible or support temporal prediction such as TCM. Reproducible spatial artifacts were seen in the initial SVD on the individually optimized data. Two subjects with individually optimized pipelines have motion artifacts, which are the two outliers identified by component 2 of the SVD (Fig. 3.9). The SVD is able to play a useful role in the visual identification stage, in addition to brain pattern identification. Even though the optimization of pipelines through (R,P) has been shown to be an effective method, it is likely to be difficult to have an optimization process that is completely automated, which was demonstrated in this dataset as well as others (Churchill et al., 2011). The only current way to ensure the optimization process selects high fSNR, artifact-free images, is by visually checking the resulting single-subject SPMs. Otherwise, the results may be at risk of being contaminated by obvious artifacts.

Individual pipeline optimization results in improved contrast SPMs compared to fixed pipeline optimization, because the number of voxels greater or less than a z-score of 3 and -3 respectively, is significantly larger for individually optimized pipelines. This indicates that the contrast SPMs from fixed pipeline optimization are suboptimal compared to individual pipeline optimization. This result was especially strong for subjects with weaker patterns, i.e. a reduced number of significant positive and negative voxels. This might also explain the distribution of subject weights in component 2 of the SVD: component 2 might reflect poorer preprocessing of the fixed pipeline data, and individual variability in the corrected SVD on individually optimized pipeline data. Further evidence is provided by the result that the amount of variance
each subject accounts for in the individually optimized SVD is significantly higher than in the fixed pipeline SVD in components 1 and 2 (Figs. 3.12A and B).

Fixed and individual pipeline optimizations have similar patterns of activation in component 1 of the SVD, see Fig. 3.12C, even though individual pipeline optimization significantly increases the amount of variance each subject accounts for, see Fig. 3.12A. Positive regions of activation are in the visual cortex, i.e. bilateral inferior, middle and superior occipital gyri, as well as the right fusiform gyrus, which has been shown to be active during visual perception and imagery tasks (Huijbers, Pennartz, Rubin, & Daselaar, 2011). The left and right precuneus, part of the retrieval success network (RSN) (Israel, Seibert, Black, & Brewer, 2010) are also active in component 1 for both fixed and individual pipeline optimization. Regions activated in the negative loadings in component 1, reflect DMN activity with negative activation in the R inferior parietal lobule and L and R frontal poles (medial PFC). These regions are referred to as negatively activated regions, instead of deactivations because they are relative reductions in BOLD signal relative to the positive regions or “activations”, but are not necessarily neuronal deactivations. Some DMN regions that are typically activated, such as the posterior cingulate cortex, do not appear in this pattern. It has been found that DMN may be subdivided into multiple components, which could be why there some regions are missing from this negative pattern in component 1 (Andrews-Hanna et al., 2010).

The distribution of the SVD subject weights from the individually optimized pipelines appear to form two subgroups, shown by the colored groups in Fig. 3.10A: as component 1 subject weights get stronger, one group’s subject weights get more positive in component 2 and the other group’s subject weights get more negative in component 2. There are no subgroups in the fixed pipeline SVD results (Fig. 3.7A) where component 2 appears to be driven by outliers, not
a differential subject effect. This means that the individually optimized subjects who strongly relate to the component 1 brain pattern, increasingly relate to two different patterns, i.e. they strongly relate to the pattern in component 1 while they differentially relate to the positive or negative pattern in component 2. This may reflect interacting brain networks that are linked to behavioral performance through component 2.

An indicator of activation strength and individual variability is the regression of SVD subject weights from components 1 and 2 with behavior. The only relationship is with RTdif and component 2 subject weights in fixed and individually optimized SVD results. Considering all the reported relationships between regions of activation and retrieval success, it was surprising that accuracy is not correlated to the first or second component in the SVD results. While fixation accuracy scores are higher than recognition accuracy scores, recognition accuracy scores are quite high with a small range of values, indicating that subjects did not find the recognition task very difficult. The small range of accuracy scores, with many at ceiling, may be the reason why recACC does not correlate with the subject weights in this data set. The recognition task may have been too easy and the range of recACC values too constrained near 100% to correlate with brain patterns or SVD subject weights. However, it is possible that regions in the RSN would be related to accuracy scores if a region of interest analysis were applied to the data, but they were not examined in this study.

