4-25-2009

Imaging schizophrenia: data fusion approaches to characterize and classify

Andrew M. Michael

Follow this and additional works at: http://scholarworks.rit.edu/theses

Recommended Citation
IMAGING SCHIZOPHRENIA: DATA FUSION APPROACHES TO
CHARACTERIZE AND CLASSIFY

by

Andrew M. Michael

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology

April 25, 2009

Signature of the author _______________________________________

Accepted by ___________________________________________

Coordinator, Doctoral Degree Program Date
The Ph.D. Degree Dissertation of Andrew M. Michael has been examined and approved by the dissertation committee as satisfactory for the dissertation required for the Ph.D. degree in Imaging Science.

Dr. Stefi Baum, Dissertation Advisor

Dr. Vince Calhoun, Dissertation Advisor

Dr. Pengcheng Shi, Dissertation Committee Chair

Dr. Peter Bajorski

Dr. Maria Helguera

Dr. David Messinger
I, Andrew M. Michael, hereby grant permission to the Rochester Institute of Technology to reproduce my dissertation in whole or in part. Any reproduction will not be for commercial use or profit.

Andrew M. Michael

Date
IMAGING SCHIZOPHRENIA: DATA FUSION APPROACHES TO CHARACTERIZE AND CLASSIFY

By

Andrew M. Michael

Submitted to the Center F. Carlson Center for Imaging Science in partial fulfillment of the requirements for the Doctor of Philosophy Degree at the Rochester Institute of Technology

Abstract

Schizophrenia is a complex, chronic and disabling mental disorder that affects about one percent of the adult population. The etiology of schizophrenia remains elusive and to date there are no image based tools to diagnose it. Advancements in magnetic resonance imaging (MRI) have enabled researchers to develop less invasive and \textit{in vivo} techniques, such as structural MRI (sMRI), functional MRI (fMRI) and diffusion tensor imaging (DTI), to construct theories about the neural underpinnings of schizophrenia. With sMRI, fMRI and DTI the distribution of tissues, the functional activity and the brain network are imaged respectively. Subjects with schizophrenia (SZ) and healthy controls (HC) are scanned with different modalities to identify differences, but the analysis of each modality has traditionally been carried out separately. Data fusion of multimodal data and an analysis of the joint information may hold the key to reveal hidden traces of this subtle disorder.

In this work we develop techniques to correlate sMRI with fMRI, fMRI with other fMRI and DTI with symptom scores. The brain is a highly interconnected organ
and local morphology can influence functional activity at distant regions. Through our methods it is possible to perform a cross correlation analysis between modalities incorporating all brain voxels. By reducing the large cross correlation matrix to useful statistics new aspects of schizophrenia are revealed. The methods introduced are simple, easy to implement and efficient. In another effort we modify canonical correlation analysis (CCA) to fuse two sets of brain data to locate brain regions with significant correlations. The new differential features identified through our fusion methods are used to classify subjects.

The sMRI–fMRI fusion indicates that the linkage between gray matter and functional activity probed by a sensorimotor task is weaker in SZ than in HC. Linkages between functional activity and structural regions in the cerebellum and the prefrontal cortex are found to be aberrant in SZ. The pair wise fusion of four different fMRI tasks shows that SZ activate to different tasks less uniquely than do HC. The above results support the ‘disconnection hypothesis’ of schizophrenia and the ‘theory of cognitive dysmetria’. DTI–symptom score fusion indicates that regions in the superior longitudinal fasciculus have high DTI–symptom correlations. Our preliminary classification efforts show high success rates in the leave–one–out scheme. The results presented in this work reveal several novel and interesting findings to better understand schizophrenia. The methods introduced are general, and can be easily applied to healthy and other pathological brain data to explore brain behavior.
Acknowledgements

Many have contributed towards developing this dissertation and during my stay at the Center for Imaging Science (CIS), Rochester Institute of Technology (RIT). My first gratitude goes to my advisor, Dr. Stefi Baum. When I was selected to work under her I felt honored since she was the director of CIS. At that time little did I know that Dr. Baum is one of the great advisors a graduate student can have. She made time to meet me every week in spite of her very busy schedule. Meetings with her always motivated me and refreshed my mind for new ideas. When results were not productive her encouraging words were sufficient to fuel research for the upcoming week. I am grateful to her for supporting me all the while even when our research took several different directions. Dr. Baum provided financial support through stipend, tuition and expenses for many conferences and training workshops. I have learnt many lessons, academic and otherwise, from her. Words cannot thank her enough for being a great advisor and a great person.

My next biggest ‘thank you’ goes to my dissertation advisor Dr. Vince Calhoun. I thank him for the confidence he had in me, to work on a brain imaging project while I did not have prior experience in brain imaging. He made substantial efforts to employ me as a research assistant at the Mind Research Network (MRN) while I was a student at CIS. During my work at MRN Dr. Calhoun provided financial support to continue my dissertation. Dr. Calhoun also made time to meet me every week amidst his very busy schedule. He is ready to help a student and encourages a student to meet and question him when in doubt. His efficiency in replying back to emails was remarkable and this
helped research progress faster. Dr. Calhoun provided resources such as books, software and training courses to help me learn a new field.

Members of the medical image analysis lab at the MRN were very helpful to me by sharing their knowledge and experience in brain imaging. They helped me to learn new brain imaging software and methods. Special thanks to Dr. Arvind Caprihan with whom I collaborated to work on diffusion tensor imaging. I learnt diffusion tensor imaging and its preprocessing software from Dr. Caprihan.

Members of my thesis committee have been helpful to me through their comments and guided me to complete my thesis. I thank Dr. Pengcheng Shi, Dr. Peter Bajorski, Dr. Maria Helguera and Dr. David Messinger who were on my committee. Their experience in diverse backgrounds was useful to tackle problems in brain imaging which is an emerging multi disciplinary field.

The CIS faculty at RIT helped me to complete the required course work. I learnt many aspects of imaging, from photons to processing, through these courses and wish to thank the faculty for their expertise in teaching these courses. Many others at RIT had helped me at CIS. Marilyn Lockwood has been very helpful in many different ways. Sue Chan was helpful to register for classes and giving out important information about CIS procedures. Jeff Cox, director of international student services at RIT, has been extremely helpful all through out my stay at RIT. He made sure that an international student at RIT felt at home by helping them through his gentle and benevolent nature. A big thank you to all my friends at RIT. Academic and non academic discussion with friends helped me to survive the long years of graduate life.
Last but not least I wish to thank my family without whom I could not have completed my doctoral degree. My parents have been always encouraging me and supported my pursuit for higher educations. A telephone call to them helped me to forget all frustrations and get back to work. I thank them for their motivation and never ceasing prayers. My biggest thank you goes to my wife Sangeetha. I cannot thank her enough for her patience in putting up with me during the many years of grad school. She took care of the family while I spent many hours of the day at the computer. Her positive thoughts and encouragements helped me immensely to reach the final goal. Very special thanks to our little daughter Aurea, for her smile would just erase all frustrations of graduate life.
To அம்மா ஆம்பா
(mom and dad)
# Table of Contents

Abstract........................................................................................................................................i

List of Figures ................................................................................................................................x

List of Tables ..................................................................................................................................xiii

List of Abbreviations ........................................................................................................................xv

1 Introduction .................................................................................................................................1

1.1 Mental Health ..............................................................................................................................4

1.2 Schizophrenia ..............................................................................................................................5

1.3 Diagnosis ....................................................................................................................................8

1.4 Brain Anatomy .............................................................................................................................9

1.5 Contribution to Imaging Science ...............................................................................................10

2 Brain Imaging .............................................................................................................................13

2.1 Fundamentals of MRI ................................................................................................................14

2.1.1 Place subject in a strong magnetic field (align protons) ......................................................15

2.1.2 Emit a radio frequency (RF) pulse (flip protons) ..................................................................17

2.1.3 Receive the reflected RF wave ..............................................................................................18

2.1.4 Construct three dimensional information ............................................................................20

2.2 Structural Magnetic Resonance Imaging (sMRI) .................................................................21

2.2.1 sMRI Imaging Parameters ...................................................................................................22

2.2.2 sMRI Preprocessing ............................................................................................................22

2.3 Functional Magnetic Resonance Imaging (fMRI) .................................................................24

2.3.1 fMRI Imaging Parameters ...................................................................................................27

2.3.2 fMRI Tasks ..........................................................................................................................27

2.3.3 fMRI Preprocessing ............................................................................................................32

2.3.4 General Linear Model (GLM) Analysis ................................................................................32

2.4 Diffusion Tensor Imaging (DTI) ..........................................................................................35

2.4.1 DTI Imaging Parameters ......................................................................................................38

2.4.2 DTI Preprocessing ..............................................................................................................39

2.5 Challenges in Brain Imaging ...................................................................................................40

2.6 Outlier Detection for Multisite Brain Data ............................................................................41

2.7 Data Fusion ..............................................................................................................................44

3 Methods ....................................................................................................................................47

3.1 A Method to Detect Outliers based on fMRI data ..............................................................48
3.2 Spatial Correlation Analysis (SCA) ........................................................................ 52
3.2.1 Histogram of $R_{12}$ ....................................................................................... 54
3.2.2 $Z$–score of $R_{12}$ along its rows/columns ..................................................... 57
3.2.3 $Z$–score of anatomical segments of $R_{12}$ .................................................... 60
3.2.4 Number of significant correlations in $R_{12}$ ................................................. 61
3.3 Multiple Linear Regression (MLR) .................................................................... 62
3.4 Canonical Correlation Analysis (CCA) ............................................................. 63
3.5 Tract based spatial statistics (TBSS) ................................................................. 68
3.6 Histogram Shift ................................................................................................. 70
3.7 Support Vector Machines (SVM) ..................................................................... 74
3.8 Discriminant Analysis (DA) ............................................................................. 77
3.8.1 Case 1: $\sum_i = \sigma^2 I$ .............................................................................. 80
3.8.2 Case 2: $\sum_i = \sum$ ............................................................................... 80
3.8.3 Case 3: $\sum_i = \sigma_i^2 I$ ........................................................................... 80
3.8.4 Case 4: $\sum_i =$ arbitrary .................................................................. 81
3.9 Classification based on DTI – PANSS Data ..................................................... 81

4 Structural – Functional Analysis ........................................................................ 83
4.1 Subjects ............................................................................................................. 86
4.2 Structural – Functional Spatial Correlation Histogram ..................................... 89
4.3 Correlation Histograms of Significant Voxels .................................................. 92
4.4 Spatial Location of Structural–Functional Correlations .................................... 95
4.5 Structural–Functional Inter Regional Correlations ......................................... 99
4.6 Regions with Significant Structural–Functional Correlations .......................... 102
4.7 Discussion ......................................................................................................... 103

5 Functional – Functional Analysis ..................................................................... 109
5.1 Subjects ........................................................................................................... 112
5.2 Behavioral Results .......................................................................................... 114
5.3 Inter–Task Correlation Histograms .................................................................. 115
5.4 Correlation Histograms of Significant Task Related Voxels ............................ 122
5.5 A Monte–Carlo Method Verification of Results ............................................... 124
5.6 Regions in the Brain Showing High Inter–Task Correlation ............................. 129
5.7 Discussion ......................................................................................................... 137
5.8 Classification .................................................................................................... 142
6 DTI – Symptom Scores Analysis ................................................................. 146
   6.1 Subjects .......................................................................................... 149
   6.2 Correlation with Atlas Regions ......................................................... 150
      6.2.1 Mean of Correlation Values ...................................................... 152
      6.2.2 Correlation of Mean Values ...................................................... 153
      6.2.3 Atlas Regions with Significant Correlation ................................. 154
   6.3 Multiple Regression for Each Voxel ................................................... 155
   6.4 Application of Canonical Correlation Analysis ................................. 157
   6.5 Classification .................................................................................. 160
7 Summary .............................................................................................. 163
8 References ........................................................................................... 169
List of Figures

Figure 1–1: The human brain and anatomical directions and planes.............................. 10
Figure 1–2: Steps of imaging .......................................................................................... 11
Figure 2–1: Behavior of hydrogen atom in normal and magnetic environment .......... 16
Figure 2–2: An MRI scanner .......................................................................................... 17
Figure 2–3: Behavior of hydrogen atom after an RF pulse ........................................... 18
Figure 2–4: Receiving RF signal and constructing the image ....................................... 20
Figure 2–5: Obtaining spatial information from an MRI signal .................................... 21
Figure 2–6: Structural MRI preprocessing Steps ......................................................... 24
Figure 2–7: Hemodynamic Response Function (HRF) ............................................... 25
Figure 2–8: fMRI activation map for the sensorimotor task ........................................ 26
Figure 2–9: Representation of the fMRI Tasks ............................................................. 30
Figure 2–10: Steps of the general linear model (GLM) ................................................ 35
Figure 2–11: Diffusion tensor measurement and DTI maps ........................................ 37
Figure 2–12: Average activation maps ($T>2$) for the sensorimotor task across the four different sites ................................................................................................................. 43
Figure 2–13: Spatial and temporal resolutions of different imaging techniques .......... 44
Figure 3–1: The $R$ value used for outlier detection for 85 patients from four different sites ........................................................................................................................................... 50
Figure 3–2: Modality 1 vs. Modality 2 Cross Correlation Matrix ($R_{12}$) .................... 53
Figure 3–3: Computing the histogram of $R_{12}$ .................................................................. 55
Figure 3–4: A scheme to check the validity of SCA histogram method ....................... 57
Figure 3–5: Collapsing $R_{12}$ along its rows/columns ................................................ 59
Figure 3–6: Segmenting $R_{12}$ to anatomical regions ...................................................... 60
Figure 3–7: Two sets of multivariate vectors to illustrate CCA correlations.............. 66
Figure 3–8: Tract Based Spatial Statistics (TBSS) Steps................................................................. 69
Figure 3–9: Histogram Shift........................................................................................................... 73
Figure 3–10: Support Vector Machine Classifier .......................................................................... 75
Figure 3–11: Conditional class probability density functions for classes 1 and 2................. 78
Figure 4–1: Structural–Functional Correlation Histogram with all Brain Voxels............. 90
Figure 4–2: Structural–Functional Correlation Difference (HC–SZ) Histograms with Voxels at Different Levels of Significance.......................................................... 94
Figure 4–3: Map of significantly correlated voxels (how a functional voxel is correlated to all structural voxels) for Healthy Controls (HC), Patients with Schizophrenia (SZ) and HC–SZ ...................................................................................................................... 96
Figure 4–4: Map of significantly correlated voxels (how a structural voxel is correlated to all functional voxels) for HC, SZ and HC–SZ............................ 98
Figure 4–5: Structural regions that have significantly higher ($P<10^{-6}$) correlations in HC than SZ ......................................................................................................................... 102
Figure 5–1: Flow chart of the functional–functional analysis.................................................... 112
Figure 5–2: Inter–Task Spatial Correlation Histograms with All Brain Voxels................. 116
Figure 5–3: Inter–Task Spatial Correlation Histograms with the Most Significant Task Related Voxels ....................................................................................................................... 123
Figure 5–4: Mean Inter–Task Spatial Correlation Histograms Found Through Monte– Carlo Test with $10^4$ voxels ........................................................................................................................ 127
Figure 5–5: Brain Regions with High ($p<0.01$) Inter–Task Correlations for the First Three Task Combinations.............................................................................................. 132
Figure 5–6: Brain Regions with High ($p<0.01$) Inter–Task Correlations for the Last Three Task Combinations.............................................................................................. 134
Figure 5–7: Classification cues for HC subjects ........................................................................ 142
Figure 5–8: Classification cues for SZ subjects ........................................................................ 143
Figure 5–9: The summation of cues ($F$) obtained from different task combinations .... 145
Figure 6–1: Flow chart of DTI–PANSS analysis........................................................................ 149
Figure 6–2: The corpus callosum segmented into five regions according to an existing atlas................................................................................................................................. 151
Figure 6–3: Mean of DTI–PANSS correlations for different atlas regions .................... 153

Figure 6–4: Correlation of mean DTI–PANSS for different atlas regions .................... 154

Figure 6–5: Voxels with significant ($p<0.05$) correlation with PANSS scores and FA and MD values ........................................................................................................................................................................... 157

Figure 6–6: JHU atlas regions with the largest number of significantly ($p<0.05$) DTI–PANSS correlations computed through CCA ................................................................. 160
List of Tables

Table 4–1: Demographics of SZ and HC with symptom scores for SZ......................... 88

Table 4–2: Significantly different brain regions between Healthy Controls (HC) and Patients with Schizophrenia (SZ) of how a functional voxel is correlated to all structural voxels................................................................. 97

Table 4–3: Significantly different brain regions between healthy controls (HC) and patients with schizophrenia (SZ) of how a structural voxel is correlated to all functional voxels ................................................................. 99

Table 4–4: Structural and functional regions that had a high correlation $z$–score difference, (HC–SZ) > 4.0 ................................................................................................................. 101

Table 5–1: Demographics and Clinical Characteristics of Patients with Schizophrenia (SZ) and Healthy Controls (HC) ......................................................................................... 114

Table 5–2 Behavioral Results for the AOD and SIRP Tasks........................................ 115

Table 5–3: Characteristics of Inter–Task Spatial Correlation Histograms with All Brain Voxels................................................................................................................................. 118

Table 5–4: Characteristics of Inter–Task Spatial Correlation Histograms with All Brain Voxels for Equal Numbers of Males (12) and Females (4) in Each Group ........ 121

Table 5–5: Characteristics of Inter–Task Spatial Correlation Histograms for Task Related Voxels................................................................................................................................. 125

Table 5–6: Characteristics of Mean Inter–Task Spatial Correlation Histogram Found through the Monte Carlo Test with $10^4$ voxels ................................................................. 128

Table 5–7: Five Largest Brain Regions with High ($p<0.01$) Inter–Task Correlations for the First Three Task Combinations ................................................................. 133

Table 5–8: Five Largest Brain Regions with High ($p<0.01$) Inter–Task Correlations for the Last Three Task Combinations ................................................................. 135

Table 5–9: Classification Results ................................................................................. 144

Table 6–1: Demographics and Clinical Characteristics of Patients with Schizophrenia (SZ) and Healthy Controls (HC) ......................................................................................... 150

Table 6–2: White matter JHU altas regions with high DTI–PANSS correlations ....... 155
Table 6-3: JHU atlas regions with the largest number of significant DTI–PANSS CCA correlation.......................................................................................................................... 159

Table 6–4: Classification success percentages with the leave–one–out scheme with different classifiers ............................................................................................................................... 161
List of Abbreviations

AAL: anatomical automated labeling
AOD: auditory oddball fMRI task
BOLD: blood oxygen level dependent
CCA: canonical correlation analysis
CSF: cerebro spinal fluid
DA: discriminant analysis
DSM: diagnostic and statistical manual of mental disorders
dof: degrees of freedom
DTI: diffusion tensor imaging
EPI: echo planar imaging
FA: fractional anisotropy
fMRI: functional magnetic resonance imaging
GM: gray matter
HC: healthy control subjects
HRF: hemodynamic response function
IA: the University of Iowa Hospital
MA: Harvard’s Massachusetts General Hospital
MCIC: mind clinical imaging consortium
MLR: multiple linear regression
MN: the University of Minnesota
MRI: magnetic resonance imaging
NM: the Mind Research Network
PANSS: positive and negative syndrome scale
PET: positron emission tomography
R/L: right / left brain hemispheres
RF: radio frequency
SCA: spatial correlation analysis
SCID: structured clinical interview for DSM
SIRP: Sternberg item recognition paradigm fMRI task
SM: sensorimotor fMRI task
sMRI: structural magnetic resonance imaging
SNR: signal to noise ratio
SPM: statistical parametric mapping toolbox
SVM: support vector machines
SZ: patients with chronic schizophrenia
T: Tesla
T1: longitudinal relaxation time
T2: transversal relaxation time
TBSS: tract based spatial statistics
TE: time to echo
TR: time to repeat
VBM: voxel based morphometry
WM: white matter
1 Introduction

Neural mechanisms behind many brain disorders, including schizophrenia, are not clearly understood (Hirsch and Weinberger 2003). The causes for these disorders remain even more obscure. The main difficulty in understanding mental disorders can be attributed to the complexity of brain structure and function. The other difficulty faced by researchers is the unavailability of diagnostic tools to uniquely identify different mental disorders. The heterogeneity within disorders and overlapping features between disorders make diagnosis of a particular disorder a difficult challenge. Current diagnosis of schizophrenia is based on patient’s self reported experiences and observed behavior. To date there are no image based or biology based laboratory tests for diagnosis. An analogy to the diagnosis of schizophrenia can be made with the following example.

Suppose a person hurts his/her leg and has a symptom of swelling in the injured region, one of the first and important steps the physician at the emergency room would take is to image the injured area with X–ray, magnetic resonance imaging (MRI) or a similar imaging modality. Results from the imaging test will help the physician to understand the nature of the injury and the cause for the symptom. The symptom of
swelling could have been caused, for example, by a bone fracture, a sprained muscle or a torn ligament. Based on this analysis the physician prescribes medication and/or therapy to the patient. A physician would almost never prescribe treatment based on the swelling, the symptom, alone. Such an action can only be due to the lack of appropriate instruments needed to image or due to the lack of knowledge to interpret the image acquired. Without a clear diagnosis recovery from the injury will be prolonged or never happen.

The state of diagnosis for schizophrenia, unfortunately, is at the stage of prescribing treatment based on symptoms alone. Strong analysis methods to diagnose which mental disorder causes the symptom or the knowledge of what neural/chemical processes are behind the symptoms are not available at present. Developments in neuroimaging techniques have enabled researchers in the past decade to perform in vivo studies of the structural and functional brains of patients with schizophrenia and to construct theories about its neural underpinnings. MRI is currently widely employed to image different brain tissue distributions and also to identify functional brain activation for a certain task. Images of patients with schizophrenia (SZ) and healthy controls (HC) are acquired and analyzed to identify differential features of structural and functional brain. The identified anatomical and functional measures may later be used as indicators of the disorder.

Between SZ and HC, studies have found differences in multiple brain regions and aberrant functional activity using a single modality, structural MRI (Pearlson and Marsh 1999), diffusion tensor imaging (Kubicki, et al. 2007) or functional MRI (Kindermann, et al. 1997). The reference cited for each modality is a review of previous study results of
schizophrenia using that modality. These results have not been sufficiently consistent to develop an image based diagnostic tool that can discriminate across subjects with wide demographics. It is becoming a common practice to scan subjects with different imaging modalities such as, structural MRI (sMRI), functional MRI (fMRI) and diffusion tensor imaging (DTI). Each modality carries both common and unique information about the brain. However, the data acquired are typically analyzed separately. Such approaches do not examine the joint information between the different modalities or tasks. Data fusion methods that enable the examination of joint information between modalities are needed and can lead to new understanding about brain behavior, especially about complex mental disorders. The brain is a vastly connected organ (Mesulam 1998) and it is reasonable to expect changes in brain morphology to result in modulations of brain activity at distant regions. It is hypothesized that neural mechanisms of schizophrenia are not circumscribed and patients may be characterized by a compromised brain network (Friston and Frith 1995). In this work we present methodologies to combine brain information from different modalities to find new features of schizophrenia.

Brain data obtained from three brain imaging modalities: sMRI, fMRI and DTI are analyzed in our work. Introduction to MRI, the three modalities with imaging parameters used in this study and preprocessing steps are presented in Chapter 2. Brain imaging studies are challenging due to subject variability and variability in activation patterns. Obtaining data with high fidelity is therefore not easy. In Chapter 2, we list challenges in brain imaging and in Chapter 3 we introduce a simple and efficient approach to detect outliers from multisite data (Michael, et al. 2009a). Introduction to and the development of methods used in this study are presented in Chapter 3 (Michael, et al.
Our work can primarily be divided into three main analyses categories: structural–functional (sMRI–fMRI), functional–functional (fMRI–fMRI) and DTI–symptoms. The sMRI and fMRI data used in this study were acquired as part of a large study on schizophrenia called the Mind Clinical Imaging Consortium (MCIC). The MCIC data were acquired from four different sites: the University of Iowa Hospital (IA), Harvard’s Massachusetts General Hospital (MA), the University of Minnesota (MN) and the Mind Research Network (NM). For the fMRI–fMRI analyses we used data from IA. DTI and positive and negative syndrome scale (PANSS) data were collected at the Olin Neuropsychiatric Research Center, Connecticut. Results and discussions of sMRI–fMRI fusion are presented in Chapter 4 (Michael, et al. in press; Michael, et al. 2008a; Michael, et al. 2008c). Results and discussions of fMRI–fMRI fusion are presented in Chapter 5 (Michael, et al. 2009b; Michael, et al. 2009c; Michael, et al. in review). DTI–symptom score results are presented in Chapter 6 (Michael, et al. 2008b; Michael, et al. 2009d; Michael, et al. 2009e). Attempts to classify subjects using the new features we discovered are also presented in this work.

1.1 Mental Health

The impact of mental disorders on productivity and quality of life is often underestimated. Increases in job related and other forms of stresses are detrimental for mental well being. The number of officially recognized mental disorders has been increasing over the past decades. In the U.S. mental disorders are diagnosed by the guidelines set by the Diagnostic and Statistical Manual of Mental Disorder Fourth Edition (DSM–IV)(Am.Psychiatric.Ass. 2000). At present there are three hundred and seventy
four mental disorders listed in the DSM–IV. Statistics show that an alarming twenty six percent of adults in the U.S. suffer from some form of diagnosable mental disorder in a given year (Kessler, et al. 2005). In the U.S and Canada mental disorders are the main cause of disability for ages 15–44 (WHO 2004). In another study (http://www.who.int/topics/) conducted by the World Health Organization, the World Bank and Harvard University reveals that the burden of mental illnesses on health and productivity is higher than the disease burden caused by all cancers combined.

1.2 Schizophrenia

Schizophrenia has been identified as one of the mental disorders that affect people across all cultures, sexes and socioeconomic groups (Bhurga 2005). It is a complex, chronic and disabling disorder and has caused distress to about 2.4 million (1.1% of adults) Americans (Regier, et al. 1993). Patients with schizophrenia (SZ) hear voices that others don’t, may believe that others are reading their mind or scheming to harm them. These experiences are terrifying and can cause fearfulness, withdrawal and some times agitations. SZ may sit for hours without moving or talking and may seem normal until they interact. Interactions with SZ will lead one to recognize that their conversations make little sense.

Symptoms of schizophrenia can be divided into three dimensions: positive, negative and disorganization (Andreasen 1982; Andreasen, et al. 1999; Andreasen and Olsen 1982; Liddle 1987). Positive symptoms include delusions, auditory hallucinations, thought disorder and disorders of movement. Positive symptoms are typically regarded as
the manifestation of the disorder. Negative symptoms, originally categorized to stress the
more enduring cognitive components, are considered the loss of normal traits and
abilities and include symptoms such as lack of emotion, lack of motivation (avolition),
poverty of speech (allogia) and failure to experience pleasure (anhedonia). Symptoms
such as affective incongruity, attentional impairment, problems with working memory
and executive functions needed to plan and organize, are categorized in the
disorganization dimension. SZ perform at lease one standard deviation below normal
population mean in various areas of neurocognitive and behavioral abilities (Heinrichs
and Zakzanis 1998), such as attention, memory, motor speed, executive functions, ability
to acquire skills, problem solving and community functioning (Green, et al. 2000).
Studies have shown that these impairments can even be present before the onset of
disease symptoms (Cornblatt and Keilp 1994).