There are several differences between the recognition and fixation conditions. These differences produce the contrast in the brain patterns. When conditions are contrasted in analysis, the differences between the conditions are highlighted, not the similarities. Both conditions required motor responses. Fixation required the selection of the same randomly positioned target (fixation cross) from two scrambled images over multiple trials, while recognition required the
selection of a randomly positioned target that they had previously studied from two foils, over multiple trials. Recognition recruited memory, visual and object identification processes, while fixation recruited predominantly motor and visual processes. Since the analysis in this study contrasts recognition versus fixation conditions, using a behavioral measure that contrasts recognition versus fixation was tested in addition to individual RTs. Thus the significant regression relationship between component 2 subject weights and RTdif, and not the other behavioral measures such as recognition or baseline RT, might be a reflection of the recognition versus fixation analysis contrast.

As mentioned above, component 2 of the fixed pipeline SVD seems to be outlier driven. Even though there appears to be a relationship between component 2 subject weights and RTdif, it appears the two outlying subjects drive the relationship. However, they were not determined to be outliers for the behavioral relationship by Cook’s distance. Cook’s distance is calculated by removing each individual point from the dataset, one at a time, and seeing the effect it has on the relationship between the two variables. Since there are two outliers in the same general direction, one will always be in the dataset when testing the impact of the other, i.e. the two outlying points are influential on each other and thus influence the overall dataset. Even though the two points were determined not to be outliers by the Cook’s distance metric, they appear to skew the relationship between the two variables in an influential way. These two subjects were not dropped from the dataset because they were not outliers in any other measure (i.e. physiology, motion, behavior and component 1 subject weights). Since SVD component 2 is isolating individual variability in a dataset with suboptimal preprocessing (i.e. fSNR is not maximized), subject heterogeneity in the results is not surprising.
As previously mentioned, subjects in component 2 of the corrected individual pipeline optimization SVD significantly account for more variance compared to the fixed pipeline SVD (Fig. 3.12B). These individually optimized subject weights are also significantly correlated with RTdif and account for 30% of the regression variance, while the fixed pipeline subject weights are barely significant at the 0.05 level and account for 20% of the regression variance. This indicates that individual subject pipeline optimization may be better at isolating brain patterns related to task behavior as it maximizes the amount of variance each subject contributes to the SVD while increasing individual variability in activation related to behavior. In addition, the one outlying subject was not determined to be an outlier via Cook’s distance, i.e. removing that subject did not significantly influence the regression relationship.

While there are some similarities in the component 2 brain patterns between fixed and individually optimized pipelines, as shown in Fig. 3.12D, there are also many differences. A large majority of the differences between the two patterns are located in the temporal lobes. For example, the fixed pipeline pattern identified bilateral parahippocampal activations anterior to the clusters found in the individually optimized pattern, in addition to hippocampal activations. On the other hand, the individually optimized pattern found activation bilaterally in the parahippocampus, L inferior temporal gyrus and L lingual gyrus. While hippocampal activation, is found in component 2 of the fixed pipeline SVD, it is not found in the individually optimized pipeline SVD. As mentioned in the introduction, this particular memory task was expected to utilize familiarity-based memory processes, not recollection. This task did not require the use of strong memory; we did not expect to see hippocampal activity since familiarity is not typically associated with hippocampal activity (Ranganath, 2010). However, there is substantial evidence suggesting the SVD results for fixed component 2 are largely
affected by outliers. Thus, these patterns in the fixed component 2 SVD are more likely noise than a differential hippocampal activation effect.