The etiology of schizophrenia remains elusive, but evidence suggests it is due to a
combination of environmental and genetic factors. Ten percent of the patients have a first
degree relative with the disorder. The identical twin of a person with schizophrenia is at
risk of developing the disorder with a 40 to 65 percent chance (Cardno and Gottesman
2000). It is unlikely that genetics alone can cause schizophrenia. Interactions between
genes and the environment are thought to be necessary for schizophrenia to develop.
Environmental factors such as exposure to viruses, malnutrition in the womb, problems
during birth and psychosocial factors such as stressful environment are possible
triggering factors. Evidences suggest it is a neuro-developmental disorder (Murray and
Lewis 1987; Weinberger 1987) rather than a neurodegenerative disorder. Evidences
include the presence of premorbid indicators (Murray, et al. 1992; Murray and Lewis

Studies separately analyzing structural and functional images have found that multiple brain regions appear to be affected in schizophrenia (Goldstein, et al. 1999; Honea, et al. 2005; Niznikiewicz, et al. 2003). There is no significant structural brain difference in patients with schizophrenia and healthy controls. Sometimes the fluid filled cavities at the center of the brain called ventricles (see Figure 1–1b) are larger in patients, overall gray matter volume can be lower in SZ and certain areas in the brain can have less or more metabolic activity. Disturbances in WM connectivity between different brain regions, ‘disconnection hypothesis’(Friston 1998), is attributed as a possible cause for schizophrenia. Studies (Foong, et al. 2002; Minami, et al. 2003), both in vivo and post–mortem, have shown WM volume, fiber number or density differences between patients with schizophrenia and healthy controls possibly due to abnormalities in the myelin sheaths around the axons. Previous studies have found that there is diminished white matter anisotropy in prefrontal (Buchsbaum, et al. 1998) and other areas of schizophrenic
brain. White matter abnormality suggests an aberrant anatomical connectivity of the brain caused by a disturbance in myelin or fiber architecture.

1.3 Diagnosis

The clinical heterogeneity of schizophrenia, characterized by a myriad of symptoms and signs, presents a difficult target for research into its neurobiological mechanisms (Andreasen 2000). No pathognomonic marker of the disorder is present and it is hard to consistently find a restricted region in the brain that is solely responsible for the disorder. Schizophrenia is currently diagnosed on the basis of behavioral symptoms. Diagnosis of the disease can be a challenge due to the overlapping nature of the symptoms with other mental disorders like bipolar disorder (Pope 1983) and borderline personality disorder (McGlashan 1987). Advancements in neuroimaging have given hope of developing an image based diagnosis tool. HC in the study were screened to ensure they were free from Diagnostic and Statistical Manual of Mental Disorders (DSM–IV) axis I or axis II psychopathology, assessed using a modified version of the comprehension assessment of symptoms and history (CASH) (Andreasen, et al. 1992a). HC were interviewed to determine that there was no history of psychosis in any first degree relatives. SZ met criteria for schizophrenia based on structured clinical interview for DSM (SCID) or CASH and were confirmed by review of the case file.

The severity of symptoms in SZ is evaluated by asking a series of symptom related questions. Positive and negative syndrome scale (PANSS) is an evaluation scheme to estimate the severity of symptoms (Kay, et al. 1987). With this method three
different scores for positive (Andreasen 1984), negative (Andreasen 1981), and disorganization symptoms are evaluated. These scores are evaluated by a professional trained in psychiatry. We use PANSS in Chapter 6 to correlate with DTI measures.

### 1.4 Brain Anatomy

The brain is the most complex organ in the human body and is the main component of the central nervous system (Nolte and Angevine 1995). It is estimated that the human brain consists of 50 to 100 billion neurons with around 100 trillion synaptic connections. At the top of the brain lies the large cerebrum and at the rear of the brain, beneath the cerebrum lies the cerebellum (see Figure 1–1a). The cerebrum contains two large cerebral hemispheres separated from each other by a deep longitudinal fissure and connected to each other by a massive bundle of nerve fibers called the corpus callosum. The cerebrum is covered with a convoluted cerebral cortex, which is composed of gray matter (GM). The cerebral cortex is the exterior surface that covers the cerebrum and its convolutions help to increase its surface area. Under the cortex is the white matter (WM) which consists of myelinated fiber tracts (see Figure 1–1b). It is understood that the cortex or the GM regions are responsible for information processing and that the WM fiber tracts carries information to different parts of GM. Hence the WM brain regions can be thought of as the network of the brain. Each hemisphere of the cerebrum is divided into four lobes: frontal, parietal, occipital and temporal. These lobes are shown in Figure 1–1a along with the fissures or landmarks that separate them. The cerebellum is even more intricately convoluted than the cerebral cortex to achieve more surface area. The cerebellum has substantial connections with the prefrontal cortex and can perform
parallel processing supported by its array structure and large number of condensed cells. 

At the center of the brain are ventricles (see Figure 1–1b), which contain cerebro spinal fluid (CSF). CSF helps to bring nutrients to the brain, take away excrement from the brain and acts as a shock absorber to the brain. The conventional directions and planes used in medical terminology are illustrated in Figure 1–1c.

Figure 1–1: The human brain and anatomical directions and planes
(a) A lateral/sagittal view of the human brain with locations of the cerebrum, the cerebellum and the four main lobes of the cerebrum as indicated. (b) An axial slice of a structural human brain map with gray matter (peripheral regions), white matter (inner regions) and the ventricles indicated. (c) Anatomical directions and planes followed in medical terminology.

1.5 Contribution to Imaging Science

A simple imaging system can be divided into three main steps: image capture, image processing and image display (Schott 2007). In the recent past images are acquired from many different sources and applied in many different fields to interpret useful

---

\(^a\) From the 20\(^{th}\) edition of Gray’s anatomy, now available copyright free in Wikipedia

\(^b\) Image of the sagittal slice of human head from Microsoft Office free clip art gallery
information. Hence the above three steps can be extended by additional steps of image post processing and extracting useful information that go far beyond visual display of images. These steps are presented in Figure 1–2.

![Image Processing Diagram](image.png)

**Figure 1–2: Steps of imaging**

In the work performed in this thesis images are acquired in a non conventional approach, where photons were not involved in any of the steps in imaging. In Section 2.1 we try to draw an analogy between photonic imaging and MRI. Apart from MRI, there are many other non photonic imaging modalities such as ultra sound imaging, positron emission tomography, electroencephalography etc. In all these different imaging modalities the primary aim is to capture the physical structure of a fixed scene or the function of a moving or changing scene. In this work we use MRI (imaging modality) to image the human brain (scene). DTI and sMRI are used to capture the anatomy of the brain and fMRI is used to capture the change in brain activation pattern. The images were acquired from patients with schizophrenia (SZ) and healthy controls (HC). In this thesis methods are developed to perform image post processing to extract useful information
using data that were previously preprocessed. We develop post processing steps or algorithms to extract useful information about schizophrenia. Our post processing methods combine images of different modalities and this is a step further pursued with the main aim of extracting useful information.
Advances in imaging techniques have enabled researchers to develop \textit{in vivo} methods to enhance the knowledge of brain anatomy and its relationship to brain function. A range of imaging modalities including X–ray, computer tomography (CT), single photon emission computed tomography (SPECT), positron emission tomography (PET), electroencephalography (EEG), magnetoencephalography (MEG) and magnetic resonance imaging (MRI) are available to scan brain regions and their functional activation. Each of these modalities has their own advantages and disadvantages and provides both unique and overlapping information. Among these techniques structural MRI (sMRI) is currently widely used to identify morphologically different regions behind a diseased or injured brain. With sMRI the spatial distribution of different tissue types of the brain are mapped. Recent developments in spatial and temporal resolution of MRI have lead to new imaging modalities such as functional MRI (fMRI) and diffusion tensor imaging (DTI). During an fMRI experiment a subject is asked to perform a task while the scanner captures the oxygen level changes in the brain. DTI uses diffusion properties of water to map the locations of white matter fibers in the brain.
In this chapter a description of the basic principles behind MRI will be provided. A brief explanation of the fundamentals of sMRI, fMRI and DTI, along with their preprocessing steps will also be presented.

2.1 Fundamentals of MRI

Magnetic Resonance Imaging (MRI) is one of the most important discoveries in medical diagnosis (Hashemi, et al. 2004). It is an important tool in radiology and can be applied to image almost any part of the human body. In the recent past it has been widely applied to image the brain, for structure and function. The physics behind MRI can be very confusing and in this section we have attempted to give the fundamentals concisely. ‘The basics of MRI’ by Joseph Hornak (http://www.cis.rit.edu/htbooks/mri/) and ‘MRI made easy’ by Christoph Specht are two excellent interactive tutorials on the fundamentals of MRI. Some of the figures presented in this section were generated using information and illustrations presented in the above tutorials.

In photography the process of imaging an object can be described by the following steps. Light from a light source falls onto the object, the object reflects it, the sensor in the camera captures the reflected light and an image is finally constructed. MRI is analogous to the above steps. Low frequency radio waves fall onto the object to be imaged, the object reflects back waves, coils in the MRI scanner pick up the signal and finally an image is computed. In the above analogy, the process of MR imaging is over simplified, but in reality it is much more complicated. To understand the concepts easily we divide the process of MRI into four main steps.
Step 1: Place subject in a strong magnetic field

Step 2: Emit a radio frequency (RF) pulse

Step 3: Receive the reflected RF wave

Step 4: Construct a three dimensional image

2.1.1 Place subject in a strong magnetic field (align protons)

The signal for MRI is obtained from the hydrogen atom or the proton. Each proton possesses a charge and a spin around its own axis. Due to these two factors a small magnetic field is created around each proton as shown in Figure 2–1a. When the protons are placed in an environment where no magnetic field is present, the protons spin in random directions and their small magnetic fields cancel each other (see Figure 2–1b). When the protons are placed in a strong magnetic field, they act as tiny compasses and align their magnetic fields parallel to the external magnetic field. A slightly larger number of protons face towards the north magnetic pole than the protons that face towards the south magnetic pole (see Figure 2–1c). The protons that face north (aligned parallel) possess a lesser energy level than the protons that face south (aligned anti parallel). When a proton is placed in a magnetic field, not only do they spin about their own axis, but also begin to wobble or precess. This phenomenon is analogous to the wobbling of a spinning top. The rate at which the proton spins is given by the Larmor equation given by Equation (2.1), where $\omega$ is the angular precession frequency, $B_0$ is the strength of the external magnetic field and $\gamma$ is a medium dependent constant known as
the gyromagnetic ratio. The modern MRI scanners (see Figure 2–2) have a field strength of 1.5 or 3.0 Tesla (T) and they are about 10,000 times stronger than the earth’s magnetic field.

$$\omega = \gamma B_0$$ \hspace{1cm} (2.1)

![Figure 2–1: Behavior of hydrogen atom in normal and magnetic environment](image)

(a) Due to the charge and spin of the hydrogen atom (protons) a small magnetic field (shown by red arrow) is created perpendicular to the spin direction. (b) In an environment free of external magnetic fields to direction of magnetic fields created by the protons are in random orientations. (c) When placed in an external magnetic field (shown by blue arrow), protons align parallel to the external field (d) In an external field, the protons spin (spin axis shown by black arrow) and precess around an axis parallel to the external field.
Figure 2–2: An MRI scanner
The scanner shown is a Siemens Magnetom Trio scanner at the Mind Research Network. This scanner was used to collect some of the data analyzed in this study.

2.1.2 Emit a radio frequency (RF) pulse (flip protons)

As a result of the higher number of protons pointing towards the north magnetic pole, a net magnetic field is created along that direction (z–direction, see Figure 2–3a). We shall refer to that resultant magnetic field as the longitudinal magnetic field and its direction will be parallel to the external magnetic field (see Figure 2–3b). When an RF pulse is emitted, at the precession frequency of the hydrogen atom, two things happen. 1) Protons that were at the lower energy state and parallel to the external magnetic field, gain energy from the RF pulse flip to the opposite direction and align anti parallel to the magnetic field (see Figure 2–3c). 2) Protons that were precessing incoherently before the

---

*From www.mrn.org, used with permission*
RF pulse become coherent or they begin to precess in–phase (see Figure 2–3d). Due to the first phenomenon the longitudinal magnetic field is nullified and due to the second a new transversal (in the xy plane) magnetic field is created.

![Image of the sagittal slice of human head from Microsoft Office free clip art gallery](image)

**Figure 2–3: Behavior of hydrogen atom after an RF pulse**
(a) The system of conventional axes: z–axis is defined by the line that connects toe to head of the subject. The xy–plane is parallel to the z–axis. (b) Before the application of the RF pulse there is a net longitudinal magnetic field in the +z direction. (c) The application of an RF pulse flips some protons to the –z direction and this reduces the longitudinal magnetic field. (d) Application of RF pulse makes protons rotate in–phase and this creates a transversal (in xy–plane) magnetic field.

### 2.1.3 Receive the reflected RF wave

Before receiving the RF signal reflected from the subject, the transmitted RF pulse is switched off. In the absence of the RF pulse, protons that are in the higher energy state flip back to their original state. This results in a gradual increase of the original longitudinal magnetic field and is characterized by the longitudinal relaxation time constant ($T_1$). In the absence of the RF pulse, protons that were precessing in–phase, fan out to different phases. As a result, the net transversal magnetic field gradually

---

*Image of the sagittal slice of human head from Microsoft Office free clip art gallery*
diminishes and is characterized by the transversal relaxation time constant ($T_2$). Examples of these two relaxations are shown in Figure 2–4b. The changes in the longitudinal and transversal magnetic fields compose the MRI signal and are picked up by the receiver coils in the scanner. $T_1$ and $T_2$ vary for different tissues in the brain. For example white matter tissues have shorter $T_1$ and longer $T_2$ than gray matter. Time to repeat (TR) and time to echo (TE) are imaging parameters of the scanning sequence and can be controlled by the user. TR is the time interval between the applications of two consecutive RF pulses and essentially determines the temporal sampling frequency of the image. The signals are measured after a short time period after the application of the RF pulse. This delay is referred to as TE. The change of signal of tissue A and B as a function of TR (Figure 2–4c top) and TE (Figure 2–4c bottom) are presented. As indicated in Figure 2–4c the contrast between tissues A and B initially increase and then decrease for longer TR and TE. TR and TE are chosen to maximize the contrast between tissues of interest. In Figure 2–4d different contrast images are shown as a function of TR and TE values. The image obtained with short TE and short TR is known as $T_1$ weighted image, where the cerebrospinal fluid (CSF) comes out as dark regions (see Figure 2–4d bottom left). The image obtained with long TE and long TR is known as $T_2$ weighted image, where CSF comes out as bright regions (see Figure 2–4d top right). The above explanations are very short descriptions for the sole purpose of understanding the concepts. Finer details of MRI imaging are beyond the scope of this work (Hashemi, et al. 2004; Westbrook 1999).
Figure 2–4: Receiving RF signal and constructing the image
(a) In the absence of the RF pulse longitudinal magnetic field gradually increases and the transversal magnetic field gradually decreases. This change is the signal that is picked up by the RF receiver coils (b) Characteristic curves for the longitudinal (top) and transversal (bottom) magnetic fields. (c) Examples of difference in contrast between tissues A and B as a function of TR (top) and TE (bottom). (d) Four different contrast images for short or long values of TR and TE.

2.1.4 Construct three dimensional information

In Section 2.1.3 we understood the basics of what the MRI signal is and how contrast between different tissues are obtained. In this section we explain how positional information of the signals are obtained. From Equation (2.1) we know that the precessing frequency is a function of the strength of the external magnetic field. By changing the strength of the magnetic field it is possible to change the frequency of the RF wave emitted from a certain location in the subject. MRI images are acquired one slice at a time. For example if the images are acquired in the axial plane (xy plane) a gradient field is applied in the z direction. While acquiring a specific axial slice, by applying gradient fields in the x and y directions, x and y positional information is obtained in the frequency space. The frequencies of the received signal convey information about its

\footnote{From ‘The Basics of MRI’ (www.cis.rit.edu/htbooks/mri/), author J. P. Hornak, used with permission}
location. This process essentially collects spatial frequency information in the Fourier space of the image and can be converted to spatial information with a two dimensional inverse Fourier transformation (see Figure 2–5 b and c). By repeating this process for different axial slices a three dimensional image can be constructed.

![Figure 2–5: Obtaining spatial information from an MRI signal](image)

(a) An example of a gradient field where the field strength linearly changes. (b) Signal obtained from a sagittal slice. (c) By computing the two dimensional Fourier transform of the signal, spatial information of the sagittal slice is obtained.

### 2.2 Structural Magnetic Resonance Imaging (sMRI)

Structural MRI is the first form of MRI images obtained and it is the most widely used MRI technique including in clinical environments. The basics of sMRI were explained briefly in Section 2.1. In the next section we present the imaging parameters used to acquire the sMRI data used in this work.

---

\[d\] From ‘The Basics of MRI’ (www.cis.rit.edu/htbooks/mri/), author J. P. Hornak, used with permission
2.2.1 sMRI Imaging Parameters

A T1–weighted sMRI was acquired at each site. For MA and NM the scans were acquired on a Siemens scanner at 1.5 Tesla (T) with the following parameters: TR/TE = 12msec/4.76msec, bandwidth = 110, flip angle = 20°, slice thickness = 1.5 mm, voxel size = 0.625×0.625×1.5 mm³, FOV = 256, and the pulse sequence = gradient echo. For IA the scans were acquired on a GE Signa scanner at 1.5 T with the following parameters: TR/TE = 20msec/6msec, bandwidth = 122, flip angle = 30°, slice thickness = 1.6 mm, voxel size = 0.6641×0.6641×1.6 mm³, FOV = 170, and the pulse sequence = gradient echo. The MN scans were acquired on a Siemens Trio at 3T with the following parameters: TR/TE = 2530msec/3.8msec, bandwidth = 110, flip angle = 7°, slice thickness = 1.5 mm, voxel size = 0.625×0.625×1.5 mm³, FOV = 256, and the pulse sequence = MP–RAGE. All four sites in the MCIC acquire brain images in a coronal orientation.

2.2.2 sMRI Preprocessing

SMRI data were derived from T₁–weighted scans. For MA and NM, three T1’s were co-registered to each other and an average T1 was computed for segmentation and smoothing. IA and MN used only a single T1 image. Preprocessing was performed using the SPM5 software package (http://www.fil.ion.ucl.ac.uk/spm/software/spm5) and consisted of the following steps: motion correction, spatial normalization, spatial smoothing and segmentation (see Figure 2–6). Motion correction was performed using INRIalign, a motion correction algorithm that is unbiased by local signal changes (Freire and Mangin 2001; Freire, et al. 2002). The brains of subjects have different shapes and
sizes and need to be spatially normalized before analyses at a group level can be made. Data were spatially normalized into the standard Montreal Neurological Institute (MNI) space (Friston, et al. 1995). We then resliced the sMRI images to match the fMRI voxel size of $3.4 \times 3.4 \times 4 \text{mm}^3$. This step is not a required step and the methods we develop can be applied to sMRI data with the original resolution size. However we reduced sMRI voxel resolution to match that of fMRI to decrease the number of computations. Spatial smoothing reduce of the images is necessary to reduce noise in the images and also to reduce remaining inter-subject variability. Spatial smoothing was carried out by convolving the images with a Gaussian kernel of full width at half-maximum (FWHM) of $9 \times 9 \times 9 \text{mm}^3$. Previous studies had shown that FWHM of $8–10 \text{mm}^3$ are optimal for cortical gray matter in fMRI studies (White, et al. 2001). We used voxel based morphometry (VBM), where spatial normalization, smoothing and tissue segmentation are combined into one unified model (Ashburner and Friston 2005). VBM was used to segment the brain into probabilistic maps of white matter (WM), gray matter (GM) and cerebral spinal fluid (CSF), with unmodulated normalized parameters as previously applied in two large voxel based morphometric studies (Meda, et al. 2008; Segall, et al. 2008). For each brain voxel the three concentrations from these three maps add up to one. Finally the voxel size of all images was resliced to $3 \times 3 \times 3 \text{mm}^3$. Subject outlier detection based on sMRI data was performed using Pearson correlation, which compared the degree to which subjects are related to the group average smoothed GM map. For further details on sMRI preprocessing and outlier detection see (Segall, et al. 2008).
Figure 2–6: Structural MRI preprocessing Steps
Data from the scanner is primarily processed with the above steps to obtain gray matter, white matter and CSF concentration maps of the brain.

2.3 Functional Magnetic Resonance Imaging (fMRI)

Functional magnetic resonance (fMRI) imaging is a technique that is used to capture the functional activation of brain regions. During an fMRI experiment a subject is asked to perform a task while the scanner records the blood oxygen level dependent (BOLD) changes of different brain regions. With efficient imaging parameters (pulse sequences) it is possible to capture the functional activity of the whole brain within a few seconds. fMRI has become a popular functional neuroimaging technique since it is less invasive, less expensive and has better spatial and temporal resolutions than previous techniques such as positron emission tomography (PET). fMRI allows investigators to study groups of patients with schizophrenia and healthy controls and identify differential brain activation between groups.

In fMRI what is imaged is the hemodynamic response due to a neural activity (Huettel, et al. 2004). A neural activity at a certain brain region requires additional oxygen consumption and oxygen present in the neighborhood of that brain region is
depleted. Brain supplies oxygenated blood to activated regions to replenish the oxygen levels. This supply over compensates the oxygen intake of the activated brain region and causes a drop in the deoxyhemoglobin concentration. Oxyhemoglobin and deoxyhemoglobin possess different magnetic properties. A change in the ratio between oxyhemoglobin and deoxyhemoglobin causes a change in the magnetic properties of the environment around the activated brain region and this change is picked up by the scanner. In fMRI what is measured is not the neuronal firing but the hemodynamic response due to it and hence is considered an indirect imaging technique of functional activation. The blood oxygen level dependent (BOLD) response to a neuronal firing is sluggish and is characterized by the hemodynamic response function (HRF) as indicated in Figure 2–7. The HRF peaks only after about 6–9s after the initial neuronal firing and hence is the limiting factor of fMRI temporal resolution.

![Figure 2–7: Hemodynamic Response Function (HRF)](image)

HRF characterizes the change of blood oxygen level at a brain region after a neuronal firing at that brain region. The typical sluggish response is indicated in the figure.
In Figure 2–8 we present an example for an fMRI map while the subjects analyzed in this study performed the sensorimotor task (see Section 2.3.2.3). The functional map is group averaged and the regions that activate according to the task are found by thresholding the voxels. The fMRI activations (shown in red) are overlaid on the structural image (shown in gray). Activations are seen in the temporal, motor and cerebellar regions in the brain and correspond to the auditory and motor functions performed during the sensorimotor task.

Figure 2–8: fMRI activation map for the sensorimotor task
The brain is shown as slices acquired in the axial plane. Slices are arranged from the top to the bottom of the brain. The Functional activations are shown in red and are overlaid onto a structural image shown in gray. In this fMRI task a subject is asked to press the right thumb after a sound. Activations are seen in auditory and motor regions.
2.3.1 fMRI Imaging Parameters

In this section we present the imaging parameters used to acquire the data used in this study. Functional data were acquired at all four sites with EPI sequences on Siemens 3.0T scanners, except at the New Mexico site where a 1.5T Siemens scanner was used. Data were collected from each participant while performing three different fMRI tasks (see Section 2.3.2). The parameters for the functional scan are as follow: Pulse sequence = PACE–enabled, single shot, single–echo echo planar imaging (EPI), scan plane = oblique axial, AC–PC; repeat time (TR) = 2s, echo time (TE) = 30ms, field of view (FOV) = 22cm, acquisition matrix = 64×64, flip angle (FA) = 90°, voxel size = 3.4×3.4×4mm³, slice thickness = 4mm, gap between slices = 1mm, number of slices = 27, ascending sequential acquisition; bandwidth (BW) = ±100kHz = 3126Hz/pixel.

2.3.2 fMRI Tasks

For this study three tasks, Auditory oddball (AOD), Sternberg Item Recognition Paradigm (SIRP) and Sensory Motor (SM) were designed based on past research results and were expected to maximize the difference between patients and healthy controls. Abnormal information processing has been hypothesized to be a core deficit in schizophrenia (Braff 1996; Callaway and Naghdi 1982; Perry and Braff 1994). This model is developed from clinical observation that patients with schizophrenia frequently complain that they are bombarded with more stimuli than they can interpret. This deficit in information processing has consequences of delusion (misinterpretation), hallucination (confuse internal with external stimuli) or negative symptoms such as alogia, anhedonia, or avolition (cease to respond). The AOD task was selected as a probe to assess the
inefficiency in information processing. In AOD the infrequent task relevant stimuli elicits a positive brain potential, measured through EEG, referred to as P3 or P3b. P3 is one of the most robust functional abnormalities found in chronic medicated schizophrenic patients (Ebmeier, et al. 1990; McCarley, et al. 1993) and manifests as a decrease in the temporal lobe amplitude. Similar finding have been shown for fMRI data as well, again particularly in temporal regions (Kiehl and Liddle 2001).

Working memory, or the ability to hold a representation and perform cognitive operations allows individuals to formulate, modify and hold a plan in mind (Baddeley 1992). A defect in this ability can explain a variety of symptoms of schizophrenia such as disorganized speech and thought, avolition or alogia (inability to maintain a plan for behavioral activities), delusions or hallucinations (inability to reference a specific external or internal experience against associative memories) (Manoach, et al. 1997). In schizophrenia, working memory deficits have been demonstrated in medicated, unmedicated patients (Park and Holzman 1992) and in healthy relatives of schizophrenic patients (Park, et al. 1995). The SIRP (Sternberg 1966) task was used as a functional probe of working memory and fMRI activation differences have been previously identified for this task (Manoach, et al. 2000; Manoach, et al. 1999).

The SM task was designed to activate the auditory cortex robustly. It was initially designed for calibration purposes for assessing and controlling between–site, within–site and within–subject variability. Data collected from this task have shown significant group differences and hence are included in the analysis.
Prior to each scan all subjects were instructed until they were able to perform the task well enough. The participants were requested to actively participate and respond as quickly and accurately as possible.

2.3.2.1 Auditory oddball (AOD).

The AOD task (Figure 2–9a) stimulates a subject with three kinds of sounds: frequent standard stimuli (1000 Hz tones with probability \( p = 0.81 \)), infrequent target stimuli (1200 Hz tones, \( p = 0.09 \)) and infrequent novel stimuli (computer generated complex tones, \( p = 0.09 \)). Sound stimuli were incorporated into E–prime scripts (http://www.pstnet.com) run on a Windows machine and presented via sound insulated, MR–compatible earphones (Avotec, Stuart, FL). Stimuli were presented sequentially in pseudorandom order for 200ms each with inter–stimulus interval (ISI) varying randomly from 500 to 2050ms with a mean of 1200ms. A subject is asked to make a speeded button–press response with their right index finger through an MR–compatible input device (http://www.mrn.org/mind–input–device/index.php) upon each presentation of the target stimulus and no response is required for the other two stimuli. The target and novel stimuli sequences were exchanged between runs to balance their presentation and ensure any differences in activity evoked by stimuli were not due to the type of stimulus used. There were four runs and each run comprised 90 stimuli and lasted for about 3.2min. Each run consisted of a different number of targets with a total of 42 targets for all the runs.
2.3.2.2 Sternberg Item Recognition Paradigm (SIRP)

In SIRP (Figure 2–9b) a subject is required to memorize a list of digits, maintain the list in memory for a brief period of time and then decide if a probe digit was or was not in the list. The stimuli were projected onto a screen positioned on the head coil. Response reaction time and accuracy were recorded. Three working memory block types: high (5t: a list of 5 digits), medium (3t: a list of 3 digits) and low (1t: a single digit) were
used in this paradigm. Each run contained two blocks of each of the three conditions presented in a pseudorandom order and blocks of each condition alternated with fixation epochs. Each working memory block began with a ‘learn’ prompt that was displayed for 2s, followed by an encoding epoch of 6s consisting of the simultaneous presentation of a set of digits/digit displayed in red. After a 1s delay a 38s recognition epoch is followed in which a series of probe digits are sequentially presented in green lasting 1.1s each. Half of the probe digits displayed was targets (digits displayed in the encoding epoch) and the other half was foils. There was a random delay between each probe digit that ranged from 0.6 to 2.48s. Subjects were asked to respond with a right thumb trigger press if the probe digit was a target or with a left trigger press if the probe digit was a foil. This was followed by a fixation baseline epoch where a fixation cross was displayed in white for a randomized duration that ranged from 4 to 20s. Prior to each scan, the subject was instructed to relax and get ready for the next trial during the fixation epoch. They were instructed to respond as quickly as possible and were rewarded with five cents for each correct response. Each subject was scanned while performing three runs and each run lasted about 6min.