Since component 2 of the fixed pipeline SVD is likely driven by subject heterogeneity due to suboptimal preprocessing, and the individually optimized component subject weights were chosen as the ‘best’ regression model via AIC, the following discussion of component 2 is for the corrected individual pipeline optimization only. Although there are some noise clusters in component 2, there is still an interesting brain pattern that significantly relates to behavior via the subject weights. The noise does not seem to confound the results since the pattern in the eigenimage results is comparable to brain patterns found in similar studies, however, additional methods to remove it, such as PHYCAA and more conservative hand editing of brain masks, should be considered for future analyses. The positive pattern reflects activations for subjects that tend to have lower RTdifs with smaller differences in reaction time between recognition and fixation, and the negative pattern reflects activations for subjects that tend to have increased RTdifs with increased differences in reaction time between recognition and fixation.

As previously mentioned, there is a trend in the behavioral data where subjects with faster RT in fixation, slow down in their recognition RT (Fig. 3.3A). Many subjects with positive weights in component 2 of the SVD have smaller RT differences. However, these subjects do not necessarily have the smallest or largest fixation or recognition RT, they just have smaller differences between their RTs. It appears this trend is subdividing subjects into roughly two groups: subjects who perform the two components of the task at different rates and subjects who perform the two components of the task at steadier rates. These two subgroups in the data are most apparent in the scatter plot between component 1 and 2 subject weights (Fig. 3.10A).
Perhaps subjects who have larger differences in RT are using one retrieval technique during recognition, while subjects with smaller RT differences who remain at a more constant pace, use another retrieval technique. Since the subject weights in component 2 of the SVD correlate with this split, further evidence is provided that there are 2 subgroups of subjects using different strategies to complete the task, in addition to the evidence of a differential network interaction in component 2 versus component 1 (Fig. 3.10A). This difference in component 2 subgroups seems to reflect a differential balance between predominantly parietal, frontal and parahippocampal activation with smaller recognition-fixation RT differences versus more emphasis on predominantly right lateral activation including fusiform, inferior and middle occipital, inferior temporal, superior frontal and rectus gyri in the larger recognition-fixation RT differences.

Subjects with small RTdifs have activation in the visual cortex (right superior and middle occipital gyri), as well as activation in regions related to motor responses such as the supplementary motor cortex. In addition, there is activation in the right parahippocampal gyrus and supramarginal gyrus, which is related to familiarity and retrieval success (Buckner et al., 2008; Daselaar et al., 2009; H. Kim, 2010; Ranganath, 2010). Contrary to typical recognition patterns, there is activity in the left middle frontal gyrus, which is related to forming memories in encoding (L. Nyberg et al., 2000). Perhaps this activity is due to the hypothesis that the localization of activation during successful encoding can vary based on the way retrieval is cued, and that the original pattern of activity during encoding can be partially reinstated if retrieval is successful (Rugg et al., 2008). Thus, subjects who perform recognition at a faster rate relative to their fixation reaction time appear to have a pattern that reflects successful retrieval and familiarity. The short length of the task and the small range of accuracy scores with many at ceiling are possible reasons this pattern did not also relate to the accuracy scores.
Subjects with larger RT differences also have activations in the visual cortex, such as the right and left middle occipital, and right inferior occipital gyri. The main cluster of activation is in the right fusiform gyrus, which is typically related to object recognition and visual processing of objects (Garoff, Slotnick, & Schacter, 2005; Simons, Koutstaal, Prince, Wagner, & Schacter, 2003). There is also activation in right frontal regions (the right rectus gyrus as well as the right superior frontal gyrus), which have been found to be preferentially involved in episodic retrieval (L. Nyberg et al., 1996; Tulving et al., 1994). In relation to behavior, subjects who perform recognition at a slower rate relative to their fixation reaction time appear to have a pattern that reflects visual processing and object recognition. This pattern could be related to weaker memory since subjects are using an inefficient strategy to perform the recognition task (i.e. recognition takes longer relative to fixation). These subjects could be taking longer to identify and process the object, requiring increased or prolonged use of the fusiform gyrus.

As hypothesized hippocampal activity, i.e. a recollection pattern, was not observed. The recognition task was not very difficult, due to task length and as reflected in the range of accuracy scores (Fig. 3.3B), leading to a pattern reflecting familiarity-based memory and object recognition.