2.3.2.3 Sensorimotor (SM) Task

The SM task was designed to activate the auditory cortex robustly since previous studies (Calhoun, et al. 2007; Kiehl and Liddle 2001; Machado, et al. 2007; Schroder, et al. 1999) have shown abnormal activation in auditory temporal lobe regions in SZ. The SM (Figure 2–9c) task consists of an on/off block design each with duration of 16seconds. During the on–block, 200ms tones were presented with a 500ms stimulus
onset asynchrony (SOA). There were eight tones at different pitches along a scale. These tones were presented in ascending and descending cycles and this pattern continued for the ‘on’ block duration. This was followed by an ‘off’ block of 16s. After each tone a subject is required to press the right thumb using the input device. There were two runs and each lasted about 4min.

2.3.3 fMRI Preprocessing

Preprocessing of fMRI was performed using the SPM5 software package (http://www.fil.ion.ucl.ac.uk/spm/software/spm5). In addition to the sMRI preprocessing steps fMRI preprocessing has a few additional steps. Since the different slices of the brain were acquired at different time points they are interpolated to construct whole brain images at all time points and this is performed during the phase fix step. The phase fix step is performed as the first step before all other sMRI preprocessing steps (see Figure 2–6).

2.3.4 General Linear Model (GLM) Analysis

General linear model (GLM, see Figure 2–9) is a multiple regression analysis (also see Section 3.3) method to find task related activation maps similar to the map presented in Figure 2–8. If an investigator is interested in identifying only the spatial locations of brain activations during an iterative task, and not in change in activation over time, the GLM method can be employed. The task related time courses are recorded and convolved with a canonical hemodynamic response function (HRF, see Figure 2–7) to get
the regressors needed for the GLM analysis, as shown in Figure 2–10. The regression coefficients for each voxel are then found using Equation (2.2).

\[
y(j) = \hat{\beta}_0 + \sum_{i=1}^{m} \hat{\beta}_i x_i(j) + e(j)
\]  

In Equation (2.2) \( y \) is the time course of the observed image, \( \beta \) are the regression coefficients, \( m \) the number of regressors, \( x_i \) are the different regressors (made by convolving the HRF with the different task time course) and \( e \) is the residual error. The values of regression coefficients (\( \beta \)) are found using minimum least squared method to minimize the residual error. From the regression coefficients activation maps are made for each of the stimuli in a task. For example for the AOD task it is possible to make three activation maps corresponding to the target, novel and standard stimuli. The GLM is a widely used and easy to implement method to find regions in the brain that activate during a task. It has two main drawbacks. The assumption that all voxels for all different tasks will have the same HRF for all subjects is unlikely. The other limitation is that only task related activation regions can be found with this method since the regressors used are derived from task time courses. Many un–task related activations occur in the brain and it is not possible to map such networks or regions with this method. GLM is often used with other techniques since it provides a straightforward method to collapse the time domain.

For the AOD task regressors were created by modeling the target, novel and standard stimuli as delta functions (Figure 2–9a) convolved with the default SPM5 canonical differences of gammas hemodynamic response function (HRF). Scanner drift, a
A gradual increase in measured signal, was modeled by a high pass filter with cutoff at 128s. Contrasts between target versus standard and novel versus standard (hereafter referred to as target and novel respectively) were created for further analysis. SIRP and SM tasks were modeled as block designs as shown in Figure 2–9b and c respectively. For SIRP the ‘recognize’ epoch was considered where working memory was involved and this block was convolved with the HRF to make the model regressor. This was done for all three loads (5t, 3t and 1t) and the average map was taken for further analysis. For SM, the ‘on’ period, where ascending and descending tones were played, was considered as one single block and was convolved with the HRF to create the model regressor. Multiple runs of the tasks were averaged for each subject to make the final activation map. Four activation maps, two from AOD (target and novel) and one each from SM and SIRP, were used for inter–task spatial correlation analysis.
1. Make Model (1 or more)  

Task time course  

HRF  

\[ x_i(j) \]  

2. Data  

3. Fit Model to the Data at each voxel  

\[ y(j) \]  

\[ y(j) = \hat{\beta}_0 + \sum_{i=1}^{M} \hat{\beta}_i x_i(j) + e(j) \]  

Regression Results  

**Figure 2–10: Steps of the general linear model (GLM)**  

1) Task time courses are recorded for the different stimuli presented within a task and are convolved with the hemodynamic response function (HRF) to construct the necessary models \((x_1)\). 2) For each voxel location activation over time \((y_j)\) is measured using MRI. 3) The models \((x_i)\) are regressed against the observed signal \((y_j)\) to find coefficients \((\beta_i)\) for each model. Using \(\beta_i\) functional activation maps for the different stimuli can be constructed.

**2.4 Diffusion Tensor Imaging (DTI)**

MRI can detect signals from protons of the water molecule (Mori, et al. 2006). DTI is a MRI technique that takes advantage of the self–diffusion of water molecules in tissues. Water molecules, when in an environment with no barriers, diffuse randomly in Brownian movements. The diffusion process can be understood with the following
example. If a drop of ink is placed on a sheet of paper and if the paper is not composed of fibers, the ink stain spreads into a circular shape. This type of diffusion is called isotropic since the ink is equally spread in all directions. However, if the paper consists of fibers in a particular direction, the stain takes an oval shape elongated along the direction of the fibers. This is due to the fact that diffusion occurs faster along the fibers than in the direction perpendicular to it. This type of diffusion is called anisotropic. In brain tissues, water molecules have a tendency to diffuse along the myelin sheath of the white matter tracts and results in higher diffusion along the axis than in directions perpendicular to the axis.

One of the advantages of MRI is that water diffusion can be measured along a predetermined direction. The extent of diffusion depends on the direction of diffusion measurement. In DTI, diffusion is measured in multiple directions (see Figure 2–11a). From these measurements the three principal eigenvalues ($\lambda_1$, $\lambda_2$ and $\lambda_3$: $\lambda_1$ is the largest) and the corresponding eigenvectors are computed (see Figure 2–11b). With these measurements it is possible to construct a three dimensional ellipsoid that is representative of diffusion at a voxel position. Once these six parameters are computed, they are typically transformed into two types of maps. One is the anisotropy map (see Figure 2–11c) and the other in the color coded orientation map (see Figure 2–11d). In the anisotropy map the white matter appears brighter (whiter) than the gray matter since its environment is more directional. It should be noted that gray matter also contains axons, but due to poor spatial resolution (large voxel size) of DTI and the convoluted nature of its axons, the directionality is lost. From the color map it is possible to identify the orientation of the fiber tracts. Red, green and blue represent fibers running along the
right–left, anterior–posterior and inferior–superior axes respectively (see Figure 1–1). A direction that is not along one of the above axes is represented by an appropriate mixture of colors.

Figure 2–11: Diffusion tensor measurement and DTI maps
(a) Diffusion of water molecules measured in multiple directions. (b) The three primary eigenvectors and their corresponding eigenvalues are computed with measurements obtained from (a). (c) Fractional anisotropy map constructed using the eigenvalues. Dark and bright regions correspond to isotropic and anisotropic diffusion regions respectively. (d) Color map indicating the direction of water diffusion or the white matter tracts.

Reproduced from Mori, et al. (2006)
DTI is a macroscopic measurement of microscopic WM properties, such as, myelin density, axonal damage, size of axons and other properties (Pierpaoli and Basser 1996). Scalar measurements that capture these tissue properties can be derived from $\lambda_1, \lambda_2$ and $\lambda_3$. Four scalar measurements, fractional anisotropy ($FA$), mean diffusivity ($MD$), radial diffusivity ($RD$) and average diffusivity ($AD$) are computed using Equation (2.3) to Equation (2.6) respectively.

$$AD = \lambda_1$$  \hspace{1cm} (2.3)

$$RD = \frac{\lambda_2 + \lambda_3}{2}$$  \hspace{1cm} (2.4)

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$  \hspace{1cm} (2.5)

$$FA = \sqrt{\frac{3(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}$$  \hspace{1cm} (2.6)

$FA$ ranges from 0 (isotropic) to 1 (anisotropic) and is a measure of fiber organization. $FA$ decreases in most diseases associated with brain connectivity impairment. A decrease in $FA$ can be caused by an increase in $RD$ (demyelination), or a decrease in $AD$ (axonal loss) or by a combination of the two effects.

### 2.4.1 DTI Imaging Parameters

The DTI images for the studies performed in the thesis were obtained on a 3.0T Siemens Allegra scanner equipped with a single channel transmit/receive head coil using
a single–shot spin–echo EPI. All DTI measurements were obtained with a 12–channel RF coil for better SNR. The DTI data were collected along the AC–PC line, with a 128x128 matrix, TR = 5900ms, TE = 84ms, FOV = 256 mm, slice thickness = 2mm (isotropic 2mm resolution), 45 slices, number of excitations (NEX) =1, generalized auto–calibrating partially parallel acquisitions (GRAPPA) = X2, 30 gradient directions, with b = 800 s/mm$^2$, and b = 0 experiment repeated 5 times. Peripheral arterial pulse gating was used to minimize effects from cerebrospinal fluid (CSF) and blood flow. The total imaging time was less than 6min per subject. The sequence is repeated twice to improve SNR.

2.4.2 DTI Preprocessing

The DTI processing follows the FSL toolbox pipeline (http://www.fmrib.ox.ac.uk/fsl/). The data are motion and eddy current corrected by registering the DTI volumes to the first volume with no diffusion gradients (b = 0). This is done with the ‘flirt’ function of FSL using 12 degrees of freedom (dof) global affine transformation and a mutual information cost function (Jenkinson, et al. 2002). The diffusion tensor is calculated using the ‘dtifit’ function of FSL (Basser and Pierpaoli 1998; Basser, et al. 2000). The gradient table for the DTI experiment is corrected for the image orientation. This correction is necessary because in the Siemens system the diffusion gradients are along the magnet axis and do not move around with the image axis. MD, AD, RD and FA (see Section 2.4) images are calculated from the eigenvalue images. An image with the principal eigenvector is also saved in a format compatible with DTI–Studio (www.mristudio.org). The FA image is normalized to the MNI template using the non–linear registration algorithm available in the ‘fnirt’ function of FSL and the
same transformation is applied to MD, AD, and RD images. These spatially normalized images are then skeletonized to the MNI template FA skeleton using the ‘tbss’ (see Section 3.5) function of the FSL toolbox (Smith, et al. 2006).

All the above analyses are done immediately after the data for each subject are collected by the automatic processing pipeline. A quality control step of examining the image of each subject after it is registered to the MNI template is also performed. The group analysis for the DTI data is done either based on the Johns Hopkins University white matter atlas included in the FSL software or based on the skeletonized images.

2.5 Challenges in Brain Imaging

Brain imaging is a new and emerging field relative to other imaging fields. Most of the advancements in brain imaging with MRI were developed in the last decade. In spite of its infancy, its impact on understanding brain behavior has been remarkable. Brain imaging is also relatively more challenging since the object to be imaged is within a living human being and opportunities to make direct measurement of its properties are very limited. Due to its complex nature, quantizing variables that characterize brain behavior are not available at present. Efforts are made to reduce the impact of known variables that can be measured. For example motion correction, spatial normalization of different subject brains onto a common template, exclusion of subjects who do not execute the fMRI task based on behavioral data and other data quality measures are taken to improve data quality. Unfortunately most of the variables are difficult to measure. For example personality differences, emotional differences, the cognitive state or the
circumstantial difference can influence functional brain activity of the subject inside the scanner. Knowledge of how these variables influence functional activity is not clearly understood. There can also be variables that we are unaware of. In this study we acquired data from multiple imaging locations and many more complications arise in multisite data as explained in the next section.

2.6 Outlier Detection for Multisite Brain Data

Multisite brain image acquisition is required to collect data from a large number of subjects to access a population with different demographic characteristics. Data from a large pool takes into account the biodiversity of human subjects and increases the potential generalizability of a research result. However, multisite studies are difficult to standardize due to differences in scanner vendors, pulse sequences and distortion corrections. Therefore combining multisite data is not easy (Pearlson 2009; Styner, et al. 2002). Steps such as, standardizing acquisition parameters, study procedures and analysis pipeline, are taken to account for inter-site differences (Friedman and Glover 2006) but it is almost impossible to keep all variables identical. One of the problems with large studies is assessing data quality, specifically identifying outliers. For this work we define an ‘outlier’ as a subject whose fMRI activation map is substantially different from that of the multisite group mean.

There are multiple ways to detect outliers at an individual level. For example outliers can be removed based on motion parameters, behavioral data or by visual inspection of all subjects’ brain images. These methods are important and help a
researcher to collect high-quality data, but they do not help to directly detect subjects who activate differently from the group mean. One approach which can be used is to spatially correlate the activation maps of individual subjects to that of the group mean. This method does not overcome two potential issues. The first is that if all voxels are included for the spatial correlation analysis it is possible that the most active, but few voxels, contribute positively towards the correlation and the less active, but many voxels, contribute negatively. In addition, by including the whole brain, non-task related voxels contribute towards the correlation. To overcome these issues, we can threshold the most active voxels and then calculate the spatial correlation using just those voxels. However, slight variations of voxels that pass the threshold can create incorrect correlations. In multisite data this issue can be further exacerbated since the threshold to select voxels can vary between sites. This is illustrated in Figure 2–12, where we show the $T$ maps ($T>2$) of an auditory sensorimotor task (see Section 2.3.2.3) obtained from subjects across four different sites with a similar number of subjects in each site. It can be seen that the number of voxels that pass the $T=2$ threshold is not the same for the different sites (even after accounting for different degrees of freedom). If the same threshold is applied to select subjects from different sites we observe that different numbers of voxels are selected. If we lower the threshold then voxels from non-task-related regions are selected for some sites. To overcome this issue in Section 3.1 we introduce a method to identify outliers appropriately which also will minimize human intervention.
Figure 2–12: Average activation maps ($T>2$) for the sensorimotor task across the four different sites. For each site the average of activation maps across all subjects from that site was calculated. The maps were thresholded at a $T$ value of 2.0. Voxels that surpassed this threshold are shown for the different sites separately.

---

$^f$ Reproduced from Michael, et al. 2009a
2.7 Data Fusion

Each brain imaging modality has its own advantages and limitations and conveys different information about the brain. For example sMRI reveals the gray matter, white matter and CSF distribution in the brain. DTI shows the location of white matter tracts or the interconnections between different brain regions. fMRI shows brain activations regions while a subject performs a certain task. Figure 2–13 shows the approximate relative spatial and temporal resolution of each modality. An imaging technique that has both high temporal and spatial resolution is yet to be found.

![Spatial and temporal resolutions of different imaging techniques](image)

**Figure 2–13: Spatial and temporal resolutions of different imaging techniques**

MEG and EEG have high temporal and low spatial resolution. DTU and sMRI have high structural and low temporal resolution and fMRI has medium spatial and medium temporal resolution.

---

Adapted from Huettel, et al., 2004
Fusing different modalities or tasks has not been easy and currently each modality or task is typically analyzed separately. Fusion is a challenge in brain data since it is difficult to combine data when all the variables are not clearly understood. Data fusion methods are desired to understand hidden features and may hold the key to develop a diagnostic tool in the future.

A wide range of studies has shown a wide variety of structural and functional brain differences between patients and controls. These results are heterogeneous, widespread and are inconsistent when applied to different data sets. The incomplete information each modality or task provides can be the reason why it has been difficult to find consistent feature differences. Combining modalities and tasks is thus sought–after to get a more complete understanding of the disorder. It is expected that incorporating multiple modalities and tasks will increase the divergence between patients and controls.

There have been a few previous efforts to fuse brain data and they examine joint information using region based approaches such as structural equation modeling and dynamic causal modeling (Friston, et al. 2003; McIntosh and Gonzalez-Lima 1994) to look at the correlation structure between regions activated by different tasks (Rajah and McIntosh 2005) or between functional and structural variables (Meyer-Lindenberg, et al. 2004). They do not take into account all brain regions and do not provide an examination of all possible combinations of brain activity. Voxel based approaches (Worsley, et al. 1998) to compare correlations are straightforward but the results are difficult to visualize due to the enormous number of possible combinations. Transformation based approaches such as singular value decomposition (Friston, et al. 1996; Friston, et al. 1993), partial least squares (McIntosh, et al. 1996) and independent component analysis use tools
(Calhoun, et al. 2001; McKeown, et al. 1998) to transform matrices into a smaller set of modes or components.

In our work presented here we correlate structural data (sMRI and DTI), functional data (from different fMRI tasks) or behavioral data (PANSS). This work is referred to as ‘data fusion’ since the correlation of data and information from single (Klein 1999) or multiple sources (Defense 1991) is included in two definitions of the term (Klein 1999; U.S. Department of Defence 1991). These definitions are from the remote sensing community where data fusion is widely applied. Another example from another field (consumer market) where correlation is considered as data fusion can be found in (Cho, et al. 2003). While combining multiple fMRI data, even though the modality is same, it can be argued that the source of data (brain activation regions) are different since different tasks are used.
3 Methods

Methods used in this work can be divided into three main categories: data fusion, data reduction and data classification. Multiple methods were employed for the above tasks and the application of the methods varied depending on the type of brain data. Spatial correlation analysis (SCA) (Michael, et al. in press; Michael, et al. in review; Michael, et al. 2008c) is an original approach developed in this work and is used to pairwise fuse brain images and to reduce them to forms that are easier to examine. Multiple linear regression (MLR) (Michael, et al. 2008b; Neter, et al. 1996) and canonical correlation analysis (CCA) (Michael, et al. 2009e; Rencher 1998) were used to combine data from different modalities and then to reduce data based on correlation significance. Tract based spatial statistics (TBSS) (Michael, et al. 2009d; Michael, et al. 2009e; Smith, et al. 2006) was used to reduce white matter data, acquired with diffusion tensor imaging (DTI), to white matter skeletons. For classification, depending on the nature of features extracted, the methods varied. Histogram shift (Michael, et al. 2008a) is a method developed in this project that uses features extracted with the SCA method to classify chronic patients with schizophrenia (SZ) from healthy controls (HC). We use support vector machines (SVM) (Hastie, et al. 2001; Michael, et al. 2009d) and discriminant
analysis (DA) (Duda, et al. 2001; Michael, et al. 2009d) to classify SZ from HC with features extracted from DTI data. In this chapter a description of the above techniques will be presented. Before we get into the development of fusion methods, we introduce an outlier detection scheme to identify subjects that activate significantly differently from that of the group mean. This issue was explained in Section 2.6 and here we introduce the method.

3.1 A Method to Detect Outliers based on fMRI data

Steps for the automated outlier detection based on fMRI data are listed below.

Step 1: Find the mean activation $t$–map of all subjects from all sites. A $t$–map of each subject is composed of $t$–values at each voxel location. A $t$–value at each voxel location corresponds to the significance at which that voxel activates according to the fMRI task.

Step 2: From the mean activation map pick voxels that surpass a user defined threshold ($Th$) and with those voxels construct a map ($Mg$).

Step 3: Find the number of voxels ($n$) in $Mg$.

Step 4: For the $i^{th}$ subject pick the $n$ voxels with the highest $t$–values and construct a map ($Ms_i$).

Step 5: Find the ratio ($R_i$) between the volume of spatial intersection between $Mg$ and $Ms_i$ and the volume of $Mg$: $\text{Volume of (} Mg \cap Ms_i \text{)} / \text{Volume of } Mg$. 
**Step 6:** Select subjects if $R_i > R_{Th}$, where $R_{Th}$ is a user specified value.

In Figure 3–1, $R$ values are indicated for a total of 85 different subjects from the 4 different sites found using a $Th$ value of 3. We also indicate the subject with the lowest (red circle) and highest (green circle) $R$ values. In the brain images below the plot, we show Mg in yellow, Ms in blue and the intersection of Mg and Ms in red. An ‘outlier’, as indicated in the left image, has larger regions of blue and yellow and smaller regions of red. A ‘good’ subject has smaller regions of blue and yellow and larger regions of red. The blue regions indicate the locations of the most task relevant voxels in an individual subject and yellow that of the group. In a ‘good’ subject these two regions should overlap as shown in the right figure. Artifacts in the left subject were probably caused by motion.
Figure 3–1: The $R$ value used for outlier detection for 85 patients from four different sites

In the brain images the activation map from an individual subject ($M_s$) is shown in blue, that of the group mean ($M_g$) in yellow and the regions that intersect ($M_s \cap M_g$) in red. The $R$ value is the ratio between $M_s \cap M_g$ and $M_g$. On the left brain image (red arrow) the activation is shown for the subject with lowest $R$ (outlier across sites) and on the right brain image (green arrow) for the subject with largest $R$ (‘good subject).

* Reproduced from Michael, et al. 2009a
In Step 1 by averaging across all subjects and all sites we are able to find a spatial map with high signal to noise ratio (SNR). At Step 2 a researcher can set the threshold \( Th \) based on the level of significant activation needed or the extent of activation region desired. In Step 4 we select the \( n \) highest \( t \)-valued voxels since our interest is on selecting the most task related voxels from each subject. The rationale behind this approach is that if the putatively highest task related voxels do not correspond to \( M_g \) then that subject’s activation is unlike that of the group mean and may be considered an outlier. The issue of slight variations in activations of high \( t \)-valued voxels contributing to the correlation is avoided by finding the spatial intersection of the two maps in Step 5. At Step 6 the user can increase the fidelity of subjects selected by increasing \( R_{Th} \). Subjects with low \( R \) values can be identified as outliers and further visual inspections can be performed on these subjects.

The dataset of 94 HC and 85 SZ was applied to this method and \( R \) values for HC had a range of 0.23 – 0.7 and for SZ 0.08 – 0.7. Twelve HC and fifteen SZ that had \( R \) values lower than 0.4 (\( R_{Th} \)) were excluded from further analysis. The \( R_{Th} \) threshold was kept at this arbitrarily selected low value to avoid detecting too many subjects as outliers. We then selected 70 HC to match age and sex to the 70 cleaned data from the SZ group. The mean and standard deviation of the \( R \) value for the 70 selected subjects were 0.56 ± 0.08 for HC and 0.55 ± 0.07 for SZ.
3.2 Spatial Correlation Analysis (SCA)

The structural and functional brain images, at a spatial resolution of $3\times3\times3$ mm$^3$, contain about 153,000 ($53\times63\times46$) voxels. After the exclusion of non–brain voxels the final number of voxels is about 60,000 ($N$). Non brain voxels include regions that lie outside the brain (in the image that is acquired as a cube) and regions in the cerebrospinal fluid (CSF) that lie within the brain. Our goal was to determine the degree to which all voxels from Modality 1 were correlated (across different subjects) to all voxels from Modality 2. Here ‘Modality’ refers to images obtained from sMRI or fMRI data. Computing such interrelationship is difficult in practice due to the need to examine the relationships between tens of thousands of voxels. For the brain resolution used in this study a cross correlation matrix ($R_{12}$) with billions of correlations has to be constructed. Constructing such a matrix is not easy due to limitations in computer memory. A random access memory (RAM) of more than 28 Giga Bytes is required to construct $R_{12}$. Even if $R_{12}$ is computed, its interpretation will be based on some reduced statistics.

The three dimensional brain images from a certain subject group (SZ or HC) and from a certain modality were vectorized to construct a matrix where subjects ($n$ number of them) are stacked along the rows and voxels along the columns (see Figure 3–2). By this process, matrices $M_1$ and $M_2$ are constructed using data from Modality 1 and Modality 2 respectively. Let $X_i$ and $Y_j$ be the column vectors across subjects for the $i^{th}$ voxel from Modality 1 and the $j^{th}$ voxel from Modality 2 respectively. Our interest was in finding the correlation between $X_i$ and $Y_j$ using Equation (3.1) where $i$ and $j$ varied independently from 1 to $N$. In subsequent subsections of this section we introduce methods to find statistics of $R_{12}$ by iteratively computing it.
Figure 3–2: Modality 1 vs. Modality 2 Cross Correlation Matrix ($R_{12}$)

Voxels from Modality 1 are vectorized and placed along the columns, and subjects along the rows, to construct the $M_1$ matrix. Similarly $M_2$ is constructed with data from Modality 2. Here ‘Modality’ refers to sMRI or fMRI data. These two matrices are computed for SZ and HC separately. The desired cross correlation matrix ($R_{12}$) has correlations between all voxels from Modalities 1 and Modality 2. $N$ is the total number of brain voxels and $R_{12}$ is the matrix product of normalized matrices $M_1$ and $M_2$.

[Equation 3.1]

$\rho_{ij} = \frac{\text{cov}(X_i, Y_j)}{\sigma_{X_i} \sigma_{Y_j}}$

---

Reproduced from Michael, et al. in review

---

\[ R_{12} = M_1^T M_2 \]
3.2.1 Histogram of $R_{12}$

The histogram of all the elements of $R_{12}$ gives a general idea of how the correlations between Modality 1 and Modality 2 voxels are distributed. Steps to find the histogram of $R_{12}$ are listed below (See Figure 3–3).

**Step1:** Find the correlation of the 1st voxel from Modality 1 with all voxels from Modality 2 (Correlation of $X_1$ and $Y_j; j = 1$ to $N$) and store them in an array ($A_1$) of length $N$.

**Step2:** Compute the histogram ($H_1$) of $A_1$ at 100 bins equally spaced between $-1$ and $+1$.

**Step3:** Repeat Step1 and Step2 for the $i^{th}$ voxel from Modality 1, $i = 1$ to $N$. Each time a histogram is found, it is added to the histogram of previous step. If $H_i$ is the histogram of correlations of the $i^{th}$ voxel from Modality 1 with all voxels from Modality 2, then the final histogram can be given by $H = \sum_{i=1}^{N} H_i$. 
Figure 3–3: Computing the histogram of $\mathbf{R}_{12}$

Rows of $\mathbf{R}_{12}$ are iteratively computed and for each row its histogram between $-1$ and $+1$ is computed. Each time a histogram is computed, it is added to the summation of previous histograms.

---

Reproduced from Michael, et al. in review
This method essentially finds the histogram of $R_{12}$ ($N \times N$) by iteratively finding the histograms of its rows and finally adding them. In Figure 3–4 we show that the histogram computed through this method and the histogram of all the elements of $R_{12}$ is identical given that the binning intervals are the same. In that sense we do not lose information, but we are reducing $N \times N$ correlations of the $R_{12}$ matrix to 100 numbers of occurrences, a form that can be easily represented and analyzed.

In Figure 3–4 the histograms of each row of $R_{12}$ is shown as number of occurrences within different intervals. The lower ($L_1$), upper ($L_2$) bounds and the bin interval ($\Delta x$) are kept constant across the rows of $R_{12}$. In this work $L_1$ and $L_2$ are equal to $-1$ and $+1$ respectively, since they are the lower and upper bounds of the correlation coefficient and $\Delta x$ equals $(L_2 - L_1)/100$. The number 100 was selected arbitrarily. $H_i$ is the histogram of the elements of $i^{th}$ row of $R_{12}$ and $f_{ij}$ is the number of occurrences in the $j^{th}$ bin of $H_i$. Histogram $H_i$ is essentially a collection of occurrences ($f_{i1}, f_{i2}, \ldots, f_{i\Delta x}$). If $L_1$, $L_2$ and $\Delta x$ are kept constant and if $R_{12}$ is fully computed one can intuitively understand that the number of occurrences within a certain bin will be the summation of occurrences of the rows of $R_{12}$ within the same bin. Hence the summation of $H_i$ is identical to the histogram of $R_{12}$ if it had been fully computed.
Figure 3–4: A scheme to check the validity of SCA histogram method

Histogram (\(H_i\)) for the \(i^{th}\) row of \(R_{12}\) is computed between \(L_1\) and \(L_2\) at bin intervals of \(\Delta x\). \(f_{ij}\) is the number of occurrences of elements in the \(i^{th}\) row of \(R_{12}\) and the \(j^{th}\) interval of \(H_i\).

### 3.2.2 Z–score of \(R_{12}\) along its rows/columns

In section 3.2.1 we find the distribution of Modality 1 – Modality 2 correlations but the spatial information of where in the brain these correlations are located is lost. In this section we introduce a technique that will retain the average spatial information. This is achieved by collapsing either the rows or the columns of \(R_{12}\) (see Figure 3–5) and then reconstructing the z–score of each row or column on to a brain map. The z–score of a row or column is the mean of the row or column divided by its standard deviation. If \(R_{12}\) is collapsed along the columns, the value at a certain voxel in the brain map corresponds to the z–score of correlations of the Modality 2 voxel at that location with all Modality 1 voxels. If collapsed along the rows the value at a certain voxel corresponds to the z–score
of correlations of Modality 1 voxel at that location with all Modality 2 voxels. To examine how each Modality 1 voxel is correlated to all Modality 2 voxels a total of $N$ maps are needed and inspecting such a large number of brain maps is not an easy task. It should be noted that this method gives an average sense of all correlations. In this method we are reducing the $N \times N$ correlations of $R_{12}$ to a map of $N$ $z$–score correlations. Steps to find the $z$–score map is listed below.

**Step1:** Compute the 1$^{st}$ row/column ($A_1$) of $R_{12}$

**Step2:** Find the mean, standard deviation and $z$–score of ($A_1$) and store the $z$–score as the 1$^{st}$ element of array $Z$ ($1 \times N$)

**Step3:** Repeat steps 1 and 2 for the $i^{th}$ row/column, $i = 1$ to $N$.

**Step4:** Reconstruct $Z$ to a brain map
Figure 3–5: Collapsing $R_{12}$ along its rows/columns\(^d\)

A row/column ($A_1$) of $R_{12}$ is iteratively computed and for each row/column its $z$–score is calculated and stored in the $Z$ array. This array can be used to construct a brain map of $z$–score correlations. If collapsed along the rows an average map of how Modality 1 voxels are correlated to Modality 2 voxels can be obtained and if collapsed along the columns vice–versa.

\(^d\) Reproduced from Michael, et al. in review
3.2.3 *Z*-score of anatomical segments of \( R_{12} \)

In section 3.2.2 we retained spatial information of correlations in an average sense but lost Modality 1 – Modality 2 inter-regional correlations. In this section a method to store significant inter-regional correlations of Modality 1 – Modality 2 is introduced and the steps are listed below.

**Step1:** Segment \( R_{12} \) into 116×116 anatomical regions as defined by the anatomical automated labeling (AAL) atlas (Tzourio-Mazoyer, et al. 2002).

**Step2:** Cluster the elements of \( R_{12} \) into 116×116 segments based on AAL atlas regions.

**Step3:** Iteratively compute each segment of \( R_{12} \), its mean, standard deviation and \( z \)–score.

![Segmenting \( R_{12} \) to anatomical regions](image)

**Figure 3–6: Segmenting \( R_{12} \) to anatomical regions**

\( R_{12} \) is segmented into (116×116) clusters defined by an anatomical atlas. Each segment is computed iteratively and its \( z \)–score is stored.

---

\^ Reproduced from Michael, et al. in review
In Figure 3–6, various segments, corresponding to different atlas regions, are represented by different colors. The orange segment, for example, corresponds to how Modality 1 voxels from the 1\textsuperscript{st} atlas region correlate to Modality 2 voxels from the 2\textsuperscript{nd} atlas region. In this method we are reducing the $N \times N$ correlations of $R_{12}$ to $116 \times 116$ inter-regional correlation z-scores.

### 3.2.4 Number of significant correlations in $R_{12}$

In sections 3.2.1 to 3.2.3 the values proposed were computed by obtaining averages in various different ways. If a Modality 1 voxel had significant negative and positive correlations with Modality 2 voxels the process of averaging can nullify the significant correlations. In this section we explain a method that will preserve only the significant correlations. For each voxel in Modality 1, the number of voxels with significant correlations (positive and negative separately) with all voxels of Modality 2 is found and stored at the Modality 1 voxel location. Using this information a brain map of Modality 1 voxels with significant Modality 2 correlations can be constructed. Significance of the correlations can then be calculated in the following manner. The ‘\texttt{tinv}’ function in the statistical toolbox of MATLAB (http://www.mathworks.com/) calculates the $t$–value ($t_i$) corresponding to a specific significant $p$–value ($p_i$). The $p$–value is a measure of an event occurring by chance. For example a $p$–value of 0.01 for a certain result indicates that the probability of that result occurring due to random chance is 1%. The significance of the result is higher for lower $p$–values. The correlation value ($r_i$) corresponding to $t_i$ was computed using Equation (3.2), where $n$ is the number of subjects.
3.3 Multiple Linear Regression (MLR)

The purpose of multiple linear regression (MLR) (Neter, et al. 1996) is to determine the relationship between multiple independent variables and a single dependent variable. The discussion here follows Neter, et al. 1996. It allows researchers to investigate the best predictor for an observation or event. Let $Y[1\times n]$ be the dependent observation, obtained from $n$ number of subjects, that is to be predicted with $p$ number of independent predictors given by the row vectors of $X[(p+1)\times n]$. The first row of $X$ is a vector of ones and helps to compute the bias. The equation for MLR is given by Equation (3.3), where $\beta[1\times (p+1)]$ contains the coefficients of the predictors and $\varepsilon[1\times n]$ is the vector of residual errors.

$$Y = \beta X + \varepsilon$$  \hspace{1cm} (3.3)

The coefficient of determination or R–square ($r^2$) is defined as one minus the ratio of the residual variability of the $Y$ variable to the original variance. If $r^2$ is equal to one it means that the predictor variables in $X$ account for all variation in observation $Y$ and if $r^2$ is equal to zero it means that no linear relationship exists between $X$ and $Y$ in the sample data.
3.4 Canonical Correlation Analysis (CCA)

Canonical correlation analysis (CCA) (Hotelling 1936) is an approach to find linear relationships between two multidimensional variables. The CCA discussion and derivation presented here is obtained from Hardoon, et al. (2004). It can be seen as the problem of finding basis vectors to project two sets of variables to mutually maximize the correlation of the projections (Hardoon, et al. 2004). Correlation between two sets of multi dimensional variables depends on the coordinate system on which they are described. A highly correlated pair of multidimensional variables in a particular coordinate system may be weakly correlated in another. CCA seeks to find a pair of linear transformations, one for each of the sets of variables, so that after the transformation the correlation of the two variables are maximized. Let $X$ and $Y$ be the two sets of multi dimensional variables. The aim of CCA is to find basis vectors, $w_x$ and $w_y$ so that the correlation $r$, given by Equation (3.4) is maximized.

$$r = \max_{w_x, w_y} \frac{\langle xw_x, yw_y \rangle}{\|xw_x\|\|yw_y\|}$$  \hspace{1cm} (3.4)

It can be shown that Equation (3.4) can be written as (Hardoon, et al. 2004),

$$r = \max_{w_x, w_y} \frac{w_x^t C_{xy} w_y}{\sqrt{w_x^t C_{xx} w_x w_y^t C_{yy} w_y}}$$  \hspace{1cm} (3.5)

where $C_{xx}$ and $C_{yy}$ are within–set–covariance matrices and $C_{xy}$ is the between–sets covariance matrix.
Maximizing $r$ can be reformulated as maximizing the numerator of Equation (3.5) with the constraints of,

$$w'_x C_{xx} w_x = 1$$
$$w'_y C_{yy} w_y = 1$$

The corresponding optimization function can be written as,

$$L(\lambda, w_x, w_y) = w'_x C_{xy} w_x - \frac{\lambda_x}{2} (w'_x C_{xx} w_x - 1) - \frac{\lambda_y}{2} (w'_y C_{yy} w_y - 1)$$  \hspace{1cm} (3.6)$$

In Equation (3.6) $\lambda$ stands for the Lagrange multiplier. Taking derivatives of Equation (3.6) with respect to $w_x$ and $w_y$ and equating them to zero results in solutions for $w_x$ and $w_y$ as given by Equations (3.7) and (3.8) respectively.

$$w_y = \frac{C^{-1}_{yy} C_{yx} w_x}{\lambda}$$  \hspace{1cm} (3.7)$$

$$C_{yy} C^{-1}_{yx} C_{yx} w_x = \lambda^2 C_{xx} w_x$$  \hspace{1cm} (3.8)$$

Equation (3.8) is a generalized eigen problem of the form $A x = \lambda B x$ and by solving for the generalized eigenvectors a sequence of $w_x$‘s can be obtained. The corresponding $w_y$ ‘s can be found by substituting $w_x$ ‘s in Equation (3.7).

In our application of CCA, the variables $X[N_1 \times n]$ and $Y[N_2 \times n]$ are obtained from $n$ number of subjects with each having $N_1$ and $N_2$ number of features respectively. If we assume that $N_1 > N_2$, CCA results in $N_2$ sets of basis vectors corresponding to $N_2$ numbers of maximized correlations in the transformed space. As a result of applying
CCA to \( \mathbf{X} \) and \( \mathbf{Y} \), we obtain vectors \( \mathbf{w}_x \) and \( \mathbf{w}_y \). From the weights of \( \mathbf{w}_x \) and \( \mathbf{w}_y \) it is possible to identify the features of \( \mathbf{X} \) and \( \mathbf{Y} \) that contribute towards the correlation in the transformed space. In this manner it is possible to explore how different features from different modalities or tasks are correlated.

To illustrate the application of CCA we use the following example. Let us assume that vectors \( \mathbf{X} \) and \( \mathbf{Y} \) have four and three features respectively with each of them having \( n = 1000 \) sample points. We construct samples of \( \mathbf{X} \) and \( \mathbf{Y} \) such that three features in \( \mathbf{X} \) and one feature in \( \mathbf{Y} \) are random variables and one feature in \( \mathbf{X} \) and two features in \( \mathbf{Y} \) are correlated. The correlated features are constructed as sinusoidals with different delays. The vectors \( \mathbf{X} \) and \( \mathbf{Y} \) are illustrated in Figure 3–7. The application of CCA onto these two data sets results in \( \mathbf{w}_x = [-1, 0, 0, 0] \) and \( \mathbf{w}_y = [0.5, -0.8, 0] \) with a maximum CCA correlation \( r \) as 1.0. In the example the random variables do not contribute towards the correlation in the transformed space and hence scalars 2, 3 and 4 in \( \mathbf{w}_x \) and the third scalar in \( \mathbf{w}_y \) result as zeros. The result indicates that the first feature in \( \mathbf{w}_x \) and the first and second features in \( \mathbf{w}_y \) contribute towards the correlation and that the other features do not.
Figure 3–7: Two sets of multivariate vectors to illustrate CCA correlations
Variable X and Y have four and three features. One feature in x and two in y have sinusoids as features and the rest are random variables.

We repeated this experiment with all features in X and Y as random variables with \( n = 1000 \) and we obtained \( r \) as 0.07. When \( n \) was reduced to 45, even with X and Y as random variables, CCA was able to transform the vectors to obtain a maximum correlation \( r = 0.45 \). From this experiment we realized that when the number of sample points gets closer to the number of features, the CCA algorithm was able to find high correlation values even with random features. Related to this issue one of the drawbacks of CCA and one that posed a serious obstacle to use CCA to extract features is the following. When the number of features \((N_1 \text{ or } N_2)\) exceeds the number of samples \((n)\) CCA gives a set of correlations \((r)\) all of which have values of 1, making it difficult to select the optimal set of features. We had hoped that by examining the weights corresponding to the maximum CCA correlation it would be possible to select optimal features that contribute towards the correlation. But when all CCA correlations result in
values equal to one, and their corresponding weights ($w_x$ and $w_y$) are not identical, the selection of optimal features is not possible. This result implies that multiple sets of basis vectors can be used to transform $X$ and $Y$ to obtain a correlation of one. To overcome this issue we modified the application of CCA in the following manner. Instead of applying CCA to all the features of $X$, we apply CCA to each of the features separately. With this method, instead of using weights with high values to select features, we use the CCA correlations to select features. Further details on the application of the CCA method will be provided in Section 6.4

The significance of CCA correlation can be calculated with Roy’s largest root statistic (Rencher 1998). Let $\theta$ be the test statistic, given by the square of the CCA correlation ($r$). We define,

\[ s = \min(N_1, N_2) \]  
\[ m_1 = \frac{(|p-q|-1)}{2} \]  
\[ m_2 = \frac{(n-N_1-N_2-2)}{2} \]

Where $n$ is the number of subjects, $N_1$ and $N_2$ the number of features in $X$ and $Y$. In Table B5 of (Rencher 1998) $\theta$ for $p = 0.05$ and given $m_1, m_2$ and $N$ can be found.
3.5 Tract based spatial statistics (TBSS)

We applied tract based spatial statistics (TBSS) (Smith, et al. 2006) to DTI data to obtain white matter skeletons. The TBSS discussion and derivation presented here follows Smith, et al. (2006). To register sMRI data onto a standard space, voxel based morphometry (VBM) is widely used where alignment, segmentation and smoothing are incorporated into a unified model (Ashburner and Friston 2000; Good, et al. 2001). SMRI data cover brain regions but DTI data consist of thin fiber tracts and are more sensitive to misalignment. Misalignments can be misinterpreted as difference between the groups in actual brain data. TBSS reduces misalignment by computing a mean skeleton from the center of the group mean fractional anisotropy (FA) map and then allows for subject variability around the mean skeleton. The TBSS algorithm is explained in the following steps.

**Step 1:** Use nonlinear registration with medium degrees of freedom (dof) to align all subjects’ FA maps. Nonlinear alignment algorithms with high dof can break brain topology. For example structurally different brains can be enforced to perfectly align, but at a specific brain region/voxel in the aligned brain, data could have come from different anatomical regions for different subjects.

**Step 2:** Compute the mean FA map of all subjects (see Figure 3–8a)

**Step 3:** Find the mean skeleton of the mean FA map. At this step, the central locations of all the tracts are computed. Most skeletal regions will have a two dimensional surface structure (see Figure 3–8b) and some a tubular structure.
**Step 4:** For each voxel location on the mean skeleton, search for the nearest maximum FA value in subject data, that lies perpendicular to the mean skeleton tract (see Figure 3–8c). The voxel value found is assigned to the subject skeleton. This process is performed on each subject separately to compute individual subject skeletons. With this method the spatial locations of the subject skeleton are kept identical to the mean skeleton but the values on the subject skeleton are found from the subject’s FA map. Since skeletons are identical across subjects, it is possible for researchers to compute statistics for a group of subjects.

![Figure 3–8: Tract Based Spatial Statistics (TBSS) Steps](image)

**Figure 3–8: Tract Based Spatial Statistics (TBSS) Steps**
(a) Center of the mean FA maps (mean skeleton) shown in red. (b) A 3–dimensional view of the mean skeleton. (c) Computing subject skeleton, values based on subject FA map and location based on mean skeleton. (d) Subject skeletons for six different subjects.

---

Reproduced from Smith, et al. (2006) with permission
3.6 Histogram Shift

In section 3.1 we introduced a method to analyze the spatial correlation between Modality 1 and Modality 2 voxels. Computing the histogram (section 3.2.1) of such correlations was one of the reduction techniques among other methods we introduced. Let us assume that data from $n_{HC}$ number of controls and $n_{SZ}$ number of patients from different modalities are available and that each modality has $N$ number of voxels. By cross correlating the voxels from different modalities, the spatial correlation matrix $R_{mn}^{SZ} [N \times N]$, $m \neq n$, is computed using data from SZ. We do not compute $R_{mn}^{SZ}$ for $m = n$ since we are not investigating the correlations between the same fMRI task. Similarly $R_{mn}^{HC} [N \times N]$ is computed using data from HC. Let $\mu_{mn}^{SZ}$ and $\mu_{mn}^{HC}$ be the mean value of the elements of matrices $R_{mn}^{SZ}$ and $R_{mn}^{HC}$ respectively. It was identified in Michael, et al. (2008c) that $\mu_{mn}^{SZ}$ and $\mu_{mn}^{HC}$ were different and this differential feature between the two groups is exploited to develop a classification algorithm.

First $\mu_{mn}^{SZ}$ and $\mu_{mn}^{HC}$ are found with one subject left–out. Then the left–out subject is included in each of the two groups to find new mean values. The expectation is that when the left–out subject is included in its own group the new mean would shift away from the mean of the other group and when it is included in the incorrect group the new mean will shift towards the mean of its own group. For better clarity, in Figure 3–9 we have shown the histogram of correlations (elements of $R_{mn}$) when (a) an HC or (b) an SZ is left out for a particular combination of two modalities. In Figure 3–9(a) the solid blue line represents the histogram for HC when a single HC is left–out; the solid red for SZ; the dashed blue when the left–out HC is included in HC; and the dashed red when the
left–out control is included in SZ. Note that the dashed line histograms are shifting in the
direction of the HC’s histogram (a right shift). In Figure 3–9(b) the same is presented
when a SZ is left out. In this case the shift in the histograms is towards that of the SZ’s
histogram (a left shift).

It is possible to make histograms similar to those shown in Figure 3–9 for
different combinations of modalities. If the shifts in both histograms are in the same
direction for all the subjects of a particular group, and in the opposite direction for all
subjects in the other group (as seen in Figure 3–9), for any one combination of modalities
then classification can be based on that combination alone. In real data there was no
single combination of modalities for which all subjects showed this trend. An algorithm
for better classification can be made by combining the histogram shifts from different
combinations of modalities and its steps are listed below.

**Step 1:** Pick a subject (S\textsubscript{i}) and leave it out from the analysis.

**Step 2:** Find $\mu^{LHC}_{mn}$, the mean correlation of HC without S\textsubscript{i}, using data from the $m^{th}$
and $n^{th}$ modalities. Calculate the mean correlation from the $N^2$ number of
correlations.

**Step 3:** Find $\mu^{LSZ}_{mn}$, the mean correlation of SZ without S\textsubscript{i}.

**Step 4:** Include S\textsubscript{i} in the HC group and find $\mu^{HHC}_{mn}$, the new mean correlation of
HC.
**Step 5:** Include $S_i$ in the SZ group and find $\mu_{mn}^{\text{SZ}}$, the new mean correlation of SZ.

**Step 6:** If $\mu_{mn}^{\text{HC}} > \mu_{mn}^{\text{LHC}}$ (right shift), $A_{mn}^{\text{HC}} = 1$ (HC) else (left shift) $A_{mn}^{\text{HC}} = -1$ (SZ). The assignment of $\pm 1$ is based on the nature of histogram shifts as shown in Figure 3–9. The direction of change in the mean value gives a cue to the left–out subject.

**Step 7:** If $\mu_{mn}^{\text{ISZ}} > \mu_{mn}^{\text{LSZ}}$ (right shift), $A_{mn}^{\text{SZ}} = 1$ (HC) else (left shift) $A_{mn}^{\text{SZ}} = -1$ (SZ).

**Step 8:** Repeat steps 2 – 7 for all combinations of $m$ and $n$.

**Step 9:** Find $F = \sum_{m=1}^{M} \sum_{n=m+1}^{M} \left(A_{mn}^{\text{HC}} + A_{mn}^{\text{SZ}}\right)$, where $M$ is the total number of modalities.

The cues are added together to make the final classification decision.

**Step 10:** If $F > 0$, $S_i$ is classified as a control, if $F < 0$, $S_i$ is classified as a patient, and if $F = 0$ no decision can be made.
Figure 3–9: Histogram Shift

Histograms between two modalities are computed with one subject (HC or SZ) left out. These histograms are shown with solid lines in red for SZ and blue for HC. If the left out subject is a HC, the histograms (represented by dashed lines) shift to the right as shown in (a) and if SZ the histograms (represented by dashed lines) shift to the left as shown in (b).

---

Reproduced from Michael, et al. 2008a
3.7 Support Vector Machines (SVM)

In this section we provide a brief introduction to support vector machines (SVM) and mathematically develop equations for its application (Hastie, et al. 2001). The discussion and derivation presented here are from Hastie, et al. (2001). SVM attempts to find a hyperplane in a higher dimensional space to separate two classes of data that are overlapping in their own space. We shall refer to this higher dimensional space as feature space. To better understand the concepts let us assume that training data consists of \( N \) points in a two dimensional space and the two classes are mostly non–overlapping and linearly separable. Let these points be \( x_1, x_2, \cdots, x_N \) and belong to either of the two classes \( y_i \in \{-1, +1\} \) as shown in Figure 3–10. We define a hyperplane given by Equation (3.12) as a linear plane that separates the two classes.

\[
x^T \beta + \beta_0 = 0
\]  

(3.12)

Where \( \|\beta\| = 1 \).

An infinite number of discriminatory planes (many possible solutions for \( \beta \) and \( \beta_0 \) ) can be constructed to classify the two classes as shown in Figure 3–10(a). To find the optimal hyperplane, SVM uses data points that are closest to the separating boundary of the two classes since they are the most difficult to classify. These critical data points determine the location of the support hyperplanes or the margin for each class as shown in Figure 3–10(b). The optimal hyperplane given by Equation (3.12) is parallel to the margins and located at a distance \( C \) from them. SVM seeks to find the hyperplane that maximizes \( C \).
Figure 3–10: Support Vector Machine Classifier
(a) Two dimensional data \( (x_i) \) from two classes \([+1, -1]\), represented by blue squares and red circles respectively, can be separated by an infinite number of hyperplanes. \( x_i^m \) are the misclassified data points. (b) SVM computes the hyperplane \( x^T \beta + \beta_0 = 0 \) which is at the midpoint between the two supporting vectors (margins) that maximally separate the two groups at a distance of \( 2C \).

If the two classes are completely separable it is possible to find the values for \( \beta \) and \( \beta_0 \) corresponding to the optimal hyperplane by solving for the optimization problem given by Equation (3.13).

\[
\max_{\beta, \beta_0} \frac{C}{\|\beta\|^2} \\
\text{subject to } y_i \left( x_i^T \beta + \beta_0 \right) \geq C, i = 1, \ldots, N
\] (3.13)

It can be show that \( C = 1/\|\beta\| \) and Equation (3.13) can be written as,

\[
\min_{\beta, \beta_0} \|\beta\| \\
\text{subject to } y_i \left( x_i^T \beta + \beta_0 \right) \geq C, i = 1, \ldots, N
\] (3.14)
Equation (3.14) is a convex optimization problem (quadratic criterion, linear inequality constraints). Well established algorithms are available to solve for the above equation (Ben-Tal 2001; Press, et al.).

The above derivation was for the case when no data points were on the incorrect side of the hyperplane. If it is not possible to perfectly separate the two classes with a hyperplane, SVM allows some points to be on the incorrect side of the two classes. For this case the same optimization problem given by Equation (3.14) is used with a penalty variable $\xi$ and is given by Equation (3.15).

\[
\begin{align*}
\min_{\beta, \beta_0} & \|\beta\| \\
\text{subject to} & \quad y_i \left( x_i^T \beta + \beta_0 \right) \geq 1 - \xi_i, \forall i \\
& \quad \xi_i \geq 0, \sum_{i=1}^{R} \xi_i \leq \text{constant}
\end{align*}
\]  

(3.15)

$\xi_i$ is the distance of the misclassified data point from its margin and is calculated for all training data points. By bounding the sum $\sum_{i=1}^{R} \xi_i$ we bound the total distance of the misclassified data points, where $R$ is the total number of misclassified training data points. From the constraints set by Equation (3.15) it is seen that the points well inside the class margins do not play a big role in determining the location of the margins.

In the above example the two classes were mostly separable in their own space. If the classes cannot be separable with a linear hyperplane, the above formulation of SVM will fail to classify data points efficiently. Linearly non–separable data sets can be linearly separable in a higher dimensional space. SVM uses kernel functions such as polynomial functions, radial basis function, sigmoid function etc. to transform the data.
points onto feature space where the two classes can be linearly separated. For example if the data points are in a two dimensional space \((x_1, x_2)\), a feature space can be found by mapping \((x_1, x_2)\) onto \((x_1^2, x_2^2, x_1, x_2)\). Efficient kernel methods (Schölkopf and Smola 2002) are used to transform data sets onto a higher dimensional space.

### 3.8 Discriminant Analysis (DA)

In the above explained SVM method, data points closest to the class margins influence heavily in determining the location of the class margin. In discriminant analysis (DA) (Duda, et al. 2001) techniques all data points of the classes help to determine the class margins. The discussion and derivation presented in this work follows Duda, et al. (2001). In DA the decision boundary is determined by the covariance of the class distribution and the locations of the class centroids. To understand the concepts of DA let us assume that our classification data set has two classes. Let the a priori probability of class 1 and 2 be \(P(\omega_1)\) and \(P(\omega_2)\) respectively. The class conditional probability density functions for these two classes are shown in Figure 3–11 and let us assume them to be \(p(x | \omega_i), i = 1, 2\) where \(x\) is a continuous random variable and it reflects the distribution of a class feature.
If \( P(\omega_1), P(\omega_2), p(x | \omega_1) \) and \( p(x | \omega_2) \) are known, it is possible to predict the class of a new data point using Bayes formula represented by Equation (3.16).

\[
P(\omega_i | x) = \frac{p(x | \omega_i)P(\omega_i)}{\sum_{j=1}^{2} p(x | \omega_j)P(\omega_j)}
\]

(3.16)

\( P(\omega_i | x) \) in Equation (3.16) is known as the a posteriori probability and gives the probability for the new data to belong to class \( \omega_i \). The denominator of Equation (3.16) can be eliminated since it is a scale factor and assures that \( P(\omega_1 | x) + P(\omega_2 | x) = 1 \). With this simplification, taking the natural logarithm and changing the feature space \( (x) \) to be
in a multidimensional space Equation (3.16) can be written as Equation (3.17), where \( g_i(x) \) is the value of \( P(\omega_i | x) \) after the above simplification steps.

\[
g_i(x) = \ln p(x | \omega_i) + \ln P(\omega_i)
\]  

(3.17)

To compute \( g(x) \), the conditional densities \( p(x | \omega_i) \) and prior probabilities \( P(\omega_i) \) for each of the classes are to be known. Out of the various density functions the normal or the Gaussian function is the density function of a distribution with a large number of sample points and is assumed to be the density function in many other applications. In a \( d \) dimensional feature space a multivariate normal density function with class mean \( \mu_i \) \((d \times 1)\) and class covariance \( \Sigma_i \) \((d \times d)\) can be given by Equation (3.18).

\[
p(x | \omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp \left[ -\frac{1}{2}(x - \mu_i)^t \Sigma_i^{-1} (x - \mu_i) \right]
\]  

(3.18)

Substituting Equation (3.18) in (3.17) results in Equation (3.19).

\[
g_i(x) = -\frac{1}{2}(x - \mu_i)^t \Sigma_i^{-1} (x - \mu_i) - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i)
\]  

(3.19)

Several cases arise depending on the nature of the covariance matrix for the different classes.
3.8.1 Case 1: $\Sigma_i = \sigma^2 I$

The simplest case occurs when the features are statistically independent and each feature from all classes having the same variance, $\sigma^2$. The covariance matrix then reduces to $\sigma^2 I$, where $I$ is the identity matrix. Equation (3.19) then reduces to Equation (3.20), which is a hyperplane.

$$g_i(x) = \frac{1}{\sigma^2} \mu'_i x - \frac{1}{2\sigma^2} \mu'_i \mu_i + \ln P(\omega_i) \quad (3.20)$$

3.8.2 Case 2: $\Sigma_i = \Sigma$

In this case the covariance matrices of the classes are assumed to be identical but are not identity matrices. For this case Equation (3.19) reduces to Equation (3.21) which again is a hyperplane.

$$g_i(x) = \mu'_i \Sigma^{-1} x - \frac{1}{2} \mu'_i \Sigma^{-1} \mu_i + \ln P(\omega_i) \quad (3.21)$$

3.8.3 Case 3: $\Sigma_i = \sigma_i^2 I$

Here the features are statistically independent but each class has a different variance, $\sigma_i^2$. Equation (3.19) reduces to Equation (3.22) which is a hyperquadric surface.

$$g_i(x) = -\frac{\|x - \mu_i\|^2}{2\sigma_i^2} + \ln P(\omega_i) \quad (3.22)$$
3.8.4 Case 4: $\Sigma_i = \text{arbitrary}$

This is the most general multivariate normal case, the covariance matrices are different for each class. Equation (3.19) reduces to Equation (3.23) and this is a hyperquadric surface.

$$g_i(x) = -\frac{1}{2}(x - \mu_i)' \Sigma_i^{-1}(x - \mu_i) - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i)$$

(3.23)

3.9 Classification based on DTI – PANSS Data

In this section we introduce an algorithm to classify patients from controls using features derived from DTI–PANSS correlations. In Section 3.4 we introduced canonical correlation analysis (CCA) to combine two sets of multivariables. We use TBSS (see Section 3.5) to reduce DTI data to skeletons and then apply CCA to find voxels with high DTI–PANSS correlations. The TBSS and the CCA steps are applied as data reduction steps to select appropriate features. The reduced features (data from reduced voxels) are used to classify subjects using support vector machines (SVM, see Section 3.7) or discriminant analysis (DA, see Section 3.8). The classification is based on the leave–one–out scheme and the steps are listed below. By this method we are trying to investigate if image values, measured through DTI, that have high correlation with symptoms of schizophrenia, measured through PANSS, can be used as features to develop a classifier.