Thus, it is possible that the split in RT differences, which correlates to subject weights, represents stronger recognition versus weaker recognition, or familiarity versus prolonged object recognition. Subjects that are increasingly stronger in component 1, differentially and increasingly grow in component 2 where subjects performing quicker in recognition relative to fixation show a pattern related to familiarity and retrieval success, and subjects performing slower in recognition relative to fixation show a pattern related to visual processing and object recognition. Individually optimizing pipelines enhanced this differential subgroup effect,
increasing individual differences in activation, which lead to inferences about how subject’s neuronal responses related to their scanner behavior.

These results demonstrate that shortened clinical-related tasks can be adapted for the fMRI scanner resulting in images with relatively high fSNR. The tablet allows subjects to complete the recognition/episodic retrieval task with minimal modifications compared to the classical way it is administered outside of the scanner. As well, the tablet does not appear to introduce extra noise or motion into the results even though a motor response with hand and some forearm movement was required.

4.1 Limitations

This study has a number of limitations. Although the specific goals of the study included examining single-subject preprocessing and analysis techniques, single-subject analysis results are often noisier than group analysis results. This might account for some of the noise seen in the data, such as the brainstem noise seen in the SVD eigenimages. Another reason the data had noise was due to the short task length. As previously mentioned, functional MRI tasks are usually longer in order to increase fSNR. The short task length partly reduced the range of accuracy scores for each subject, limiting the way behavior related to the brain patterns. However, these limitations, which stemmed from the task being shortened in order to simulate a clinical test environment, did not seem to seriously impact our ability to obtain useful results. In addition, the recognition task was not very difficult, which limited the range of accuracy scores and thus the conclusions that could be drawn from these scores.

Another limitation of this study is the sample size. Unfortunately, 7 subjects had to be dropped from the group due to abnormal behavioral results, task-correlated motion and missing
physiological information, decreasing the sample size from 27 to 20. A larger group size would have been more ideal for exploring the behavior-brain relationships, although our group of 20 subjects was still able to identify useful, significant results. It was critical to drop the 4 subjects with behavioral abnormalities because abnormal behavioral scores indicate something else was going on during the task, whether it was equipment error, subject difficulties with the equipment, inattention or fatigue. These are primary indicators that the subject will have brain activation patterns that are at least somewhat unrelated to the task due to distraction. Dropping the one subject with TCM was justified because their fMRI results were likely to be strongly influenced by motion artifacts.

4.2 Future Directions

The results of this study have led to important questions about how fMRI patterns change with individual pipeline optimization and how this optimization can improve brain-behavior relationships.

Although there is relatively low motion in the data (Fig. 3.4), motion correction was one of the two most important preprocessing steps in individual pipeline optimization (Fig. 3.8), and has been shown by others to be one of the most important preprocessing steps available (e.g., Churchill et al., 2011). This led us to test if there are better ways to optimize this preprocessing step.

In preliminary testing we found that optimizing motion correction by registering each scan in a run to the one with minimum mean displacement significantly reduced the median maximum
displacement in both runs of the encoding and recognition tasks for young subjects. Altering motion correction significantly reduces median maximum displacement in the first run of encoding by -0.2682 ± 0.2054mm ($p<0.0001$, two-tailed paired Wilcoxon) and the second run of encoding by -0.2230 ± 0.2681mm ($p<0.001$, two-tailed paired Wilcoxon). This result replicates in the recognition data, with altered motion correction significantly reducing median maximum displacement by -0.1416 ± 0.1547mm ($p<0.001$, two-tailed paired Wilcoxon) in the first run of recognition and by -0.1507 ± 0.2608mm ($p<0.001$, two-tailed paired Wilcoxon) in the second run of recognition.

While the reduction of speech-related motion was not addressed in this thesis, altering the motion correction procedure may provide a simple yet significant way of reducing the amount of motion in the data. Careful target scan selection for within-subject alignment means each individual volume moves less during motion correction, perhaps decreasing potential distortion and error due to the alignment technique. As well, we believe median maximum displacement provides a better estimate of the amount of motion occurring in a run when compared to the maximum displacement within a run. For example, a subject might have one sharp movement reflected in the maximum displacement, but low motion for the rest of the run. If maximum displacement is used as the main criteria for assessing motion, that subject’s run might be dropped. On the other hand, if median maximum displacement is used as the main criteria for assessing motion, a suitable technique for correcting motion that occurs in a brief time frame is dropping 1-2 volumes around that time point, and the rest of the data can still go through to analysis.