**Step 1:** Find white matter TBSS skeletons for all subjects using the mean FA map of all the subjects as a spatial template.
Step 2: Leave out a HC ($S_i^{HC}$) and a SZ ($S_i^{SZ}$) from the analysis.

**Step 2:** Find the CCA correlations for each voxel on the white matter skeleton using the four DTI values (AD, FA, MD and RD, see Section 2.4) and the three PANSS scores (+ve, -ve and general, see Section 1.3). The CCA correlations are computed using data from SZ. The PANSS test is not administered on HC and hence PANSS data from HC are not available. Note that the left–out patient ($S_i^{SZ}$) is not used to find voxels with high DTI–PANSS correlations.

**Step 3:** Threshold voxels with high DTI–PANSS correlations. Use the four DTI values from these voxels to form the training data set for the classifier.

**Step 4:** The training data set is fed into SVM or DA as features. For Case 2 (see Section 3.8.2) and Case 4 (see Section 3.8.4) of DA the inverse of the covariance matrix of the feature data set is required. The number of features is very large ($>>$ 1000) and computing the covariance matrix for a large number of features is not feasible. We reduce the feature matrix using singular value decomposition (Press, et al. 1992) and then compute the covariance matrix.

**Step 5:** The left–out subjects ($S_i^{HC}$ and $S_i^{SZ}$) are tested with the classifiers.
4 Structural – Functional Analysis

The great architect Frank Lloyd Wright built his career on the belief that “form and function should be one”. Wright implies a close relationship between structure and activity. In this chapter we investigate the association between structural and functional brain data. Our hypothesis is that brain pathology may be associated with a disassociation between structure and function. Most of the discussion and results presented here are from Michael, et al. (2009b), Michael, et al. (2009c) and Michael, et al. in review.

In most functional magnetic resonance imaging (fMRI) studies structural MRI (sMRI) is also acquired. These two modalities provide unique information about the brain; however their analysis is typically performed separately. The sMRI images are often used to co–register different subjects’ brains onto a common template (Toga and Thompson 2001) or used as a rendering surface to visualize overlaid functional activation (Corbetta, et al. 1998). Studies separately analyzing structural and functional images have found that multiple brain regions appear to be affected in schizophrenia (Goldstein, et al. 1999; Honea, et al. 2005; Niznikiewicz, et al. 2003). Structural findings suggest enlarged ventricles, reduced volume of temporal lobe and superior temporal gyrus, with moderate
evidence for frontal lobe volume reduction, and some evidence for persistent cavum septi pellucidi and abnormalities in basal ganglia, corpus callosum, thalamus and cerebellum. Functional results suggest abnormal connectivity between temporal and frontal brain regions and aberrant activation in the dorsolateral prefrontal cortex (DLPFC) (Manoach, et al. 2000; Weinberger, et al. 1986). An overwhelming number of studies report that for a given task, abnormal function is not localized to a specific brain region, but instead the abnormalities involve a network of brain regions.

The brain is a vastly interconnected organ and it is reasonable to expect that changes in local morphological structures may result in modulations of brain activity in distant regions (Mesulam 1998). However, the correlation or the connection between structural regions and functional activation has not been well established. Studies have been designed to examine localized correlations between gray matter volumes with fMRI (Siegle, et al. 2003), to relate gray matter homogeneity with fMRI (Mitchell, et al. 1988), or to correlate structural and functional data from certain regions in the brain (Hasnain, et al. 2001; Wen, et al. 2004). Existing tools for examining joint information include region–based approaches such as structural equation modeling (McIntosh and Gonzalez-Lima 1994) and dynamic causal modeling (Friston, et al. 2003). These studies and techniques are important; however, they are limited by lack of detailed examination of the relationship between distributed brain voxels and assume a priori knowledge of the implicated brain regions. Independent component analysis (ICA) is another tool to fuse multi–modal information to find independent brain networks (Calhoun, et al. 2006a; Calhoun and Adali In Press). The problem wherein anatomical differences or registration error contribute to differential functional activation was addressed by integrating tissue
mismatch as a nuisance variable into the general linear model (Oakes, et al. 2007). This method addresses the issue of functional activation differences due to underlying structure. The analysis investigates the influence of structure on function only within the same voxel, and does not examine relationships between structure and function at distant brain regions. There are other methods (Fan, et al. 2007; Ford, et al. 2002) that use structural and functional data for classification purposes, however these do not examine the inter–relationship between structurally and functionally derived data.

A complementary approach to the above methods is to assess cross–correlations among all sMRI and fMRI voxels. This task is straightforward for a small number of voxels, but becomes challenging for whole brain studies due to the magnitude of necessary computations and temporary storage needed. In this chapter we use techniques developed in section 3.1 to identify new features of schizophrenia through the fusion of sMRI and fMRI. sMRI data were derived from $T_1$–weighted scans and voxel–based morphometry (Ashburner and Friston 2000) was used to segment the brain to obtain gray matter (GM) concentration map. GM concentration is the percentage of GM content at a voxel location. We selected GM since GM brain regions are the information processing center of the brain. Our interest for this study was in investigating the relationship between GM and functional activity and not in the physical connection between brain regions that are measured with white matter concentration or diffusion tensor imaging (DTI). The fMRI data were acquired while subjects performed an auditory sensorimotor (SM) task (see section 2.3.2.3) and were reduced to activation maps by regressing the time course of the task against the data (see section 2.3.4).
Schizophrenia is likely associated with disruptions in the connectivity between different cortical regions (Andreasen, et al. 1999; Andreasen, et al. 1998; Breakspear, et al. 2003; Friston 1998; Kubicki, et al. 2007; Lim, et al. 1999; Ross and Pearlson 1996). Studies that investigate the linkage between structure and function while taking into account all brain voxels are needed to investigate the differences between healthy and patient groups. The term ‘linkage’ is used to denote the relationship, association or correlation between two variables. A joint analysis of this nature enables one to do a comprehensive study of the interactions between sMRI and fMRI. It is reasonable to expect that large functional networks are associated with features of gray matter to different degrees in HC and SZ. The disconnection hypothesis of schizophrenia (Friston 1998) states that neural mechanisms in schizophrenia are not localized to any one area alone, but rather that it is the integrity of the interconnections between brain regions that is compromised. Thus, we predict that the relationship between structure and function will be weaker in patients with schizophrenia than in healthy controls.

4.1 Subjects

Subjects used in this analysis were scanned at four different sites: the University of Iowa Hospital (IA), Harvard’s Massachusetts General Hospital (MA), the University of Minnesota (MN) and the Mind Research Network (NM) and are part of a larger study called the Mind Clinical Imaging Consortium (MCIC). Patients were recruited from inpatient and outpatient psychiatric clinics, group homes, referrals from physicians, advertisements, and by word–of–mouth. Controls were recruited from advertisements, fliers, and word–of–mouth. All participants provided written, informed, IRB approved
consent at their respective sites and were compensated for their participation. Subjects analyzed in this study had normal hearing (assessed by self report) and were able to carry out the SM task. There were a total of 94 healthy controls (HC) and 85 patients with chronic schizophrenia (SZ) and the site break down of subjects is as follows: IA 29 HC and 19 SZ, MA 16 HC and 16 SZ, MN 21 HC and 23 SZ and NM 28 HC and 27 SZ. Outliers were determined based on sMRI and fMRI data. sMRI outliers were checked based on spatial correlations with the average map (see section 2.2.2) and fMRI outliers were detected using a new method developed in this study (see section 2.6). No outliers were identified based on sMRI data but 12 HC and 15 SZ were removed based on fMRI data. Out of the 82 cleaned HC data, 70 HC were picked to match (as much as possible) the cleaned 70 SZ for age and sex. The number of subjects in each group is matched to keep the significance of structural–functional correlation constant between the two groups. The demographics of 70 HC and 70 SZ are presented in Table 4–1 and results presented in this chapter predominantly pertain to these subjects. The socioeconomic status (SES) in Table 4–1 is a measure of a family’s economic and social position calculated using the family’s income, education, occupation, wealth etc. (Hollingshead and Redlich 1958). In Table 4–1 all demographics except years of education have insignificant group differences. A large number of subjects have to be removed to exactly match all subject demographics between the two groups. To avoid this loss we select another subset of subjects from the cleaned 82 HC and 70 SZ to match education years to show the robustness of our result.

Healthy subjects in the study were screened to ensure they were free from Diagnostic and Statistical Manual of Mental Disorders (DSM–IV) axis I or axis II
psychopathology, assessed using a modified version of the comprehension assessment of symptoms and history (CASH) (Andreasen, et al. 1992a). They were interviewed to determine that there was no history of psychosis in any first-degree relatives. Patients met criteria for schizophrenia based on structured clinical interview for DSM (SCID) or CASH and confirmed by review of the case file. Average patient symptom measures, positive (Andreasen 1984), negative (Andreasen 1981), and disorganization are reported in Table 4–1.

<table>
<thead>
<tr>
<th>Table 4–1: Demographics of SZ and HC with symptom scores for SZ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SZ</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td><strong>Male, Female</strong></td>
</tr>
<tr>
<td><strong>Handedness (non–right hand)</strong></td>
</tr>
<tr>
<td><strong>Education</strong></td>
</tr>
<tr>
<td><strong>Parental Socioeconomic Status</strong></td>
</tr>
<tr>
<td><strong>Years Since Diagnosis</strong></td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
</tr>
</tbody>
</table>

Note: SZ = patients with schizophrenia; HC = healthy controls; $n_p =$ number of patients; $n_c =$ number of controls; NA = not applicable
4.2 Structural – Functional Spatial Correlation Histogram

The method introduced in Section 3.2.1 was applied to the group of HC and SZ separately. Modality 1 and 2 used in Section 3.2.1 to formulate the problem is replaced by structural and functional data respectively, to compute the structural – functional correlation histogram ($R_{\text{SF}}$) with all brain voxels. In Figure 4–1, the y–axis represents the number of occurrence of different correlation values given along the x–axis. The summation of the number of correlation occurrences in the histogram equals the number of elements of $R_{\text{SF}}$ ($N \times N$). A mean correlation value of zero in the histogram indicates that on average the structural gray matter concentration (across subjects) and functional activation (across subjects) are uncorrelated. A positive correlation value means that an increase or decrease in gray matter concentration corresponds to an increase or decrease in functional activation respectively. A negative correlation value means that a change in gray matter concentration has a change in functional activation in the opposite direction.
Figure 4–1: Structural–Functional Correlation Histogram with all Brain Voxels

(a) Histograms for all possible combinations of correlations between structural (gray matter concentrations) and functional (activation maps for an auditory sensorimotor task) voxels of the whole brain for seventy patients with schizophrenia (SZ) in red and seventy healthy controls (HC) in blue. The y–axis represents the number of occurrences of correlation values given along the x–axis. In (b) the difference (HC – SZ) histogram is shown and it is evident from this plot that SZ have a higher number of correlations close to zero and HC have a higher number of correlations close to +0.2 and –0.2.

Reproduced from Michael, et al. in review
In Figure 4–1(a) the SZ histogram is shown in red and the HC histogram is shown in blue. The difference between SZ and HC histograms is not very clear when they are drawn on the same plot. To enhance the difference in Figure 4–1(b) the difference histogram (HC – SZ) is illustrated. From this plot it is evident that the number of zero correlations is higher in SZ than in HC and the number of correlations close to +0.2 and –0.2 are higher in HC than in SZ. To find the significance of the correlation between functional activation and gray matter concentration we converted the correlation value (±0.2) that showed group difference into a t value using Equation (3.2), where \( r = \pm 0.2, n = 70 \) (number of subjects). We then converted the t value to its corresponding p value. With this method for an \( r = \pm 0.2 \) we obtained a p value of 0.048. This value does not indicate the significance of the difference between the two groups, but that the significance of correlation between structural and functional data within a group is \( p < 0.05 \).

The result presented above was based on a cohort of subjects with insignificant group differences in age and parental socioeconomic status and significant group difference in years of education. To match educations we selected 67 subjects from each group from cleaned 82 HC and 70 SZ. The mean, standard deviation and the significance of difference for years of education for the two groups are as follows: HC = 14.6 ± 1.5, SZ = 14.1 ± 2.4, t–value = 1.4 and p–value = 0.15. The structural – functional correlation histograms (as in Figure 4–1) were computed and the result that HC having higher number of correlations than SZ was consistent with this subject group as well.
4.3 Correlation Histograms of Significant Voxels

In the previous analyses structural–functional correlations were found using all brain voxels. We were interested in checking the above result for functional voxels with significant activation and structural voxels with high gray matter concentration. Our test was to find if the pattern of HC having higher correlations than SZ changed across different levels of significant voxels. Functional voxels that had the highest group (HC+SZ) mean absolute $t$-values for the SM task and structural voxels with the highest group mean gray matter concentration were used for this analysis. The voxels that pass a certain threshold in the group mean map were chosen from each subject to compute the histogram. In Figure 4–2(a) the difference histograms (HC–SZ) are shown as a colormap. Shades of red and yellow correspond to HC having higher number of correlations than SZ and shades of blue to SZ having higher number of correlation than HC at correlation values given by the x-axis. We compute the histograms for different combinations of structural and functional thresholds and these are shown in the rows of Figure 4–2 (a). In Figure 4–2 (a) SM $t$–value is changed (between 0 and 4.0 in steps of 0.25, separated by dashed lines) in the outer loop and GM concentration is changed (between 0 and 0.8 at steps of 0.1) in the inner loop. In Figure 4–2(a), the first row corresponds to the histogram computed with all structural and functional voxels and the last row to the histogram computed with structural voxels with GM concentration above 0.8 and functional voxels with $t$–values above 4.0

At lower SM $t$–value thresholds (0.0 to 1.5) and lower GM thresholds (0.0 to 0.5) HC show higher number of correlations around correlation values of +0.2 and –0.2, and SZ show higher number of correlations around correlation value 0.0. This pattern is
consistent with the histogram found in the above section using all brain voxels. At higher GM thresholds (0.5 to 0.8), independent of SM $t$-values, HC show higher number of positive correlations and SZ a higher number of negative correlations. This is seen in Figure 4–2 (a), as GM thresholds are increased from 0.5 to 0.8, blue and red shades move to the left and right of zero correlation respectively. In Figure 4–2(a) color intensity reduces at higher GM and SM thresholds. This is due to the fact that at higher thresholds a lower number of voxels are selected to compute the histogram and hence the number of occurrences is reduced. To find the significance of the histograms at different thresholds we normalize the histogram by dividing it by the total number of correlation occurrences. A normalized histogram is essentially a probability density function (pdf). In Figure 4–2(b) the difference between HC and SZ normalized histograms are illustrated in a manner similar to Figure 4–2(a). From Figure 4–2(b) we see that more significant difference occurs at higher SM and GM thresholds and at these thresholds controls show more positive and patients more negative correlations.
Figure 4–2: Structural–Functional Correlation Difference (HC–SZ) Histograms with Voxels at Different Levels of Significance

Individual histograms are computed for different sets of structural and functional voxels that are selected using different thresholds. In (a) difference (HC–SZ) histograms are represented by different rows where the color denotes number of occurrences at correlation values of x–axis. The SM $t$–values are incremented in the outer loop and the GM concentration iteratively incremented in the inner loop. For example the first nine rows (separated by the dashed line) correspond to histograms of functional voxels of the whole brain and structural voxels selected with gray matter concentrations at nine thresholds (0 to 0.8 at steps of 0.1). From (a) it is seen that only at lower SM $t$–values and lower GM concentrations HC show a higher number of negative correlations (around −0.2) but the trend of HC having higher number of positive correlations (around +0.2) continues across all levels. At higher levels of SM and GM thresholds the color intensity is reduced since fewer voxels are chosen to compute the histogram. In (b) normalized histograms are computed and it is seen that the significance of group difference between HC and SZ is higher at voxels selected at higher thresholds.

$b$ Reproduced from Michael, et al. in review
4.4 Spatial Location of Structural–Functional Correlations

In the two previous sections we identified the distribution of structural–functional correlations using the method introduced in Section 3.2.1, but lost the spatial content of where in the brain these correlations were located. In this section we present results, using the method introduced in 3.2.2, where the spatial content of the correlations is retained. In Figure 4–3 we present the $z$–scores of how a functional voxel is correlated to all structural voxels. For a certain voxel from functional data its correlations (a total of $N$) with all structural voxels is found. The mean and standard deviation of these $N$ correlations are computed and the corresponding $z$–score is calculated with the null hypothesis mean as zero. This $z$–score is presented at the brain location at that particular functional voxel. This process is repeated with all functional voxels and voxels with absolute $z$–scores above 1.0 are mapped in Figure 4–3a, Figure 4–3b and Figure 4–3c for HC, SZ and HC–SZ respectively.

In HC there were large regions of positive correlations in the following anatomical regions: superior temporal gyrus, pre and postcentral gyri, superior frontal gyrus, and insula. Negative correlations were seen in the following regions in HC: middle temporal gyrus, inferior temporal gyrus and middle occipital gyrus and cuneus. In SZ the following regions showed positive correlations: inferior parietal lobule, lentiform nucleus, middle frontal gyrus, superior temporal gyrus and insula. Negative correlations in SZ were seen in precuneus, middle occipital gyrus, cuneus, fusiform gyrus, cingulate gyrus, middle temporal gyrus, culmen and anterior cingulate. In Table 4–2 we present regions that show significant differences ($|z| > 1.0$) between HC and SZ of how a certain functional voxel is correlated to all structural voxels. These regions were converted from
MNI space to Talairach coordinates and entered into a database (http://ric.uthscsa.edu/projects/tdc/) to provide the labels and volumes of contiguous regions.

**Figure 4–3: Map of significantly correlated voxels (how a functional voxel is correlated to all structural voxels) for Healthy Controls (HC), Patients with Schizophrenia (SZ) and HC–SZ**

For a certain voxel from functional data its correlation with all structural voxels are computed and using their mean and standard deviation a \( z \)-score is calculated and stored at the location of the functional voxel. This process is repeated with all functional voxels to construct a mean \( \bar{z} \)-structural correlation brain map. Regions with \(|z| > 1.0\) are shown in the brain map. The map for healthy controls (HC) is shown in (a) and patients with schizophrenia (SZ) in (b) and HC–SZ in (c). Yellow and red shades represent positive correlations and blue shades negative correlations.

\(^{c}\) Reproduced from Michael, et al. in review
Table 4–2: Significantly different brain regions between Healthy Controls (HC) and Patients with Schizophrenia (SZ) of how a functional voxel is correlated to all structural voxels.

<table>
<thead>
<tr>
<th>Region</th>
<th>R / L Volume (cc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior Temporal Gyrus</td>
<td>3.2/5.1</td>
</tr>
<tr>
<td>Superior Frontal Gyrus</td>
<td>5.2/0.9</td>
</tr>
<tr>
<td>Precentral Gyrus</td>
<td>3.1/2.2</td>
</tr>
<tr>
<td>Middle Frontal Gyrus</td>
<td>3.1/1.9</td>
</tr>
<tr>
<td>Cingulate Gyrus</td>
<td>1.5/2.8</td>
</tr>
<tr>
<td>Inferior Parietal Lobule</td>
<td>1.0/3.1</td>
</tr>
<tr>
<td>Medial Frontal Gyrus</td>
<td>1.9/1.9</td>
</tr>
<tr>
<td>Insula</td>
<td>2.9/0.8</td>
</tr>
<tr>
<td>Precuneus</td>
<td>1.0/3.1</td>
</tr>
</tbody>
</table>

| HC – SZ > 0                    |                    |
| Inferior Frontal Gyrus         | 10.0/3.7           |
| Middle Frontal Gyrus           | 10.8/2.6           |
| Middle Temporal Gyrus          | 10.8/1.4           |
| Inferior Parietal Lobule       | 8.4/3.3            |
| Superior Temporal Gyrus        | 5.8/0.8            |
| Lentiform Nucleus              | 4.7/0.6            |
| Superior Frontal Gyrus         | 1.6/1.6            |

For a functional voxel its correlations with all structural voxels are found and from the mean and standard deviation the $z$–score is computed. This process is repeated for all functional voxels for HC and SZ separately and the volumes of regions with a $|z|$ score difference >1.0 are listed in the table. These regions correspond to Figure 7c where HC–SZ>0 is shown in yellow and HC–SZ<0 in blue.

In Figure 4–4 we present the $z$–score of how a structural voxel is correlated to all functional voxels. Note that the cerebellum in HC has stronger positive correlations than SZ and the prefrontal regions in SZ have stronger positive correlations than HC. These $z$–scores were found in the same manner as explained for the previous result (functional voxel with all structural voxels). In Figure 4–4a, Figure 4–4b and Figure 4–4c regions with $|z|>1.0$ are presented for HC, SZ and HC–SZ respectively.
Figure 4–4: Map of significantly correlated voxels (how a structural voxel is correlated to all functional voxels) for HC, SZ and HC–SZ

For a certain voxel from structural data its correlation with all voxels from functional data are found and using their mean and standard deviation a z-score is calculated and stored at the location of the structural voxel. This process is repeated to all structural voxels and regions with $|z| > 1.0$ are shown as a brain map. The map for healthy controls (HC) is shown in (a) and patients with schizophrenia (SZ) in (b) and HC–SZ in (c). Yellow and red shades represent positive correlations and blue shades negative correlations.

An interesting result was that in HC, cerebellar regions showed positive correlations and prefrontal regions showed negative correlations. From Figure 4–4 (a) and (b) it is seen that HC show larger and more significant correlated regions than SZ. The following brain regions showed positive correlations in HC: cerebellum, middle occipital gyrus, pyramis, superior temporal gyrus, inferior semi–lunar lobule, precuneus, thalamus and cuneus. In the following brain regions, negative correlations were seen in HC: medial frontal gyrus, superior frontal gyrus, middle frontal gyrus, precentral gyrus, inferior frontal gyrus, cingulate gyrus, anterior cingulate and insula. SZ did not exhibit large contiguous regions with either positive or negative correlations. In Table 4–3 we present the regions and their volumes, that show group differences of $|z|>1.0$ of how a certain structural voxel is correlated to all functional voxels.

---

d Reproduced from Michael, et al. in review
Table 4–3: Significantly different brain regions between healthy controls (HC) and patients with schizophrenia (SZ) of how a structural voxel is correlated to all functional voxels

<table>
<thead>
<tr>
<th>Region</th>
<th>R / L Volume (cc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerebellum</td>
<td>22.9/23.5</td>
</tr>
<tr>
<td>Middle Occipital Gyrus</td>
<td>4.2/3.8</td>
</tr>
<tr>
<td>Middle Frontal Gyrus</td>
<td>3.8/3.5</td>
</tr>
<tr>
<td>Cuneus</td>
<td>3.9/3.1</td>
</tr>
<tr>
<td>Middle Temporal Gyrus</td>
<td>2.6/3.8</td>
</tr>
<tr>
<td>Precuneus</td>
<td>2.0/2.5</td>
</tr>
<tr>
<td>Supramarginal Gyrus</td>
<td>2.8/0.9</td>
</tr>
<tr>
<td>Superior Temporal Gyrus</td>
<td>2.1/1.2</td>
</tr>
<tr>
<td>Inferior Temporal Gyrus</td>
<td>2.0/1.1</td>
</tr>
<tr>
<td>Inferior Occipital Gyrus</td>
<td>1.9/1.2</td>
</tr>
<tr>
<td>Thalamus</td>
<td>1.7/1.4</td>
</tr>
<tr>
<td>Insula</td>
<td>2.9/0.1</td>
</tr>
<tr>
<td>Medial Frontal Gyrus</td>
<td>9.9/7.4</td>
</tr>
<tr>
<td>Superior Frontal Gyrus</td>
<td>5.4/6.7</td>
</tr>
<tr>
<td>Middle Frontal Gyrus</td>
<td>3.7/5.4</td>
</tr>
<tr>
<td>Cingulate Gyrus</td>
<td>3.6/5.4</td>
</tr>
<tr>
<td>Anterior Cingulate</td>
<td>4.7/3.2</td>
</tr>
<tr>
<td>Precentral Gyrus</td>
<td>1.1/4.0</td>
</tr>
<tr>
<td>Lingual Gyrus</td>
<td>1.4/3.1</td>
</tr>
<tr>
<td>Inferior Frontal Gyrus</td>
<td>2.0/1.9</td>
</tr>
<tr>
<td>Insula</td>
<td>0.2/3.3</td>
</tr>
<tr>
<td>Middle Temporal Gyrus</td>
<td>1.6/1.6</td>
</tr>
<tr>
<td>Cuneus</td>
<td>0.4/2.6</td>
</tr>
<tr>
<td>Posterior Cingulate</td>
<td>1.0/2.0</td>
</tr>
</tbody>
</table>

For a structural voxel its correlations with all functional voxels are found and from the mean and standard deviation the $z$–score is computed. This process is repeated for all structural voxels for HC and SZ separately and the volumes of regions with a $|z|$ score difference $>1.0$ are listed in the table. These regions correspond to Figure 8c where HC–SZ$>0$ is shown in yellow and HC–SZ$<0$ in blue.

### 4.5 Structural–Functional Inter Regional Correlations

In previous sections we presented the distribution of correlations and an average of where high correlations were located. In this section using the method introduced in Section 3.2.3 we identify regions in the brain that have significant structural–functional
linkages. With this method the structural–functional cross correlation matrix \((R_{SF})\) is reduced to a matrix of 116x116 \(z\)-scores of inter–regional correlations. The cerebellar vermis from the structural data had significant positive correlation in HC with the following regions from the functional data: calcarine, cuneus, lingual gyrus, paracentral lobule, and Heschl’s gyrus. In SZ there were no positive correlation above a \(z\)-score of 4.0 but the basal ganglia from the structural data had significant negative correlation with the posterior cingulate from the functional data. We computed the significance of the above inter–regional correlations and they were all \(p < 0.01\).

Our next interest was to determine the structural–functional inter–regional correlations that were significantly different between HC and SZ. There were no inter–regional correlation where the \(z\)-score of SZ–HC was greater than 4.0. There were a few inter–regional correlations where HC–SZ was greater than 4.0 and are listed in Table 4–4. The results show that structural regions in the cerebellum have significantly higher structure – function correlations in HC than in SZ.
Table 4–4: Structural and functional regions that had a high correlation z–score difference, (HC–SZ) > 4.0

<table>
<thead>
<tr>
<th>Structural Regions</th>
<th>Functional Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal Ganglia (Lentiform Nucleus)</td>
<td>Cingulum Gyrus / Precuneus (L)</td>
</tr>
<tr>
<td>Superior Temporal Gyrus / Transverse Temporal Gyrus / Insula (L)</td>
<td>Cerebellum (Culmen)</td>
</tr>
<tr>
<td>Cerebellar Tonsil (L)</td>
<td>Cerebellum (Vermis, Uvula, Culmen, Declive)</td>
</tr>
<tr>
<td>Cerebellum (Culmen)</td>
<td>Precuneus / Cuneus / Lingual Gyrus / Insula / Transverse Temporal Gyrus / Superior Temporal Gyrus</td>
</tr>
</tbody>
</table>

The structural–functional cross correlation matrix ($R_{SF}$) was segmented into regional correlations as defined by the AAL atlas. For each segment a $z$–score was computed using the mean and standard deviation. This was done for controls and patients separately and inter–regions that had $z$–score differences greater than 4.0 are listed in the table. For example structural regions in the Superior/Transverse Temporal Gyrus and the insula had a $z$–score correlation difference (HC–SZ) of greater than 4.0 with functional regions in the Culmen.
4.6 Regions with Significant Structural–Functional Correlations

In the above methods we found an average value for the correlations. The question of how significant these correlations are was not answered. In this section we store the location of voxels that have significant structural–functional correlations. For this purpose we use the method explained in Section 3.2.4 to find the number of functional voxels that have a significant \( p < 10^{-6} \) correlation with a certain structural voxel. This process is then repeated to all structural voxels. The difference (HC–SZ) images of how structural regions that have significant \( p<10^{-6} \) structural–functional correlation are shown in Figure 4–5. HC show larger significant correlation regions and in addition they have higher number of significant correlations with functional voxels than in SZ.

\[ R \quad HC - SZ \quad L \]

\[ \begin{array}{cccc}
  +48 & +39 & +30 & +21 \\
  +12 & +3 & +6 & -15 \\
 -24 & -33 & -42 & -51 \\
\end{array} \]

Figure 4–5: Structural regions that have significantly higher \( (P<10^{-6}) \) correlations in HC than SZ. Shades of red and blue denote \((HC–SZ)>0\) and \((HC–SZ)<0\) respectively. The color intensity corresponds to the number (as indicated by the colorbar) of functional voxels with significant \( (P<10^{-6}) \) correlation with structural voxel at that location.\(^e\)

\(^e\) Reproduced from Michael, et al. in review
4.7 Discussion

The main purpose of this work was to present fusion results of structural and functional data through cross–correlations across the whole brain. The approaches were applied to a group of patients with chronic schizophrenia and matched healthy controls to investigate if the degree of association between the gray matter concentration and activation for an auditory sensorimotor task were different in these two groups. Correlations were evaluated across subjects for all gray matter voxels on a sensorimotor task activation map and reduced to statistical measures using different methods. The large structural–function cross–correlation matrix is reduced to metrics through three different methods. With the histogram of all possible correlations between gray matter and the fMRI activation maps we were able to demonstrate the distribution of the correlations, however, the spatial location of the correlations were not retained. By collapsing the rows/columns of cross correlation matrix we preserved the spatial information of the correlations. The question of whether to collapse along the rows/columns to obtain the difference between the groups can be answered in the following manner. By collapsing in two different directions two different questions are being answered. By collapsing the columns we find how a certain structural voxel is correlated to all functional voxels and the corollary is obtained by collapsing across the rows. By computing the average correlation across the whole brain useful information can be lost. For example some regions in the brain can have negative correlations and other positive and an average can indicate a correlation value close to zero. To account for this type of variation in correlation we report the $z$–score. With the help of the AAL atlas we made an effort to find significantly correlated links between structural and functional brain regions. To find
brain regions that had significant structural–functional correlations we stored the number of significantly correlated voxels. In all of the above methods, the result of weaker structural–functional correlation in SZ than in HC was consistent. The result of structural regions in the cerebellum of HC having stronger positive correlation with functional activation was also consistent across different methods.