Unfortunately no other conclusions can yet be drawn from these preliminary results as this revised method has not been implemented in the full pipeline and analysis. Until its effects have
been evaluated based on its pipeline metrics such as reproducibility and prediction, there is no way to know how it affects fSNR. It would be useful to find out if it causes significant changes in the final pipeline optimization results, as it is a small, easy alteration to implement.

In addition to the preliminary results above, there are still many more questions that could be addressed within this dataset. A region of interest analysis could be applied to the data to see how the strength of activation in specific regions relates to scanner behavior. For example, the z-scored SPMs from the NPAIRS analysis might show changes in the parahippocampal gyrus that relate to individual differences in RT. This could support the hypothesis that two different strategies are occurring during recognition.

Next, it would be interesting to see how the pipeline results change if the new data-driven PHYCAA physiological noise correction approach was substituted for AFNI’s RETROICOR as Churchill et al. (2012) have shown that it significantly outperforms RETROICOR, which was used in this study. Perhaps it would lead to increases in (R,P), and the increased ability to identify behavioral relationships in fMRI data. In addition, since the variability of cardiac rates has been related to pipeline ordering, it could be useful to examine how cardiac variability affects pipeline optimization in this data.

Individually optimized pipelines have not been utilized in the overt speech literature. While the common preprocessing techniques on their own cannot remove speech-related noise, individually optimized pipelines should be able to cause significant increases in fSNR, increasing the detection of meaningful activation. Using individual pipeline optimization in conjunction with speech-related noise removal techniques, such as the ones by Birn et al. (2004) and Goodyear et al. (2004), could lead to reductions in speech-related artifacts (Birn et al., 2004;
Goodyear, Zhu, Brown, & Mitchell, 2004). There are many research and clinical studies that would benefit from robust speech-related motion removal techniques.

The results of this study leave us with questions and ideas for new studies. It would be interesting to extend these findings into the aging and stroke populations. These two groups are often plagued by noisy results (Seto et al., 2001). Individual preprocessing pipelines have the potential to increase the detection of activation and signal in these populations, allowing researchers to potentially draw stronger conclusions from additional data that might have otherwise been dropped as a result of unresolved artifacts.

It may be beneficial to alter this task for future studies. The ambiguity regarding subject strategy during the task, i.e. recollection versus familiarity, can be controlled for in recognition studies via a “remember/know” paradigm. Instead of having perceptual and semantic foils like this study, one picture would be shown to each subject and they would assess whether the picture was ‘remembered’, ‘known’, or ‘new’ (H. Kim, 2010; Spaniol et al., 2009). In addition, this was a relatively simple memory task; increasing the difficulty of the task may lead to stronger conclusions about the memory strategies subjects are employing to complete the task.

Finally, the applications of this study benefit the potential uses of clinical fMRI. As previously mentioned, clinical fMRI has the potential to aid in the diagnosis of age-related disorders such as Alzheimer’s disease, and the assessment of treatment in age-related disorders, such as stroke. The shortened tasks used in this study, in conjunction with the individual subject optimization techniques, have the ability to mimic neuropsychological assessments used in the clinic while producing fMRI images with high fSNR. These are the two main reasons clinical fMRI has not been implemented. In order to move forward with clinical fMRI, these methods would have to
be tested and refined on clinical populations. Overcoming these obstacles pushes fMRI one step closer to the clinic.

4.3 Conclusions

This study validates the use of individually optimized pipelines over fixed pipelines in fMRI studies. The results support the initial hypothesis that individual pipeline optimization would result in stronger and more reliable activation patterns compared to fixed pipeline optimization. This was supported by showing individually optimized pipelines lead to significant increases in reproducibility, significant increases in the detection of significant positive and negative voxels, and significant correlation of brain patterns with behavior.