The results reported in this study were derived from a total of 140 subjects, seventy in each group, scanned at four different brain imaging centers. The large number of subjects makes our result more robust. The auditory sensorimotor functional task was kept identical across all four sites and the raw images were preprocessed with the same script. Data quality was ensured with a multi–step QA process and one of them was the application of the simple yet effective and efficient outlier detection method introduced in Section 2.6.

The application of the methods introduced, onto chronic patients with schizophrenia and matched healthy controls, revealed several interesting findings. The main finding was that patients with schizophrenia showed less correlation between structural data (gray matter concentration) and functional data (obtained from the activation map of a sensorimotor task) than healthy controls. This result was confirmed in different ways: the histogram of structural–functional correlations, a spatial map of correlations and a linkage map between structure and function.

We used three different methods to measure the relationship between structure and function between patients and controls, and all three approaches yielded consistent results. In our first result we found that at correlation values centered around +0.2 and –
0.2 the controls have a higher number of correlations than patients and at zero correlation patients had a higher number of correlations than controls. This finding suggests that the influence of gray matter regions on functional activity as probed by the sensorimotor task is weaker in patients than in controls. The difference between patients and controls at +0.2 was greater than that at −0.2. This result was obtained when all brain voxels were included in the analysis. To examine how this correlation difference changed across different levels of gray matter concentration and functional activation level we computed an image of histograms (see Figure 4–2). We found that voxels with higher gray matter concentration voxels had more positive correlations in controls and more zero and negative correlations in patients. This result was consistent across different thresholds of $t$–values of the sensorimotor task.

Results from Section 4.4 and Section 4.6 results show continuous regions of high correlations in both patients and controls. This indicates that highly correlated regions were not located in randomly distributed brain regions, but were spatially clustered, which increases the validity of our results. The spatial locations of the correlations were obtained by collapsing the $R_{SF}$ matrix along the rows or along the columns. If collapsed along the rows, the $z$–scores map indicates how a functional voxel is correlated to all structural voxels. This map (Figure 4–3) showed larger and more significant correlation regions in controls than in patients. The functional task consisted of auditory input and hence it was reasonable to expect high correlations in the temporal regions. In SZ the superior temporal gyrus did show positive correlations but the volume was smaller compared to HC. In superior temporal gyrus and superior frontal gyrus HC show more positive correlation than SZ (Table 4–2). In inferior frontal gyrus, middle frontal gyrus
and middle temporal gyrus SZ show more positive correlation than HC. In the superior temporal gyrus and superior frontal gyrus there were subregions where either HC or SZ showed more positive correlations. The question of whether this difference was caused by inaccurate registration of subject data is reasonable. The spatial normalization step in preprocessing used the same template to normalize patient and control subject data. Minimal topological differences can be attributed to spatial normalization but volume differences as large as listed in Table 4–2 are unlikely to be due to registration issues. Another advantage of the method introduced is that it negates issues pertaining to registration. Instead of performing region of interest (ROI) analyses our method investigates voxels from the whole brain. In ROI methods precise registration is crucial and even the slightest misalignments can cause errors in results. In our method voxels are taken from subjects at all spatial locations of the template and errors due to registration will be minimized.

In Figure 4–4 and Table 4–3 regions with high correlation between structural voxels and functional voxels are presented. Here controls demonstrate significantly more correlated regions than patients. The main findings are that cerebellar regions in HC had more positive correlations and prefrontal regions in HC had more negative correlations than SZ. In controls the correlations in cerebellar regions in the left hemisphere were larger and stronger than the corresponding right region (Figure 4–4a). The question of the sensorimotor task influencing this result is unlikely given that the task requires a right thumb press which activates right cerebellar regions (and not left) as shown in Figure 2–8. We performed a two sample $t$–test on gray matter concentration and did not find cerebellar regions to be significantly different between the two groups. This indicates that
between group differences in concentration did not contribute to the differences in correlations.

To retain the linkage between different structural and functional regions we segmented the brain with the AAL atlas. In controls, structural regions in the cerebellum show significantly more positive correlations with functional regions (Table 4–4) than in patients. While computing structural–functional inter regional correlations, we do not find significant correlations between the same structural and functional regions. This should encourage researchers to investigate correlations between distant brain regions. This point supports the hypotheses that local morphological structures can affect distant functional activation.

It is possible that higher correlations can be introduced as a result of head motion while correlating voxels within the same brain. Let us assume that while data are being acquired group 1 has still heads inside the scanner and group 2 has head motion. If in group 2 a certain voxel is active, due to head motion, its activation will spread to neighboring voxels in the activation map. This can make correlations higher in group 2 than in group 1. In our results we see less correlation in patients than in controls. However, in reality it is the patients who have more head motion or tremors than the controls (Caligiuri, et al. 1993). If head motion had contributed to the results then our results should have been flipped with patients showing more correlation than controls.

In the theory of cognitive dysmetria (Andreasen, et al. 1998) it is hypothesized that the system that is disturbed in schizophrenia is more distributed and it is modeled as a dysfunction in cortical–subcortical–cerebellar circuitry. This theory states that deficits
in the complex distributed neural circuitry within the brain can express itself with a broad range of symptoms. In our results we report regions that are widely distributed across the whole brain, especially in the cerebellar, subcortical and prefrontal regions, which interestingly are the nodes that fall under the rubric of the cognitive dysmetria model. The cerebellum has substantial connections with the prefrontal cortex and can perform parallel processing supported by its array structure and large number of condensed cells (Andreasen, et al. 1998). In our Method 2 results of how gray matter concentration is correlated with functional voxels of the whole brain and Method 3 results of significantly correlated brain regions, controls showed higher correlation in cerebellar regions than do patients. The results reported in this study also support the disconnection hypothesis (Friston 1998) of schizophrenia. The histogram of correlations suggests less connection between structure and function in schizophrenia. The spatial location of correlations show larger and stronger regions of correlations in controls than in patients. Results of the correlation between structural and functional regions as defined by an anatomical atlas suggest that controls have stronger positive inter–connections than patients. The disconnection hypothesis states that the disconnection is explicitly functional, not anatomical. This effective connectivity, as opposed to anatomical connectivity, is hypothesized as a result of differences in synaptic efficacy (Friston 1998). In this work we do not directly investigate the anatomical connections between different regions of the brain but the correlation between gray matter concentration and functional activation. This investigation does not report on synaptic efficacy but provides evidence that further sources can exist for sMRI/fMRI linkage differences in schizophrenia.
5 Functional – Functional Analysis

Single functional MRI (fMRI) task analysis methods of functional MRI brain data, though useful, are not able to evaluate the joint information between tasks. Data fusion of multiple tasks that probe different cognitive processes provides knowledge of the joint information and may be important in order to better understand complex disorders such as schizophrenia. In this chapter we apply methods introduced in Section 3.1 to fuse imaging data from two tasks at a time to compute the histogram of correlations for all possible combinations of whole brain voxels. Most of the discussion and results presented here are from Michael, et al. (2008a), Michael, et al. (2008c) and Michael, et al. in press.

Many studies have found brain region differences of patients with schizophrenia using a single task. It is becoming a common practice to scan subjects while they perform different tasks but each task is typically analyzed separately (Derbyshire, et al. 1998; Liu, et al. 2004). Such approaches can be used to understand differential activation in the same voxels but they do not examine the joint information between different tasks and different voxels. Tools that enable the examination of joint information between tasks
that probe different functional domains can lead to new understanding of the complex disorder.

In this study auditory oddball (AOD), Sternberg item recognition paradigm (SIRP) and sensorimotor (SM), are utilized. The paradigms are described in Section 2.3.2. We analyze the target and novel stimuli in the AOD task separately and these two stimuli will be referred to as separate tasks to simplify the language. Abnormal information processing has been hypothesized to be a core deficit in schizophrenia (Braff 1996; Callaway and Naghdi 1982; Perry and Braff 1994). The AOD task was selected as a probe to assess the inefficiency in information processing. In AOD the infrequent task relevant stimuli elicits a positive brain potential, measured through EEG, referred to as P3 or P3b. P3 is one of the most robust functional abnormalities found in chronic medicated schizophrenic patients (Ebmeier, et al. 1990; McCarley, et al. 1993) and manifests as a decrease in the temporal lobe amplitude. Similar findings have been shown for fMRI data as well, again particularly in temporal regions (Kiehl and Liddle 2001).

Working memory, or the ability to hold a representation and perform cognitive operations allows individuals to formulate, modify and hold a plan in mind (Baddeley 1992). In schizophrenia, working memory deficits have been demonstrated in medicated, unmedicated (Andreasen, et al. 1992b) and in healthy relatives of schizophrenic patients (Ashton, et al. 1995). The SIRP (Sternberg 1966) task was used as a functional probe of working memory and fMRI activation differences have been previously identified for this task (Manoach, et al. 2000; Manoach, et al. 1999). The SM task was designed to activate the auditory cortex robustly. It was initially designed for calibration purposes for assessing and controlling between–site, within–site and within–subject variability. Data
collected from this task has shown significant group differences and hence is included in the analysis.

We show how the fusion of tasks can help to find new differential features between groups of schizophrenic patients and matched healthy controls. A widespread analysis incorporating all brain voxels of different tasks is done to provide further insight into the distributive nature of schizophrenia. We expected to replicate the finding of (Calhoun, et al. 2006) without excluding brain regions or components and hypothesize that patients will show more positive inter–task correlations for all task combinations. We will then check this result by several methods including a Monte Carlo (Metropolis and Ulam 1949) method to confirm that this result did not occur by chance. Brain regions that show high inter–task correlations are found to identify regions that activate differently for the two groups of subjects.

An overview of our functional–functional analysis is given in Figure 5–1. In this analysis subjects with schizophrenia and healthy controls are scanned with a set of identical fMRI tasks. We then pair wise combine these tasks to find a differential feature between the patients and controls. The new differential feature found is then applied to develop a classification algorithm.
Figure 5–1: Flow chart of the functional–functional analysis

5.1 Subjects

The subjects for this study were scanned at the University of Iowa Hospital and are part of a larger study (MCIC: MIND Clinical Image Consortium). These subjects are a subset of subjects that were used in the previous chapter (see Section 4.1). Subjects were recruited via advertisements and by word–of–mouth, provided written, informed, IRB (University of Iowa) approved consent and were compensated for their participation. Among the subject pool there were 28 healthy controls (HC) and 19 ($n_p$) patients with chronic schizophrenia (SZ). After careful examination of the behavioral data it was found that two HC had not successfully completed the AOD task and they were removed from the analysis reducing the number of HC to 26 ($n_c$). Three subjects performed poorly in some of the runs of the AOD task and these runs were excluded from the analysis.
Healthy subjects in the study were screened to ensure they were free from diagnostic and statistical manual of mental disorders (DSM–IV) axis I or axis II psychopathology, assessed using a modified version of the comprehension assessment of symptoms and history (CASH) (Andreasen, et al. 1992a). They were interviewed to determine that there was no history of psychosis in any first–degree relatives. Patients met criteria for schizophrenia in the DSM–IV based on CASH and review of the case file. All participants analyzed in this study had normal hearing (assessed by self report) and were able to carry out all tasks successfully. There was no significant age and socioeconomic status differences between the two groups and are given in Table 5–1 along with other demographics. Average patient symptom measures, positive (Andreasen 1984), negative (Andreasen 1981), and disorganization, are also reported in Table 5–1.

Most patients, except three (for whom information was not recorded) were receiving antipsychotic medication: clozapine (3), quetiapine (3), olanzapine (3), ziprasidone (2), risperidone (2), haloperidol and olanzapine (1), quetiapine and aripiprazole (1), perphenazine and aripiprazole (1).
Table 5–1: Demographics and Clinical Characteristics of Patients with Schizophrenia (SZ) and Healthy Controls (HC)

|                      | SZ  
|----------------------|-----
|                      | \(n_p=19\)  
| **Age**             | 34 ± 11 (range: 19–59)  
|                      | 33 ± 11 (range: 19–57)  
| **Male, Female**     | 15 males  
|                      | 12 males  
|                      | 4 females  
|                      | 14 females  
| **Handedness**       | 2  
| (non–right hand)     | 1  
| **Socioeconomic**    | 2.6 ± 0.8  
| **Status**           | 2.8 ± 0.5  
| **Average Years**    | 13.3  
| **Since Diagnosis**  | NA  
| **Symptoms**         | positive = 3.89 ± 3.46  
|                      | negative = 7.58 ± 3.45  
|                      | disorganization = 1.74 ± 1.37  
|                      | NA  

5.2 Behavioral Results

The behavioral results for the AOD and SIRP tasks are listed in Table 5–2. For the AOD task 42 targets were presented in all four runs. The percentage of correct responses and the percentage of responses to non–target stimuli (Errors of commission) did not show significant difference between SZ and HC. For the SIRP task high success rates (correctly identifying a probe digit) were recorded for all three difficulty levels for both groups but again no significant group differences were noted. In Table 5–2 we also report Cohen’s \(d\) values which indicate the degree of overlap between the two groups. A \(d\) value of 0.61 corresponds to an overlap of 61.3% and a \(d\) value of 1.05 corresponds to an overlap of 42.6 %.
Table 5–2 Behavioral Results for the AOD and SIRP Tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>SZ</th>
<th>HC</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AOD Task</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Hits (%)</td>
<td>95 ± 7.2</td>
<td>98.2 ± 1.8</td>
<td>0.61</td>
</tr>
<tr>
<td>Errors of Commission (%)</td>
<td>2.8 ± 2.6</td>
<td>2.4 ± 2.1</td>
<td></td>
</tr>
<tr>
<td><strong>SIRP Task</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Hits (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–t</td>
<td>97.1 ± 4.6</td>
<td>99.2 ± 1.1</td>
<td>0.62</td>
</tr>
<tr>
<td>3–t</td>
<td>96.0 ± 4.3</td>
<td>98.7 ± 1.3</td>
<td>0.85</td>
</tr>
<tr>
<td>5–t</td>
<td>96.4 ± 2.6</td>
<td>98.6 ± 1.4</td>
<td>1.05</td>
</tr>
</tbody>
</table>

5.3 Inter–Task Correlation Histograms

The approach explained in Section 3.2.1 was separately applied to the set of HCs and SZs. In Section 3.2.1 we formulated the problem as combining data from Modality 1 and Modality 2. For this chapter Modality 1 and 2 data are obtained from different fMRI tasks to compute inter–task correlation histograms by pairing two tasks at a time. We analyze four tasks, two from AOD and one each from SM and SIRP, and combine them two at a time to give six possible task combinations. In Figure 5–2 the correlation histograms are shown for all six combinations by different sub plots.
Figure 5–2: Inter–Task Spatial Correlation Histograms with All Brain Voxels

Histograms for different task combinations shown for patients (red) and controls (blue). Solid blue line represents when all 26 controls were included in the analysis and dotted blue line when 19 (to match the number of patients) randomly selected controls were analyzed. Note that in four out of six inter–task combinations the patients’ histogram is shifted to the right indicating higher inter–task similarity than controls.

In Figure 5–2, the y–axis represents the number of occurrences of correlation values given along the x–axis. From the histogram it is possible to understand the nature and the distribution of elements of the large correlation matrix. A mean value of zero means that on average the activation of the voxels of the two tasks are uncorrelated, a

---

*a* Reproduced from Michael, et al. in press
positive value means that when activations increase or decrease in one task, activations in the other task also increase or decrease, and a negative value means that if activations increase in one task the activations decrease in the other and vice versa.

In Figure 5–2 the histogram of HC \((n_c=26)\) and SZ \((n_p=19)\) are shown by blue (solid) and red lines respectively. In four task combinations, AOD Target + SM, AOD Target + SIRP, AOD Novel + SM and SM + SIRP the mean value of SZ histogram is more positive than that of HC. In the AOD Target + AOD Novel comparison the mean value of the HC is more positive than that of the SZ and in the AOD Novel + AOD SIRP comparison the histograms almost overlap showing no significant difference. In Table 5–3 statistical measures (mean, standard deviation, skewness, kurtosis, the percentage of correlations higher than a threshold \((\rho_{th})\) and the percentage of correlations below \(-\rho_{th}\) are listed for SZ, HC with the difference. The percentages above or below a threshold correspond to the area under the histogram curve and \(x > \rho_{th}\) and the area under the curve and \(x < -\rho_{th}\) respectively. Since the two groups have different numbers of subjects (and hence different degrees of freedom), we compute the appropriate correlation threshold corresponding to a significance of \(p < 0.01\) in the following manner. For a \(p\) value of 0.01 we found the \(t\) value for \(n_p = 19\) and \(n_c = 26\) using the ‘tinv’ function in the statistical toolbox of MATLAB. These \(t\) values were then converted to correlation coefficient values using Equation (3.2). For SZ \((n = 19)\) the threshold \(\rho_{th}\) is equal to 0.57 and for HC \((n = 26)\) the threshold \(\rho_{th}\) is equal to 0.493. Neither 0.493 nor 0.57 are bin values in the histogram and we interpolate the neighboring bin values to find the number of correlations above 0.493 and 0.57 for HC and SZ respectively.
Table 5–3: Characteristics of Inter–Task Spatial Correlation Histograms with All Brain Voxels

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>+Corr. (%) (p &lt; 0.01)</th>
<th>−Corr. (%) (p &lt; 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target + Novel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>–0.09</td>
<td>0.07</td>
<td>0.29</td>
<td>0.17</td>
<td>2.56</td>
<td>1.49</td>
</tr>
<tr>
<td>HC</td>
<td>0.07</td>
<td>0.09</td>
<td>0.09</td>
<td>0.22</td>
<td>2.76</td>
<td>1.25</td>
</tr>
<tr>
<td>Δ</td>
<td>–0.16</td>
<td>–0.17</td>
<td>0.24</td>
<td>–0.06</td>
<td>–0.66</td>
<td>–0.22</td>
</tr>
<tr>
<td>Target + SM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>0.12</td>
<td>0.12</td>
<td>0.25</td>
<td>0.19</td>
<td>0.58</td>
<td>0.30</td>
</tr>
<tr>
<td>HC</td>
<td>–0.03</td>
<td>0.23</td>
<td>0.25</td>
<td>–0.04</td>
<td>0.37</td>
<td>0.43</td>
</tr>
<tr>
<td>Δ</td>
<td>0.15</td>
<td>0.02</td>
<td>0.09</td>
<td>–0.54</td>
<td>0.27</td>
<td>–0.54</td>
</tr>
<tr>
<td>Target + SIRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>0.16</td>
<td>0.16</td>
<td>0.22</td>
<td>–0.26</td>
<td>0.26</td>
<td>0.09</td>
</tr>
<tr>
<td>HC</td>
<td>–0.03</td>
<td>0.23</td>
<td>0.24</td>
<td>–0.23</td>
<td>0.75</td>
<td>0.09</td>
</tr>
<tr>
<td>Δ</td>
<td>–0.19</td>
<td>0.23</td>
<td>0.26</td>
<td>–0.29</td>
<td>0.91</td>
<td>0.12</td>
</tr>
<tr>
<td>Novel + SM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>0.05</td>
<td>0.05</td>
<td>0.26</td>
<td>–0.12</td>
<td>0.67</td>
<td>1.18</td>
</tr>
<tr>
<td>HC</td>
<td>–0.09</td>
<td>0.14</td>
<td>0.22</td>
<td>–0.12</td>
<td>0.76</td>
<td>1.35</td>
</tr>
<tr>
<td>Δ</td>
<td>0.14</td>
<td>0.03</td>
<td>0.08</td>
<td>–0.09</td>
<td>0.73</td>
<td>–0.73</td>
</tr>
<tr>
<td>Novel + SIRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>–0.02</td>
<td>0.21</td>
<td>0.12</td>
<td>0.12</td>
<td>0.34</td>
<td>0.31</td>
</tr>
<tr>
<td>HC</td>
<td>–0.01</td>
<td>0.21</td>
<td>0.05</td>
<td>0.02</td>
<td>0.34</td>
<td>0.31</td>
</tr>
<tr>
<td>Δ</td>
<td>0.07</td>
<td>0.01</td>
<td>0.07</td>
<td>–0.63</td>
<td>0.43</td>
<td>–0.63</td>
</tr>
<tr>
<td>SM + SIRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>0.23</td>
<td>0.23</td>
<td>0.21</td>
<td>2.91</td>
<td>5.66</td>
<td>0.02</td>
</tr>
<tr>
<td>HC</td>
<td>0.06</td>
<td>0.21</td>
<td>0.24</td>
<td>2.91</td>
<td>5.66</td>
<td>0.02</td>
</tr>
<tr>
<td>Δ</td>
<td>0.16</td>
<td>0.01</td>
<td>0.18</td>
<td>0.19</td>
<td>3.05</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Statistical measures for SZ, HC and their difference (Δ) for inter–task correlation histograms (Figure 5–2) are listed. The HC histograms were found with all 26 controls and a random selection of 19 controls to match the number of patients. +Corr. and −Corr. percentages were found using a correlation threshold of +0.57 and −0.57 respectively, when 19 subjects were used and +0.49 and −0.49 respectively, when 26 subjects were used. These thresholds correspond to a significance of p=0.01.

Standard deviations for all combinations of tasks are higher in SZ than in HC. For the same four combinations where the mean of SZ is higher than that of HC, skewness is more negative in SZ than HC. Skewness can also be recognized Figure 5–2 where we have marked the mode (box) and mean (circle) of each histogram. If a histogram is skewed its mode and mean do not overlap over each other. If the mode is larger than the mean, then it is more positive in SZ than HC. Skewness can also be recognized Figure 5–2 where we have marked the mode (box) and mean (circle) of each histogram.
mean, as in four of the SZ histograms, it indicates that the curve is tilted to the right. This tilt implies that in SZ correlations are more spread over negative values and more concentrated over positive values. The negative difference in skewness between SZ and HC for the four task combinations indicates that SZ histograms are more tilted towards positive correlations than HC histograms. The difference in Kurtosis does not show a pattern consistent to the pattern seen with the mean and skewness. The percentage of correlations greater than $\rho_{th}$ is higher in SZ for the same four combinations where mean of SZ is higher. Percentage of correlations less than $-\rho_{th}$ is higher in HC than SZ for all task combinations except AOD Target + AOD Novel. In all of the above comparisons it is seen that SZ have higher positive correlations than HC for AOD Target + SM, AOD Target + SIRP, AOD Novel + SM and SM + SIRP.

The above comparisons were computed using data from 26 HCs and 19 SZs. The shape of the curves, specifically standard deviation and kurtosis, is dependent on the number of subjects. According to the central limit theorem a small number of data points or few subjects will make the histogram look flatter and thereby increase and decrease the standard deviation and kurtosis respectively. In that sense the assessments in the previous paragraph were not made in a consistent manner since the number of subjects in each group was different. To make a fair comparison 19 HCs were randomly chosen to match the number of SZs and histograms were recomputed with this new data set. The histograms are shown in Figure 5–2 by blue dotted lines. The curve properties of this histogram are listed in Table 5–3 next to the values corresponding to 26 HCs. For this test the $\rho_{th}$ for each group is 0.57 corresponding to $p$ value of 0.01 for 19 subjects. Even though for some task combinations the mean of the histograms shifted, the difference in
group means shows the same pattern as before. The standard deviation increases and kurtosis decreases (see Table 5–3) for all task combinations in HC. As expected from the central limit theorem, the standard deviation and kurtosis of our histograms are dependent on the number of subjects. For this reason and due to the fact that these parameters do not show a consistent pattern in terms of the difference between SZ and HC they will not be considered for further comparisons. All other parameters of group differences followed the same pattern seen previously. From these two tests, in terms of mean, skewness and percentage of correlations we have shown that in four inter–task combinations SZ shows more positive inter–task correlations than HC, in one combination HC shows more positive correlation than SZ and in another there was no significant difference. One could argue the validity of the consistency check since it was done on just one random subset of subjects. The point of this comparison was not to prove the robustness of the result but to exclude statistical measures that are clearly dependent on the number of subjects. In a later section we will evaluate robustness with a more rigorous method.

Another potential ambiguity of the results arises from the fact that the two subject groups have unequal number of males and females, 15 males and 4 females in SZ and 12 males and 14 females in HC. An obvious question is whether the group differences we observe are due to schizophrenia or due to the gender disparity. To answer this question we randomly selected equal number of males (12) and females (4) from each group and ran the inter–task correlation algorithm. The histogram parameters are listed in Table 5–4. For the same four tasks, AOD Target + SM, AOD Target + SIRP, AOD Novel + SM and SM + SIRP the mean and skewness shows that SZ had more positive correlations than did HC and for the AOD Target + AOD Novel and AOD Novel + SIRP the vice
versa. These results are consistent and the values of the differences are comparable to previous results. Threshold ($\rho_{th}$) for this test was found using 16 subjects and was 0.65 for both groups. Percentage of positive correlations above the threshold is higher in SZ for all combinations except AOD Novel + SIRP. The difference between the percentages of negative correlations below the threshold does not show large margins except for AOD Target + AOD Novel. This test confirms that the difference between SZ and HC we detect does not appear to be due to the gender disparity in the two groups.

Table 5–4: Characteristics of Inter–Task Spatial Correlation Histograms with All Brain Voxels for Equal Numbers of Males (12) and Females (4) in Each Group

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Skewness</th>
<th>+Corr. (%) ($p &lt; 0.01$)</th>
<th>–Corr. (%) ($p &lt; 0.01$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target + Novel</td>
<td>SZ</td>
<td>–0.10</td>
<td>0.19</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.01</td>
<td>–0.03</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Δ</td>
<td>–0.11</td>
<td>0.22</td>
<td>0.18</td>
</tr>
<tr>
<td>Target + SM</td>
<td>SZ</td>
<td>0.14</td>
<td>–0.24</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.05</td>
<td>–0.05</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Δ</td>
<td>0.09</td>
<td>–0.19</td>
<td>1.16</td>
</tr>
<tr>
<td>Target + SIRP</td>
<td>SZ</td>
<td>0.20</td>
<td>–0.37</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.03</td>
<td>–0.07</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Δ</td>
<td>0.18</td>
<td>–0.30</td>
<td>0.70</td>
</tr>
<tr>
<td>Novel + SM</td>
<td>SZ</td>
<td>0.04</td>
<td>–0.11</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>–0.08</td>
<td>0.08</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Δ</td>
<td>0.11</td>
<td>–0.19</td>
<td>0.67</td>
</tr>
<tr>
<td>Novel + SIRP</td>
<td>SZ</td>
<td>–0.06</td>
<td>0.19</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.04</td>
<td>–0.05</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Δ</td>
<td>–0.10</td>
<td>0.24</td>
<td>–0.32</td>
</tr>
<tr>
<td>SM + SIRP</td>
<td>SZ</td>
<td>0.22</td>
<td>–0.29</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.02</td>
<td>–0.04</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>Δ</td>
<td>0.20</td>
<td>–0.25</td>
<td>1.24</td>
</tr>
</tbody>
</table>

Statistical measures for SZ, HC and their difference (Δ) for inter–task correlation histograms are listed with equal numbers of males and females in each group. +Corr. and –Corr. percentages were found using a correlation threshold of +0.65 and –0.65 respectively, corresponding to a significance of $p=0.01$ for 16 subjects.
5.4 Correlation Histograms of Significant Task Related Voxels

In the previous analyses inter–task correlations were found for all voxels of the brain. During the preprocessing step we had removed voxels that were considered ‘noisy’ as a result of scanner issues. Another set of ‘noisy’ voxels appear when voxels do not activate consistently across all subjects. We were interested in checking the above result for the statistically significant activation voxels. Voxels that had the highest $t$ values for each task were chosen as significant voxels. For a certain brain voxel its $t$ value was found using its mean and standard deviation using information from all subjects, both SZ and HC. By this method inter–task correlations are found between the same set of voxels for the two groups. Inter–task correlation histograms were computed for the $N = 10^4$ and $N = 10^3$ most significant voxels, corresponding to ~20% and ~2% of the total number of voxels. We selected these large numbers arbitrarily since with a large number of voxels it enables us to investigate the functional connectivity of brain networks. These histograms are shown in Figure 5–3 and for this analysis all 26 controls were included.
**Figure 5–3: Inter–Task Spatial Correlation Histograms with the Most Significant Task Related Voxels**

Histograms computed with $N = 10^4$ (solid lines) and $N = 10^3$ (dotted lines) most significant voxels found through $t$–test of all subjects (26 controls, blue and 19 patients, red). The total number of correlations is different depending on the number of voxels used. The histograms with 1000 most significant voxels were scaled by 100 to depict them on the same plot. Higher inter–task similarity in patients than controls in four out of six task combinations is true for task related voxels.

In Table 5–5 the statistical measures for the histogram curves are listed. The histograms corresponding to $N = 10^4$ and $N = 10^3$ voxels are shown by solid and dotted lines respectively. The threshold ($\rho_{th}$) to find the percentage of correlations was 0.57 for SZ and 0.49 for HC. For $N = 10^4$ voxels the mean, skewness and percentage of

---

\[b\] Reproduced from Michael, et al. in press
correlations showed that AOD Target + SM, AOD Target + SIRP, AOD Novel + SM and SM + SIRP SZ (same four tasks as in previous analyses) have a higher number of positive correlations than HC and the AOD Target + AOD Novel and AOD Novel + SIRP the vice versa. In Figure 5–3 the mean (circle) and mode (square) of the histograms are marked for $N = 10^4$ voxels and from their locations we see that SZ histograms are more tilted towards positive correlations than HC for all task combinations except AOD Target + AOD Novel and AOD Novel + SIRP. For $N = 10^3$ voxels the difference in mean follows the same pattern as for the $N = 10^4$ voxels except for AOD Novel + SIRP. The other parameters too follow the same pattern for most of the task combinations with a few exceptions. The mean and mode are not marked on the histograms for better clarity of the Figure.

5.5 A Monte–Carlo Method Verification of Results

The results presented above were based on a single dataset and outliers in either group could have created the difference between SZ and HC. In our method outliers cannot be easily found since it requires a group of subjects to find the inter–task correlation histogram. The most straightforward and accurate, but expensive and time consuming, way to check the validity of this result would be to try and replicate the same result on a completely different set of data. An alternative method is to select different subsets within each group, run the correlation algorithm and check if results follow the same pattern.
Table 5–5: Characteristics of Inter–Task Spatial Correlation Histograms for Task Related Voxel

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Skewness</th>
<th>+Corr. (%) (p &lt; 0.01)</th>
<th>–Corr. (%) (p &lt; 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=10^4</td>
<td>N=10^3</td>
<td>N=10^4</td>
<td>N=10^3</td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>SZ</td>
<td>-0.08</td>
<td>-0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.12</td>
<td>0.17</td>
<td>-0.13</td>
</tr>
<tr>
<td>Novel</td>
<td>Δ</td>
<td>-0.21</td>
<td>-0.36</td>
<td>0.32</td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>SZ</td>
<td>0.12</td>
<td>0.09</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>-0.05</td>
<td>-0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>SM</td>
<td>Δ</td>
<td>0.17</td>
<td>0.11</td>
<td>-0.27</td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>SZ</td>
<td>0.16</td>
<td>0.03</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>SIRP</td>
<td>Δ</td>
<td>0.17</td>
<td>0.07</td>
<td>-0.25</td>
</tr>
<tr>
<td><strong>Novel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>SZ</td>
<td>0.03</td>
<td>0.07</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>-0.13</td>
<td>-0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>SM</td>
<td>Δ</td>
<td>0.16</td>
<td>0.10</td>
<td>-0.24</td>
</tr>
<tr>
<td><strong>Novel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>SZ</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.02</td>
<td>-0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>SIRP</td>
<td>Δ</td>
<td>-0.03</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>SM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>SZ</td>
<td>0.21</td>
<td>0.15</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.04</td>
<td>0.06</td>
<td>-0.03</td>
</tr>
<tr>
<td>SIRP</td>
<td>Δ</td>
<td>0.17</td>
<td>0.10</td>
<td>-0.26</td>
</tr>
</tbody>
</table>

Statistical measures for SZ, HC and their difference (Δ) for inter–task correlation histograms (Figure 4) are listed. The N significant voxels were found by ordering the t values found using subjects from both groups. +Corr. and –Corr. percentages were found using a correlation threshold of +0.57 and –0.57 respectively, corresponding to a significance of p=0.01 for 19 subjects.

For this purpose nineteen HCs were randomly selected from the group of 26 HCs to match the number of SZs. From the two groups of 19 subjects a subset of 10 subjects were randomly and iteratively selected for each group. The number of subjects in each group was made the same so that the variability of subsets in both groups will be similar. The reason to select 10 subjects from each group was to allow more variability to the subsets without making the subsets too small. If a larger number of subjects had been
selected, more of them will be repeated in the different subsets. We did not use the popular ‘leave–one–out’ method since it can check for the presence of only one outlier. A total of $19 \choose 10 \approx 90k$ combinations of subsets are possible for each of the groups. It is not feasible to execute such an exhaustive search with the inter–task correlation algorithm. Instead one hundred random subsets were selected for each group ensuring that no subset was repeated more than once. For faster computation the analysis was done with the 10,000 most task related voxels found through $t$ values as explained in the 'Correlation Histograms of Significant Task Related Voxels' section. For each of the subsets the $t$ values were found individually. Histograms for all 100 subsets were made and at each bin value the mean and standard error were found. Figure 5–4 shows the mean histogram of all hundred histograms with standard error. In Table 5–6 the statistical measures of the mean histogram are listed along with $p$–values. The mean and skewness of the histograms reconfirm results previously found. In the same four combinations of tasks (AOD Target + SM, AOD Target + SIRP, AOD Novel + SM and SM + SIRP) SZ have more positive correlations than HC. Again, for the AOD Target + AOD Novel combination HC have more positive correlations than SZ and the AOD Novel + SIRP combination does not show significant difference. For both groups the threshold for the percentage of correlation is 0.75 corresponding to 10 subjects. These percentages show that except for AOD Target + AOD Novel in all other combinations SZ have higher number of positive correlations and lower number of negative correlations than HC.
Figure 5–4: Mean Inter–Task Spatial Correlation Histograms Found Through Monte–Carlo Test with $10^4$ voxels

Mean and standard deviation for histograms made with different combinations of subsets. Subsets were made by randomly selecting 10 subjects each from groups of 19 patients and 19 controls. Statistical measures were found from 100 different combinations of subsets. The color bar under each histogram shows the corresponding $p$–values (in log10, values shown in the colorbar on the right) found using two sample $t$–test. The robustness of previous results is further validated by this test.

Reproduced from Michael, et al. in press

---

$^c$ Reproduced from Michael, et al. in press
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Skewness</th>
<th>% of + Corr. with $p &lt; 0.01$</th>
<th>% of –Corr. with $p &lt; 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Novel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>–0.07</td>
<td>0.12</td>
<td>0.84</td>
<td>1.93</td>
</tr>
<tr>
<td>HC</td>
<td>0.13</td>
<td>–0.18</td>
<td>2.19</td>
<td>0.32</td>
</tr>
<tr>
<td>Δ</td>
<td>–0.19</td>
<td>0.30</td>
<td>–1.35</td>
<td>1.61</td>
</tr>
<tr>
<td>$p$</td>
<td>2.4E–5</td>
<td>2.9E–5</td>
<td>4.2E–3</td>
<td>4.9E–5</td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ SM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>0.08</td>
<td>–0.12</td>
<td>1.59</td>
<td>0.53</td>
</tr>
<tr>
<td>HC</td>
<td>–0.10</td>
<td>0.18</td>
<td>0.39</td>
<td>1.50</td>
</tr>
<tr>
<td>Δ</td>
<td>0.18</td>
<td>–0.30</td>
<td>1.21</td>
<td>–0.97</td>
</tr>
<tr>
<td>$p$</td>
<td>5.1E–5</td>
<td>1.9E–4</td>
<td>1.6E–3</td>
<td>1.7E–4</td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ SIRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>0.13</td>
<td>–0.26</td>
<td>1.36</td>
<td>0.26</td>
</tr>
<tr>
<td>HC</td>
<td>–0.06</td>
<td>0.09</td>
<td>0.48</td>
<td>1.14</td>
</tr>
<tr>
<td>Δ</td>
<td>0.19</td>
<td>–0.35</td>
<td>0.88</td>
<td>–0.88</td>
</tr>
<tr>
<td>$p$</td>
<td>1.9E–6</td>
<td>1.3E–5</td>
<td>5.3E–4</td>
<td>7.7E–7</td>
</tr>
<tr>
<td><strong>Novel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ SM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>0.02</td>
<td>–0.05</td>
<td>0.83</td>
<td>0.76</td>
</tr>
<tr>
<td>HC</td>
<td>–0.08</td>
<td>0.18</td>
<td>0.42</td>
<td>1.11</td>
</tr>
<tr>
<td>Δ</td>
<td>0.10</td>
<td>–0.23</td>
<td>0.41</td>
<td>–0.34</td>
</tr>
<tr>
<td>$p$</td>
<td>2.4E–3</td>
<td>4.0E–4</td>
<td>3.7E–3</td>
<td>9.8E–2</td>
</tr>
<tr>
<td><strong>Novel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ SIRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>–0.02</td>
<td>0.01</td>
<td>0.48</td>
<td>0.71</td>
</tr>
<tr>
<td>HC</td>
<td>–0.06</td>
<td>0.11</td>
<td>0.37</td>
<td>0.90</td>
</tr>
<tr>
<td>Δ</td>
<td>0.05</td>
<td>–0.10</td>
<td>0.11</td>
<td>–0.19</td>
</tr>
<tr>
<td>$p$</td>
<td>6.2E–2</td>
<td>6.9E–2</td>
<td>1.7E–1</td>
<td>1.7E–1</td>
</tr>
<tr>
<td><strong>SM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ SIRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>0.18</td>
<td>–0.33</td>
<td>1.93</td>
<td>0.14</td>
</tr>
<tr>
<td>HC</td>
<td>0.07</td>
<td>–0.12</td>
<td>1.25</td>
<td>0.43</td>
</tr>
<tr>
<td>Δ</td>
<td>0.11</td>
<td>–0.21</td>
<td>0.68</td>
<td>–0.28</td>
</tr>
<tr>
<td>$p$</td>
<td>1.0E–4</td>
<td>3.5E–4</td>
<td>3.1E–3</td>
<td>1.1E–4</td>
</tr>
</tbody>
</table>

100 different subsets of 10 SZ and 10 HC were randomly selected from a population of 19 SZ and 19 HC. For each of the subset statistical measures for SZ, HC and their differences were found and the mean of those values are listed in the Table (Figure 5). +Corr. and –Corr. percentages were found using a correlation threshold of +0.75 and –0.75 respectively, corresponding to a significance of $p=0.01$ for 10 subjects.

The statistical significance of the result was found with $p$–values computed with a two sample $t$–test. The $p$ values are found by first converting the statistics of the two groups into a $t$ value using $(n_c + n_p – 2)$ degrees of freedom where $n_c = n_p = 10$. These values
are shown in logarithmic scale and represented by different colors in Figure 5–4. The highest $p$–values are found for the AOD Novel + SIRP combination and this is expected due to the overlapping nature of its histograms. The second highest values are found for the AOD Novel + SM combinations. All other task combinations showed very low $p$–values except for correlation values (close to 0, +1 and –1) where the histograms are overlapping.

In all of the above analyses, from the histogram plots and calculated statistical measures, we have shown that for the same four combinations of tasks (AOD Target + SM, AOD Target + SIRP, AOD Novel + SM and SM + SIRP) SZ show more positive inter–task correlations than do HC. For AOD Target + AOD Novel combination HC show more positive correlation than SZ. The AOD Novel + SIRP show inconsistent or insignificant difference.

5.6 Regions in the Brain Showing High Inter–Task Correlation

Our next goal was to locate regions in the brain that showed high inter–task correlation. In the method we introduced in Section 3.2.1, correlations of all possible combinations of voxels from two tasks are reduced to a histogram and in this process the spatial information is lost. We selected the 10,000 most significant voxels for each of the tasks found through $t$ values as explained in Section 5.4. 10,000 voxels is about 20% of the total number of brain voxels and is an arbitrary number. The reduction of voxels was to reduce computations and to investigate the correlation between task related voxels. For a particular significant voxel in Task1 (the task before the ‘→’ in Figure 5–5 and Figure
we found the fraction of voxels, out of significant voxels in Task2 (the task after the ‘→’), that had a correlation of greater than a positive threshold or less than a negative threshold. The threshold was ±0.57 for SZ and ±0.49 for HC. These thresholds correspond to a $p$ value of 0.01 for 19 subjects (SZ) and 26 subjects (HC). In Figure 5–5 and Figure 5–6 the locations of significant voxels of Task1 that have a high correlation with all significant voxels of Task2 are shown for both SZ and HC. High inter–task correlation regions are overlaid on a high resolution anatomical image (T1) that is included in SPM toolbox package. In Figure 5–5 and Figure 5–6 a scale of [0.1 to 0.3] is chosen for better representation. Regions with positive correlation are represented with shades of red and yellow and regions with negative correlations are represented with shades of blue. In Figure 5–5 and Figure 5–6 a voxel in dark red in SZ indicates that out of the correlations of this voxel in Task1 with the 10,000 most significant voxels of Task2, 10% of them have correlations higher than 0.57. A voxel in blue in HC indicates that out of the correlations of this voxel in Task1 with the 10,000 most significant voxels of Task2, 20% of them have correlations less than −0.49. Note that in previous descriptions of two task combinations, the tasks were connected by the ‘+’ sign and here we use the ‘→’ sign. While finding the inter–task correlation histograms all possible combinations of voxels from the two tasks are taken into account and the order of the two tasks does not change the final histogram of correlations. To find the locations of voxels with high inter–task correlations each voxel in Task1 is used to find the fraction of high correlation voxels in Task2. Since the fraction is found for each voxel in Task1 we use the ‘→’ to link the two tasks. Due to space limitations we do not show the alternative
result when interchanging the two tasks (which may result in same or additional brain regions with high inter–task correlations).

The five largest contiguous brain regions (sum of both left and right hemisphere volumes) of Task1 that have high fractions of significant positive and negative correlations with Task2 are listed in Table 5–7 and Table 5–8. These regions were converted from MNI space to Talairach coordinates and entered into a database (http://ric.uthscsa.edu/projects/tdc/) to provide the labels.
Figure 5–5: Brain Regions with High ($p<0.01$) Inter–Task Correlations for the First Three Task Combinations

For a significant voxel in Task1, the fraction of voxels (out of 10,000 significant voxels) in Task2 that have a correlation greater than $\rho_{th}$ or less than $-\rho_{th}$ are represented by shades of red and yellow and shades of blue respectively. $\rho_{th}$ was 0.49 for controls and 0.57 for patients. Results are shown for AOD Target $\rightarrow$ AOD Novel, AOD Target $\rightarrow$ SM and AOD Target $\rightarrow$ SIRP. Patients are on the left column and controls are on the right. Note that except for the AOD Target $\rightarrow$ AOD Novel combination patients show more positively correlated (red) regions and less negatively correlated regions (blue) than controls.

Reproduced from Michael, et al. in press

\[\text{d}\]
Table 5–7: Five Largest Brain Regions with High ($p<0.01$) Inter–Task Correlations for the First Three Task Combinations

<table>
<thead>
<tr>
<th>Task Combination</th>
<th>Region</th>
<th>R / L Volume (cc)</th>
<th>Region</th>
<th>R / L Volume (cc)</th>
</tr>
</thead>
</table>
| **AOD Target→AOD Novel**
| Positive Correlation |
| Sub–Gyral        | 0.1/0.4 |
| Extra–Nuclear    | 0.1/0.1 |
| Parahippocampal Gyrus | 0.0/0.1 |
| Cerebellar Tonsil| 0.1/0.0 |
| Declive           | 1.1/1.9 |
| Extra–Nuclear    | 0.9/1.7 |
| Lentiform Nucleus| 0.6/0.4 |
| Superior Temporal Gyrus | 0.1/0.5 |
| Thalamus          | 0.0/0.5 |
| Middle Frontal Gyrus | 1.8/0.0 |
| Thalamus          | 0.9/0.5 |
| Declive           | 1.3/0.0 |
| Inferior Frontal Gyrus | 0.9/0.0 |
| Extra–Nuclear    | 0.2/0.5 |

| Negative Correlation |
|----------------------|------------------|
| Superior Temporal Gyrus | 1.4/1.9 |
| Inferior Parietal Lobule | 0.8/0.6 |
| Insula               | 0.1/0.6 |
| Culmen               | 0.6/0.0 |
| Cerebellar Tonsil    | 0.0/0.5 |
| Extra–Nuclear        | 0.1/1.0 |
| Thalamus             | 0.0/0.7 |
| Inferior Frontal Gyrus | 0.0/0.3 |
| Inferior Parietal Lobule | 0.0/0.2 |
| Sub–Gyral            | 0.0/0.2 |

| **AOD Target→SM**
| Positive Correlation |
|----------------------|------------------|
| Inferior Parietal Lobule | 3.4/1.0 |
| Superior Temporal Gyrus | 1.9/0.9 |
| Postcentral Gyrus     | 1.1/0.1 |
| Supramarginal Gyrus   | 0.5/0.0 |
| Sub–Gyral             | 0.3/0.2 |
| Declive               | 0.2/0.0 |
| Extra–Nuclear         | 0.1/0.0 |
| Thalamus              | 0.1/0.0 |

For the first three task combinations the labels of the five largest contiguous brain regions in each group with high correlations (positive and negative) in Figure 6 are listed along with their volumes. A blank cell in the Table indicates the absence of such a region.
Figure 5–6: Brain Regions with High ($p<0.01$) Inter–Task Correlations for the Last Three Task Combinations

For a significant voxel in Task1, the fraction of voxels (out of 10,000 significant voxels) in Task2 that have a correlation greater than $\rho_{th}$ or less than $-\rho_{th}$ are represented by shades of red and yellow and shades of blue respectively. $\rho_{th}$ is 0.49 for controls and 0.57 for patients. Results are shown for AOD Novel $\rightarrow$ SM, AOD Novel $\rightarrow$ SIRP and SM $\rightarrow$ SIRP. Patients are on the left column and controls are on the right. Note that except for the AOD Novel $\rightarrow$ SIRP combination patients show more positively correlated (red) regions and less negatively correlated regions (blue) than controls.

* Reproduced from Michael, et al. in press
Table 5–8: Five Largest Brain Regions with High ($p<0.01$) Inter–Task Correlations for the Last Three Task Combinations

<table>
<thead>
<tr>
<th>Task Combination</th>
<th>Region</th>
<th>R / L Volume (cc)</th>
<th>Region</th>
<th>R / L Volume (cc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive correlation</td>
<td>Inferior Parietal Lobule</td>
<td>0.9/0.4</td>
<td>Cuneus</td>
<td>1.4/2.0</td>
</tr>
<tr>
<td></td>
<td>Middle Frontal Gyrus</td>
<td>0.8/0.0</td>
<td>Lingual Gyrus</td>
<td>0.8/1.1</td>
</tr>
<tr>
<td></td>
<td>Inferior Frontal Gyrus</td>
<td>0.6/0.0</td>
<td>Inferior Frontal Gyrus</td>
<td>1.4/0.2</td>
</tr>
<tr>
<td></td>
<td>Superior Temporal Gyrus</td>
<td>0.5/0.0</td>
<td>Sub–Gyral</td>
<td>0.3/0.8</td>
</tr>
<tr>
<td></td>
<td>Postcentral Gyrus</td>
<td>0.5/0.0</td>
<td>Middle Frontal Gyrus</td>
<td>0.8/0.0</td>
</tr>
<tr>
<td>negative correlation</td>
<td>Sub–Gyral</td>
<td>0.0/0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fusiform Gyrus</td>
<td>0.0/0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inferior Frontal Gyrus</td>
<td>0.0/0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive correlation</td>
<td>Middle Frontal Gyrus</td>
<td>0.8/0.0</td>
<td>Middle Frontal Gyrus</td>
<td>1.1/0.0</td>
</tr>
<tr>
<td></td>
<td>Insula</td>
<td>0.2/0.0</td>
<td>Inferior Frontal Gyrus</td>
<td>0.6/0.0</td>
</tr>
<tr>
<td></td>
<td>Precentral Gyrus</td>
<td>0.1/0.0</td>
<td>Superior Frontal Gyrus</td>
<td>0.1/0.0</td>
</tr>
<tr>
<td></td>
<td>Superior Temporal Gyrus</td>
<td>0.1/0.0</td>
<td>Superior Temporal Gyrus</td>
<td>0.1/0.0</td>
</tr>
<tr>
<td></td>
<td>Postcentral Gyrus</td>
<td>0.1/0.0</td>
<td>Extra–Nuclear</td>
<td>0.1/0.0</td>
</tr>
<tr>
<td>negative correlation</td>
<td>Middle Temporal Gyrus</td>
<td>1.1/0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Superior Temporal Gyrus</td>
<td>0.1/0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sub–Gyral</td>
<td>0.1/0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive correlation</td>
<td>Precentral Gyrus</td>
<td>0.8/2.0</td>
<td>Precentral Gyrus</td>
<td>0.4/0.8</td>
</tr>
<tr>
<td></td>
<td>Declive</td>
<td>2.2/0.5</td>
<td>Postcentral Gyrus</td>
<td>0.0/0.9</td>
</tr>
<tr>
<td></td>
<td>Superior Temporal Gyrus</td>
<td>1.6/0.7</td>
<td>Declive</td>
<td>0.8/0.0</td>
</tr>
<tr>
<td></td>
<td>Postcentral Gyrus</td>
<td>0.4/1.4</td>
<td>Middle Frontal Gyrus</td>
<td>0.3/0.0</td>
</tr>
<tr>
<td></td>
<td>Insula</td>
<td>0.2/1.4</td>
<td>Culmen</td>
<td>0.3/0.0</td>
</tr>
</tbody>
</table>

For the last three task combinations the labels of the five largest contiguous brain regions with high correlations (positive and negative) in Figure 7 are listed along with their volumes. A blank cell in the Table indicates the absence of such a region.

For AOD Target $\rightarrow$ SM, AOD Target $\rightarrow$ SIRP, AOD Novel $\rightarrow$ SM and SM $\rightarrow$ SIRP combinations, HC shows no or smaller regions of positive correlations (regions in red and yellow) than SZ. For AOD Target $\rightarrow$ SM, AOD Target $\rightarrow$ SIRP and AOD Novel...
SM combinations no or smaller regions of negative correlations (regions in blue) are seen in SZ than HC and the SM→SIRP combination does not show any regions of negative correlations in both SZ and HC. In the AOD Target→AOD Novel combination HC show larger regions of positive correlations than SZ and SZ show regions of negative correlations while HC does not show any. Positive correlation regions for the AOD Novel → SIRP combination show regions of comparable sizes in SZ and HC and SZ show slightly larger negatively correlated region than HC.

For the AOD Target → AOD Novel combination, even though, both SZ and HC show regions with positive correlations the regions are not the same. In AOD Target → SM, AOD Target → SIRP and AOD Novel → SM where patients show regions of positive correlations the superior temporal gyrus and the inferior parietal lobule are repeated. In the SM→ SIRP combination where both SZ and HC show regions of positive correlations the precentral gyrus, postcentral gyrus and declive are present in both groups. In the AOD Target→ AOD Novel combination where SZ show regions of negative correlations subcortical and cerebellar regions are present. In HC where regions of negative correlations are present frontal gyrus, thalamus and extra nuclear regions are repeated. From Figure 5–5 and Figure 5–6, based on the size of positively and negatively correlated regions, the results that, for the AOD Target + SM, AOD Target + SIRP, AOD Novel + SM, and SM + SIRP tasks, SZ show more positive correlations than HC and for the AOD Target + AOD Novel task vice versa, is reconfirmed.
5.7 Discussion

Consistent with our hypothesis we found four task combinations (AOD Target + SM, AOD Target + SIRP, AOD Novel + SM and SM + SIRP) where SZ were performing different tasks in a more positively correlated manner than did HC. These combinations included the AOD Target + SIRP combination for which Calhoun et al. (2006) found the same result with a different data set and method. Contrary to our hypothesis, in one task combination (AOD Target + AOD Novel) HC showed more positive correlation than did SZ and in another combination (AOD Novel + SIRP) the difference in correlation was insignificant or inconsistent over the different tests. In the four task combinations where SZ histograms were to the right of HC histograms, the skewness showed that they were more tilted towards positive correlations, the percentages of correlations above a positive threshold were higher in SZ than HC and the percentages of correlations less than a negative threshold were higher in HC than SZ. In one combination these results were reversed. These results were confirmed in different ways: a random selection of HCs to match the number of SZs, a check with equal number of males and females in each group, analyses with significant task related voxels and a Monte–Carlo method to remove the possibility of the result occurring due to outliers. A few exceptions were noted in the test for equal number of males and females and the test with \( N = 10^3 \) significant voxels.

In combinations where SZ shows more positive inter–task correlation the tasks occurred at different scanning sessions. From the mean of the histograms a clear statement about the strength (either positive or negative) of inter–task correlation is not easy to make since this varied for the different tests. For example for the AOD Target + SM task in Table 5–6 (Monte Carlo test) the absolute value of the mean is higher in HC.
than SZ. But in all other tests for this task combination the absolute value of the mean is higher in SZ than HC. In this work we do not make the claim that the strength of inter–task correlation is higher in SZ than HC but a clear result in all tests is that SZ show a more positive correlation than HC for four task combinations and in one the opposite. Results from the percentage of significant correlations imply that in four task combinations SZ show a more common pattern of activation among tasks than HC. This result implies that the manner in which SZ perform different or specialized tasks is more similar or less unique than HC. A breakdown of specialized wiring between cognitive domains may be a possible cause. It can be hypothesized that due to the break down of different networks patients were activating similar networks for different tasks. The only combination where patients show less positive inter–task correlation than controls is the AOD Target + AOD Novel. It should be noted that data for this combination were collected within the same session of scanning. This may be the reason why for this combination SZ were having a less positive inter–task correlation than HC. An inefficient resource allocation process or the inability to efficiently switch from one task to another in schizophrenic patients may be causing this unpredicted result.