In summary, when a pipeline is kept fixed, preprocessing steps are applied to all subjects even if the steps are suboptimal for those subjects. Some subjects might need more, less, or different preprocessing than other subjects. Keeping preprocessing pipelines fixed often introduces error and noise in subject images that is preventable by optimizing pipelines on an individual subject basis. This result is reflected in the significant increase in reproducibility, or decrease in distance from perfect (R,P), in individually optimized pipelines compared to fixed pipelines. As well, fixed pipelines do not maximize the detection of significant positive and negative voxels compared to individually optimized pipelines. This is why individually optimized pipelines detect significantly more significant positive and negative voxels compared to fixed pipelines, especially for weaker patterns.

Finally, individual pipeline optimization maximizes individual variability in brain patterns that relate to behavioral variables obtained during the task. This is demonstrated when individually optimized pipelines result in a significant correlation with the difference between recognition
and fixation RT while fixed pipeline optimization does not. The relationship between individually optimized subject weights from the SVD and RT differences revealed a split in the data where subjects who tend to perform recognition faster relative to fixation (smaller RTdifs) have a pattern related to retrieval success and familiarity, while subjects who tend to perform recognition slower relative to fixation (larger RTdifs) have a pattern related to visual processing and object recognition.

These results allow the conclusion that individual pipeline optimization leads to significantly stronger and more reliable activations correlated with scanner behavior compared to fixed pipeline optimization. In addition, these results are achieved on much shorter tasks that are more similar to classic neuropsychological tests than typical fMRI experiments. Functional MRI has not been implemented as a clinical tool because of discrepancies between neuropsychological assessments and common fMRI tasks, fMRI noise, low fSNR and fixed pipeline optimization. Since this study significantly increased fSNR by using individually optimized pipelines with shortened neuropsychological tasks, it is reasonable to start testing these methods on clinical populations to evaluate the outcome. Hopefully, we are one step closer to implementing fMRI in the clinical evaluation of age-related disorders.
References


Langley, L. K., & Madden, D. J. (2000). Functional neuroimaging of memory: Implications for cognitive aging. *Microscopy Research and Technique, 51*(1), 75-84. doi:2-6


subcortical structures and cerebellum but not in nonvisual cortex. *Journal of Cognitive Neuroscience, 9*(5), 624-647.


Appendices

Appendix A

In order to detect insignificant SVD components, a simulated dataset was created to model the results of a ‘random’ SVD. The modeled noise was created as following:

1. A sample dataset was created by simulating random Gaussian noise with a mean of 0 and variance of 1. This dataset had 20 ‘subjects’ with each subject composed with a different random noise estimate.

2. The group mask used for actual subjects was applied in order to create a noise ‘volume’ comparable in size to the subject SPMs.

3. The SVD was performed in Matlab on the 20 subject simulated noise dataset – producing eigenimages, singular values and component weights.

Intensity values from the eigenimages of all 20 components were examined, see Fig. A.1. All 20 curves for each component are similar to each other. This is expected since the simulated noise is random for each ‘subject’, i.e. the SVD did not reveal any internal structure. The curves appear to be centered upon 0, since the modeled noise has a mean of 0 and variance of 1.
Figure A.1 Histogram of all 20 components of the SVD eigenimages performed on a simulated 20 subject Gaussian noise dataset.
Appendix B

The following section describes the process in which significant components and null components were determined. The intensities of the fixed and individually optimized SVD results were plotted in the same way as the simulated results, see Fig. B.1. The first two components in the fixed pipeline results were visually different than the other components in the SVD. The individually optimized results had three components that were visually different from the other components. The distribution of the first few components appeared different from the others in the center and tails.

The 95% threshold values were calculated on all the SVD results: simulated, fixed and individually optimized. These were plotted against each other, see Fig. B.2. After component 4 in the fixed pipeline optimization, the data curves converge into the simulated noise curves, indicating that the components reflect random noise, see Fig B.2A. The same for the individually optimized pipelines, after component 4, see Fig. B.2B. Since the values for the later components match up with the simulated noise dataset, there is sufficient evidence to conclude that the components reflect random noise.

To provide further evidence for dropping components, several distributions were fit to each component in the real data in order to find the line of best fit. A t-location-scale distribution fits both sets of results very well. This distribution is different from a regular t-distribution, as it does not assume a mean equal to 0; it fits the location/mean instead. As well, it fits a scale parameter describing the spread of the distribution. The parameters from each curve for each component were plotted in order to see the trends in curve parameters across components. As
mentioned above, the way in which the components differed from each other visually was in the center and tails. Thus, the curve parameters and skewness of the curves should provide accurate

Figure B.1 Histograms of all 20 components of the SVD eigenimages for fixed (A) and individually optimized (B) datasets. The first three components were plotted in different colors compared to the other 17 plotted in black.
Figure B.2 95% Threshold intensity/eigenvalues from the SVD on fixed (A) and individually optimized (B) data. Intensity values from the Gaussian data simulation are on the plots. The positive Gaussian threshold is the blue line and the negative Gaussian threshold is the purple line.

estimates of these visual observations, see Fig. B.3. The curve parameters evaluated were log-likelihood, mean (equivalent to mu/location), sigma/scale (related to variance), skewness, and nu (degrees of freedom/shape). Log likelihood is the cumulative sum of the probability each
point belongs to the distribution. A higher number implies a better fit. Log-likelihood, sigma, and skewness appeared to capture the differences between distributions across the components the best, see Fig. B.3 (A-F).

In order to quantitatively define which components were significantly different from the null components, an upper and lower boundary were created. The upper bound is the third quartile plus 1.5 times the interquartile range and the lower bound is the first quartile minus 1.5 times the interquartile range. Straight black lines on the plots represent the upper and lower bounds for defining whether an individual point is different from the rest of the data.

The measures for fixed pipeline optimization were consistent across components for log-likelihood, Fig. B.3A, sigma, Fig. B.3B, and skewness, Fig B.3C: components 1 and 2 have different values from the other 18 components, according to the upper and lower bounds that were set. The other 18 components converge towards one value while the first two components have different values. Components 1 and 2 have better fits than the other components as identified by log-likelihood, narrower scale (sharper peaks), and higher skewness. There is one outlier in skewness at the lower bound (component 19). The plots provided reasonable evidence to conclude that components 1 and 2 contain significant patterns while components 3-20 do not in fixed pipeline and are null distributions (excluding component 19).
B.3 Plots of parameters from the fitted t-location-scale distributions to each component in fixed (A,B,C) and individually optimized (D,E,F) pipelines. Upper and lower lines represent the third or first quartile, plus or minus 1.5 times the interquartile range respectively. Log likelihood (A), sigma/scale (B), and skewness (C) from the fitted t-location-scale distributions were plotted across all components for fixed pipeline optimization. Log likelihood (D), sigma/scale (E), and skewness (F) from the fitted t-location-scale distributions were plotted across all components for individually optimized pipelines. Points
for components outside the upper and lower bounds are considered different from the rest of the components.

The measures for individual pipeline optimization were not as consistent. Log-likelihood, Fig. B.3D isolated components 1, 2, and 3 as different from the other 17. This means that their t-location-scale distributions have better fits than the other components. However, sigma isolated components 1 and 2 as having different scale than the rest of the components, although component 3 was on the border of the boundary, Fig. B.3E. Skewness, on the other hand, only distinguished component 1 from the other 19 components, see Fig. B.3F. Even though these measures are not consistent, individual pipeline optimization has more variability between subjects, and while the distributions for components 1-3 may not differ from the other 17 in all measures, they still appear to be characteristically different. The plots provided reasonable evidence to conclude that components 1, 2 and 3 contain significant patterns while components 4-20 do not in the individually optimized pipelines and are null distributions.

The plots in Fig. B.3 were used to obtain the components that were truly random, i.e. the ones that stopped fluctuating and converged to a similar value, since some null components can contain random structure. These were selected from the skewness plots, Figs. B.3 C and F, where the components converge around 0. These were components 9-20 (except 19) in fixed and 10-20 in individually optimized pipelines. These null components were used to create a null distribution, for each set of SVD results, which was used to create the 95% threshold.