The above results are further validated by brain regions that showed significant inter–task correlations. In Figure 5–5 and Figure 5–6, except for the first row (AOD Target + AOD Novel), SZ show no regions of negative correlations (areas in blue) and HC has either no or smaller regions with positive correlations (areas in red). In Figure 5–6 and Table 5–8, except for the second row (AOD Novel + SIRP), SZ show no or smaller regions of negative correlations and HC has no or smaller regions with positive
correlations. For SM→SIRP combinations neither group show regions of positive correlations.

For the AOD Target→AOD Novel combination both groups show subcortical and cerebellar regions with significant inter–task correlations. Here extra–nuclear shows both positive and negative correlations in SZ and just positive correlations in HC. Declive and thalamus shows positive correlations in HC but show negative correlations in SZ. The theory of cognitive dysmetria (Andreasen, et al. 1998) schizophrenia is modeled as a dysfunction in cortical–subcortical–cerebellar circuitry. Our results show group differences in subcortical and cerebellar regions and may support the theory of cognitive dysmetria.

The superior temporal gyrus show positive correlations in SZ for five task combinations and in HC for one task combination (AOD Novel→SIRP) where other tests do not show significant group difference. Three out of the four tasks had an auditory input and hence in all of the task combinations the temporal lobe regions are involved. This is a possible explanation of the repeated presence of the superior temporal gyrus in SZ. But the superior temporal gyrus does not show inter–task correlation in the AOD Target + AOD Novel combination even though they were both auditory tasks. This shows that the regions shown as high inter–task correlations are not merely due to the activation but due to the fact that they are task correlated. The inferior parietal lobule shows positive correlations for three task combinations in SZ and negative correlation in HC for one task combination. The inferior frontal gyrus or the middle frontal gyrus shows either positive or negative correlation regions for four task combinations in HC and in SZ the inferior frontal gyrus is present in only AOD Target→AOD Novel.
It is not straightforward to extend our approach to combine more than two tasks since correlation for more than two variables is not established. However, our simple approach can fuse \( n \) different tasks from the activation maps to form histograms in the \( n \) or less dimensional space. For example three tasks (T1, T2 and T3) can be combined to compute a three dimensional histogram with correlations between T1 and T2 as the x–axis, correlations between T2 and T3 as the y–axis and frequency as the z–axis. Four tasks can be combined as in the previous example in a four dimensional space or in a three dimensional space with T1 and T2 as one axis and T3 and T4 as the other. The histograms we have obtained in our work can be considered the marginal histograms of higher dimensional histograms. Higher dimensional histograms may detect group differences that marginal histograms cannot detect.

The data were preprocessed using the well established SPM5 software to produce the activation maps. One of the drawbacks of using activation maps made with regressors is that they are biased or forced to pick voxels that show activity similar to the regressors. This assumption can be relaxed by using data–driven approaches such as independent component analysis (ICA) (Calhoun, et al. 2001). In the method introduced in this paper, what is computed is the variation in task specific activation across different subjects of a certain group correlated for two different tasks. The analysis examines if the degree of activation between voxels of two different tasks are correlated or not across subjects to find differences between two groups. Since the time domain is collapsed at the first level analysis, while generating the activation maps, no analyses in the time domain, such as standard approaches for functional or effective connectivity, are performed here. Studies of that nature are primarily aimed at understanding the connectivity within brain regions
while performing a certain task and not necessarily to understand correlations between
tasks. It should be noted that two voxels (from different tasks) that show correlated
contrast estimates across subjects may be uncorrelated in the temporal domain within
each of the subjects. Correlating two voxels from two different tasks in the temporal
domain to examine inter or intra–subject variability is not trivial in our case since the
tasks used in our study consist of multiple stimuli presented in random order. The random
presentation of stimuli also contributes towards inter and intra–subject variability that
may or may not have existed if the correlations had been found in the time domain across
different subjects.

It is possible that higher correlations could be introduced due to head motion
while correlating voxels within a single brain. However group differences in head
movements do not appear to contribute to the results obtained here since our correlation
analysis is across subjects and between different tasks. We evaluate this directly by
correlating motion parameters over time and as expected our results showed that motion
across subjects is uncorrelated. In addition, if a higher degree of head motion in patients
did contribute to our results it is reasonable to expect that its contribution will be higher
when stimuli within the same task are correlated than when two different tasks are
correlated. In the AOD Target + AOD Novel task combination (stimuli within the same
task), we observe that patients do not show a more positive correlation than controls but
the opposite.
5.8 Classification

For classification of SZ from HC we apply the histogram shift method introduced in Section 3.6. In Figure 5–7 (a) and (b) we present the results of $A_{sz}$ and $A_{hc}$ ($N = 1000$ voxels) found when an HC was left out one at a time. The matrices are shown as images where the x-axis represents the left–out subject ($S_i$, $i = 1$ to $n_c$) and the y-axis the six different fMRI task combinations. Blue corresponds to +1, a cue for an HC and red to −1, a cue for a SZ.

Figure 5–7: Classification cues for HC subjects

A cue for HC is represented by blue and for SZ by red. The cues were obtained from different task combinations while testing for the different control subjects (x-axis). Cues obtained from SZ histogram shift is shown in (a) and cues obtained HC histogram shifts in (b).

---

Reproduced from Michael, et al. 2008a
Figure 5–8: Classification cues for SZ subjects\textsuperscript{g}

A cue for HC is represented by blue and for SZ by red. The cues were obtained from different task combinations while testing for the different SZ subjects (x–axis). Cues obtained from HC histogram shift is shown in (a) and cues obtained SZ histogram shifts in (b).

Figure 5–8 carries the same information as Figure 5–7, but here the left–out subject $S_i$ is taken from the SZ group ($i = 1$ to $n_{SZ}$). Figure 5–7(a) and (b) correspond to $A_{mn}^{HC}$ and $A_{mn}^{SZ}$ respectively. Here again red corresponds to a cue for SZ and blue HC.

Note that, in Figure 5–7 and Figure 5–8, a higher number of correct cues are obtained when $S_i$ is placed in the ‘wrong’ group. This is due to the fact that, the shift in histogram is higher when an alien (incorrect) subject is combined to a group.

In Table 5–9 the classification rates are listed (HC in blue and SZ in red) as percentages based on cues obtained from each of the six task combinations separately and the summation of cues from all six task combinations. The percentages are found with respect to the number of subjects in each group, since the testing is performed by leaving out each subject one–at–a–time. The percentage of ‘correct’ classification was low for all

\textsuperscript{g} Reproduced from Michael, et al. 2008a
individual task combinations. High ‘unsure’ rates in individual task combinations indicate that there was not sufficient information to categorize most of the subjects as a HC or SZ. The last row of Table 5–9 shows the classification rates using the summation ($F$) of cues (see step 9 of Section 3.6) from all combinations used to classify subjects. Compared to using cues from individual task combinations, using the summation of cues from all task combinations, results in a significantly higher correct classification with 73.1% for HC and 78.9% for SZ.

### Table 5–9: Classification Results

<table>
<thead>
<tr>
<th>Task Combination</th>
<th>Correct</th>
<th>Incorrect</th>
<th>Unsure</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOD Target + AOD Novel</td>
<td>50.0</td>
<td>3.8</td>
<td>46.1</td>
</tr>
<tr>
<td>AOD Target + SM</td>
<td>19.2</td>
<td>15.3</td>
<td>65.3</td>
</tr>
<tr>
<td>AOD Target + SIRP</td>
<td>34.6</td>
<td>7.7</td>
<td>57.7</td>
</tr>
<tr>
<td>AOD Novel + SM</td>
<td>23.1</td>
<td>3.8</td>
<td>73.1</td>
</tr>
<tr>
<td>AOD Novel + SIRP</td>
<td>23.1</td>
<td>15.4</td>
<td>61.5</td>
</tr>
<tr>
<td>SM + SIRP</td>
<td>30.8</td>
<td>3.8</td>
<td>65.4</td>
</tr>
<tr>
<td>All</td>
<td><strong>73.1</strong></td>
<td><strong>15.4</strong></td>
<td><strong>11.5</strong></td>
</tr>
<tr>
<td>All</td>
<td><strong>78.9</strong></td>
<td><strong>10.5</strong></td>
<td><strong>10.5</strong></td>
</tr>
</tbody>
</table>

Classification results for HC is in blue and SZ in red. Classification results are calculated as a percentage of correctly classified subjects to the total number of subjects in a particular group.
In Figure 5–9 the $F$ values are shown in the y–axis while trying to classify the different subjects shown in the x–axis. The 19 patients are shown in red circles and the 26 controls are shown in blue dots. The decision boundary is shown by the black line.

![Figure 5–9: The summation of cues ($F$) obtained from different task combinations](image)

**Figure 5–9: The summation of cues ($F$) obtained from different task combinations**

Cues from individual task combinations are added to obtain an overall cue. Cue for the different HC ($n_c=26$) and SZ ($n_p=19$) are indicated with blue dots and red circles respectively. $F = 0$ is the decision boundary.

---

*Reproduced from Michael, et al. 2008a*
6 DTI – Symptom Scores Analysis

Disturbances in white matter (WM) connectivity between different brain regions, is attributed as a possible cause for schizophrenia and is termed the ‘disconnection hypothesis’ (Friston 1998). Studies, both in vivo and post–mortem, have shown WM volume, fiber number or density differences between patients with schizophrenia and healthy controls possibly due to abnormalities in the myelin sheaths around the axons (Foong, et al. 2002; Minami, et al. 2003). Diffusion tensor imaging (DTI, see Section 2.4) is used to capture WM tract locations and its properties. In this chapter we investigate if the severity of symptoms experienced by a patient with schizophrenia is correlated to white matter properties measured through DTI. WM properties are measured through four scalar quantities of DTI: fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD) (see Section 2.4). The symptom scores are evaluated using psychometric tests and are clinically used in the diagnosis process. Positive and negative syndrome scale (PANSS) is one such evaluation scheme that measures symptoms of schizophrenia in three dimensions: positive, negative and general. The goal of this analysis is to find potential image based biomarkers of schizophrenia that are correlated to behavioral measures that are currently used to diagnose schizophrenia.
Previous studies (Shin, et al. 2006; Skelly, et al. 2008) have found correlations between DTI parameters and PANSS scores. Skelly and collaborators (Skelly, et al. 2008) found WM deficits in SZ and showed that FA values were negatively correlated to positive PANSS scores. Shin and collaborators used apparent diffusion coefficients (ADC) to show that regions in the right insula were correlated with negative PANSS scores. In another study, it was found that SZ show WM reductions in frontal lobes and that negative symptom scores are negatively related to WM volumes in the cingulate and in the right internal capsule (Paillère-Martinot, et al. 2001). Skelly, et al. (2008) found WM deficits in SZ and showed that DTI FA values are negatively correlated to positive PANSS scores. In all of the above studies, correlations were found on a voxel by voxel basis with a maximum number of 25 subjects. The correlations were found using either the FA or ADC with some PANSS scores.

In this study, we used 41 SZ subjects to correlate all four different DTI values with all three PANSS scores. Note that a correlation analysis of this nature cannot be performed using HC data, since PANSS scores are not recorded for HC. Previous studies analyze FA values alone to investigate WM disruptions. However, AD, MD and RD can hold valuable information that FA cannot reveal. For example, low FA values can occur due to interruptions in the axons or defects in the myelin sheath around axons. A low AD value would indicate that diffusion along the axon is impeded and may be due to axonal damage. A high RD value indicates that diffusion along the perpendicular directions of the axon is high and that it may be due to damages in the myelin sheath. By analyzing all four DTI measures, additional information about the nature of diffusion can be provided to the analyses algorithms.
First we investigate the correlations between WM regions segmented through two different atlases. Then DTI values that have significant correlations with the PANSS scores are combined through multiple regression to predict the PANSS scores to find regions in the brain that have high correlation. This is done on a voxel by voxel basis. The above analysis of multiple regression has to be done on each PANSS scores separately.

We use canonical correlation analysis (CCA) to combine all four DTI measures and all three PANSS scores to select voxels with high DTI–PANSS correlations (see Section 3.4). An overview of the DTI–PANSS analysis is given in CCA was applied to patients’ WM skeletons derived using tract based spatial statistics (TBSS, see Section 3.5). The TBSS step is used to reduce data as well as to address subject variability more appropriately. We then use voxels with high DTI – PANSS CCA correlations as features to classify SZ and HC subjects. For classification we used support vector machines (see Section3.7) and discriminant analysis (see Section 3.8). Most of the discussion and results presented here are from Michael, et al. (2008b), Michael, et al. (2009d) and Michael, et al. (2009e).
6.1 Subjects

The subjects in this study were scanned at the Olin Neuropsychiatry Research Center, Institute of Living, Connecticut. All subjects provided written informed consent to participate after the procedures were explained to them. These procedures were approved by Yale University and Hartford hospital institutional review boards. DTI data from 44 patients with chronic schizophrenia (SZ) and 44 matched healthy controls (HC) were collected at the Olin Neuropsychiatry Research Center. SZ were diagnosed using the structured clinical interview for DSM–IV (SCID) and review of the case file. Out of the SZ, 25 were chronic patients, 14 had relapse, and 6 were first–break. Among the SZ, 10 were taking first–generation antipsychotics, 31 were taking second–generation antipsychotics, 8 were taking mood stabilizers or antidepressants, and 5 were taking no
medications or the information was not recorded. More demographics of the subjects are listed in Table 6–1.

Table 6–1: Demographics and Clinical Characteristics of Patients with Schizophrenia (SZ) and Healthy Controls (HC)

<table>
<thead>
<tr>
<th></th>
<th>SZ ( n_p = 44 )</th>
<th>HC ( n_c = 44 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>39 ± 12</td>
<td>38 ± 12</td>
</tr>
<tr>
<td></td>
<td>(range: 19–59)</td>
<td>(range: 19–60)</td>
</tr>
<tr>
<td><strong>Male, Female</strong></td>
<td>22 males</td>
<td>18 males</td>
</tr>
<tr>
<td></td>
<td>12 females</td>
<td>26 females</td>
</tr>
<tr>
<td><strong>Handedness</strong></td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>(non–right hand)</td>
<td></td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
<td>positive = 15.6 ± 5.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>negative = 15.7 ± 6.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>disorganization = 31.8 ± 9.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

6.2 Correlation with Atlas Regions

The WM of the whole brain was segmented into 20 regions defined by the Johns Hopkins University (JHU) atlas (Mori, et al. 2008), These regions are pairs of the following 10 regions where the pairs correspond to the left and right hemispheres of the brain: Anterior thalamic radiation (ATR), Corticospinal tract (CST), Cingulum – cingulate gyrus (CNGg), Cingulum – hippocampus (CNGh), Forceps major and minor (FMaM), Inferior fronto–occipital fasciculus (IFOF), Inferior longitudinal fasciculus (ILF), Superior longitudinal fasciculus (SLF), Uncinate faciculus (UNF), Superior longitudinal fasciculus – temporal part (SLFt). Special attention was given to corpus
callosum (CC), the largest WM structure that connects the left and right cerebral hemispheres. The CC was further segmented into 5 sub regions according to the CC atlas (Hofer and Frahm 2006) as shown in Figure 6–2. The CC was segmented in the following manner: Region 1 (CC1): prefrontal; Region 2 (CC2): premotor and supplementary motor; Region 3 (CC3): motor; Region 4 (CC4): sensory; Region 5 (CC5): parietal, temporal, and occipital.

Figure 6–2: The corpus callosum segmented into five regions according to an existing atlas
The corpus callosum is the largest nerve bundle in the brain that connects the left and right hemispheres of the brain. Region 1: prefrontal; Region 2: premotor and supplementary motor; Region 3: motor; Region 4: sensory; Region 5: parietal, temporal, and occipital.

The regional correlation to PANSS scores can be found in two different ways. The first is to find correlations for each of the voxels within a region–of–interest (ROI) and then find the mean correlation (mean of correlations). The second is to correlate the

---

* Reproduced from Michael, et al. 2008b
mean DTI value of an ROI (correlation of means). It can be shown mathematically that the two methods are not equivalent if the correlations are not homogenous across the ROI. The first method is computationally more intense, but is better since it allows us to analyze the spatial variation of correlation across the ROI. We show that for our data sets the two methods give similar results as presented in preceding sections.

6.2.1 Mean of Correlation Values

Correlations across subjects for each voxel within an atlas region were found with the 3 different PANSS scores separately. Then the mean and the standard deviation of all correlations within a region were found. In Figure 6–3 the correlations are shown as an image for all 4 DTI values and the 3 test scores. The DTI values are presented in columns and the PANSS scores in sub columns. The correlations are represented as colors according to the values indicated in the adjacent color bar.

A clear result is that FA values of almost all regions are negatively correlated with all 3 scores. This implies that in schizophrenia the scores are positively correlated with the degree of WM disruptions. This result supports the ‘disconnection hypothesis’ and has been previously reported in other studies (Kubicki, et al. 2007). MD and AD, for most of the regions, shows positive correlations with the scores and RD does not show strong correlations for any of the regions.
For each brain voxel, the correlation between the four DTI values and the three PANSS scores is found. The mean of the correlations that falls within an anatomical segment defined by the twenty JHU atlas regions and the five segments of the corpus callosum is represented with a color. Each JHU atlas region row has two colors, top and bottom corresponding to the left and right hemispheres.

6.2.2 Correlation of Mean Values

In Figure 6–4 we present the correlation of mean DTI values of atlas regions correlated with the scores. The correlation values found in this method show results similar to results of Section 6.2.1. The FA of different regions shows negative correlations and MD, AD and RD regions show positive correlations with the PANSS scores. Again, compared to other DTI values, RD does not show strong correlations with the PANSS scores. Correlation values obtained with this method are slightly stronger than that of Section 6.2.1.

---

\[ \text{Reproduced from Michael, et al. 2008b} \]
The mean values of the four DTI values were found for the twenty different anatomical regions defined by the JHU atlas and the five segments of corpus callosum. For each regional mean DTI value its correlation to the PANSS scores are found and represented as a color. Each JHU atlas region row has two colors, top and bottom corresponding to the left and right hemispheres.

6.2.3 Atlas Regions with Significant Correlation

The WM atlas regions that have the highest (positive/ negative) correlation value for each of the DTI values are listed in. An absolute correlation value of greater than 0.31 is significant at \( p \)-value less than 0.05 with 41 as the number of subjects. The significance of the correlation was found using Equation (3.2). The values listed in blue correspond to mean of correlations along with standard deviation as explained in Section 6.2.1. Values in red correspond to the correlation of mean values as explained in Section 6.2.2. The atlas regions with the highest correlations are repeated for different DTI value and PANSS score combinations. The same atlas regions are found by the two different

---

\(^{\text{c}}\) Reproduced from Michael, et al. 2008b
methods for most DTI–PANSS combinations. SLFt Right and ILF Left and all five segments of the corpus callosum show the highest correlations with the scores as listed in Table 6–2.

<table>
<thead>
<tr>
<th></th>
<th>Positive score</th>
<th>Negative score</th>
<th>General score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SLFt(R) −0.28 ± 0.06</td>
<td>CC5 −0.34 ± 0.08</td>
<td>CC1 −0.26 ± 0.04</td>
</tr>
<tr>
<td></td>
<td>SLFt(R) −0.33</td>
<td>CC5 −0.40</td>
<td>CC5 −0.29</td>
</tr>
<tr>
<td><strong>MD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC1 0.25 ± 0.06</td>
<td>SLFt(R) 0.20 ± 0.09</td>
<td>CC1 0.17 ± 0.07</td>
</tr>
<tr>
<td></td>
<td>ILF(L) 0.37</td>
<td>SLF(R) 0.27</td>
<td>ILF(L) 0.24</td>
</tr>
<tr>
<td><strong>AD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC2 0.21 ± 0.07</td>
<td>SLFt(R) 0.24 ± 0.10</td>
<td>CC4 −0.17 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>ILF(L) 0.33</td>
<td>SLF(R) 0.34</td>
<td>CC4 −0.22</td>
</tr>
<tr>
<td><strong>RD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC3 0.16 ± 0.01</td>
<td>CC5 0.12 ± 0.09</td>
<td>CC5 0.11 ± 0.11</td>
</tr>
<tr>
<td></td>
<td>CC3 0.16</td>
<td>CC5 0.15</td>
<td>CC5 0.13</td>
</tr>
</tbody>
</table>

Atlas regions with the most significant (highest/lowest) mean value found using the mean of correlation method is listed in blue along with the standard deviation. The same calculated using the correlation mean DTI is given in blue. SLFt(R): superior longitudinal fasciculus (temporal part) right, ILF(L): inferior longitudinal fasciculus left and CC: corpus callosum.

6.3 Multiple Regression for Each Voxel

From the correlations of atlas regions we find that FA and MD have higher correlation values with the PANSS scores than the other DTI values. We are interested in studying how well these DTI measures explain the three PANSS scores. We use multiple linear regression (MLR, see Section 3.3) as shown in Equation (6.1).
\[ \beta_0 + \beta_1 FA_k + \beta_2 MD_k = Y^i_k \] (6.1)

where \( k = 1, \ldots, N \) and \( N = 41 \), the number of subjects and \( Y^i \) is the vector of scores, \( i = 1, 2, 3 \) corresponding to the +ve, –ve and gen scores. MLR is done for each WM voxel separately. For each voxel the \( r^2 \) value, the coefficient of multiple determination (see Section 3.3), and the \( p \)–value of its significance are calculated.

The voxels that had \( p \) values less than 0.05 for positive, negative and general scores are presented in Figure 6–5(a), (b) and (c) respectively. The significantly correlated voxels are shown with shades ranging from orange to red. Orange indicates a \( p \)–value of 0.05 and red a \( p \) value close to zero as indicated by the color bar in Figure 6–5. The \( p \)–values were overlaid on WM tracts and slices shown have an inter–slice gap of 6mm. The following atlas regions have the highest number of voxels (number of voxels in parenthesis) for the positive score: SLF Left (909), SLF Right (671), ILF Left (551), SLF Left (581), ILF Left (373) and CST Left (303) have the highest number of voxels for the negative score. SLF Left (518), ILF Left (354) and IFOF Left (336) were the regions for the general score. The atlas regions are indicated with purple, yellow, green, and blue colors. The CC regions do not show up in this ranking scheme since the total number of voxels within CC regions is much smaller compared to other WM atlas regions.
6.4 Application of Canonical Correlation Analysis

In Sections 6.2 and 6.3, the correlations were computed for each of the PANSS scores separately. Our next goal was to find brain voxels with significant DTI–PANSS correlations with a multivariate approach. We use canonical correlation analysis (CCA) that was introduced in Section 3.4. CCA is applied to find relationships between two sets of multi-dimensional variables, DTI measures of WM integrity and concomitantly measured PANSS scores for SZ. AD, RD, MD and FA were used as DTI measures and positive, negative and general scores were used as measures from PANSS. CCA seeks to find a pair of linear transformations to maximize correlation between the four DTI measures and the three PANSS summary scores. Another difference of the application of CCA from Sections 6.2 and 6.3 is that here we select voxels that lie at the center of WM tracts and not over regional WM locations. The center tracts or the skeletons are
computed with tract–based spatial statistics (TBSS). In Section 3.5 we introduced TBSS and listed its advantages. MD, AD and RD skeletons for the different subjects were calculated from the mean skeleton of FA values of all subjects. The skeletons, of all subjects and for all four DTI measures, were spatially identical. The DTI value at the skeleton locations varied from subject to subject.

The CCA correlation value \( r \) corresponding to significant \((p<0.05)\) DTI – PANSS scores was calculated using the Roy’s largest root statistic test (See section 3.4). In this case of CCA application \( N = 41, N_1 = 4 \) and \( N_2 = 3 \). With these values it is found that a correlation value of \(|r|>0.6\) corresponds to a significance of \( p<0.05 \). Voxels on the skeleton with DTI – PANSS CCA correlation of \(|r|>0.6\) were identified and the number voxels on the skeleton and within a JHU atlas region are listed in Table 6-3. The left anterior thalamic radiation (ATR–L), splenium of the corpus callosum (SCC) and right superior longitudinal fasciculus (SLF–R) are the atlas regions with the largest number of significantly correlated voxels.

In Figure 6–6 the atlas regions are shown with descending brightness of red, shades sorted by the ranking of total number of significant voxels. In the atlas definition SCC and forceps major (FMAJ) are overlapping and we show FMAJ. The locations of the skeleton and the significantly correlated voxels are shown by orange and white (some circled) respectively.
Table 6-3: JHU atlas regions with the largest number of significant DTI–PANSS CCA correlation

<table>
<thead>
<tr>
<th>White matter tract</th>
<th>Number of significant voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior thalamic radiation left (ATR–L)</td>
<td>187</td>
</tr>
<tr>
<td>Superior long. fasciculus right (SLF–R)</td>
<td>158</td>
</tr>
<tr>
<td>Splenium of corpus callosum (SCC)</td>
<td>158</td>
</tr>
<tr>
<td>Forceps major (FMAJ)</td>
<td>132</td>
</tr>
<tr>
<td>Forceps minor (FMIN)</td>
<td>116</td>
</tr>
<tr>
<td>Anterior thalamic radiation right</td>
<td>111</td>
</tr>
<tr>
<td>Inferior fronto–occipital fasciculus right</td>
<td>107</td>
</tr>
<tr>
<td>Superior longitudinal fasciculus left</td>
<td>102</td>
</tr>
</tbody>
</table>

In this section we investigated the correlation between DTI and PANSS values using CCA, a multivariate analytic approach. We found correlation locations on a skeleton found with TBSS which accounts for spatial variability of WM tracts across different subjects. We report regions with significant correlation between DTI measures and PANSS scores. The finding in the superior longitudinal fasciculus related to symptom severity is consistent with findings of Section 6.3 and two other independent recent studies. (Seok, et al. 2007; Skelly, et al. 2008). Our study as a whole supports the “disconnection” hypothesis of schizophrenia.
Figure 6–6: JHU atlas regions with the largest number of significantly ($p<0.05$) DTI–PANSS correlations computed through CCA.
JHU atlas regions are presented in shades of red, brightest for the region with highest number of significant voxels. The skeletons are indicated in orange color. The location of significant voxels are shown in white (some circled)

6.5 Classification

In this section we present classification results using the algorithm presented in Section 3.9. We use support vector machine (SVM, see Section 3.7) and discriminant analysis (DA, see Section 3.8) as classifiers. We use SVM with linear or radial basis kernels and for DA we use four different discriminant boundary locating methods as explained in Sections 3.8.1 to Section 3.8.4. To classify using DA ‘classify’ function in the statistics toolbox of MATLAB was used. We used ‘LBSVM’ (http://www.csie.ntu.edu.tw/~cjlin/libsvm) software toolbox to perform classification based on SVM. The classification results are presented in Table 6–4.
Table 6–4: Classification success percentages with the leave–one–out scheme with different classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>CCA Correlation Threshold</th>
<th>$r = 0.4$</th>
<th>$r = 0.5$</th>
<th>$r = 0.6$</th>
<th>$r = 0.7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM (linear kernel)</td>
<td>HC</td>
<td>60</td>
<td>58</td>
<td>55</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>SZ</td>
<td>63</td>
<td>58</td>
<td>63</td>
<td>55</td>
</tr>
<tr>
<td>SVM (RBF kernel)</td>
<td>HC</td>
<td>70</td>
<td>75</td>
<td>70</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>SZ</td>
<td>48</td>
<td>55</td>
<td>58</td>
<td>55</td>
</tr>
<tr>
<td>DA (Case 1)</td>
<td>HC</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>SZ</td>
<td>60</td>
<td>63</td>
<td>63</td>
<td>60</td>
</tr>
<tr>
<td>DA (Case 2)</td>
<td>HC</td>
<td>60</td>
<td>63</td>
<td>60</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>SZ</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>50</td>
</tr>
<tr>
<td>DA (Case 3)</td>
<td>HC</td>
<td>68</td>
<td>68</td>
<td>70</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>SZ</td>
<td>58</td>
<td>58</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>DA (Case 4)</td>
<td>HC</td>
<td>50</td>
<td>55</td>
<td>55</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>SZ</td>
<td>53</td>
<td>45</td>
<td>58</td>
<td>60</td>
</tr>
</tbody>
</table>

The classification percentages were computed as a percentage of correctly classified subjects out of the total number of subjects (44) in each group. The HC classification success is presented in blue and SZ in red. The number of voxels selected as features is reduced by increasing the CCA thresholds.

The classification success rates were lower for SZ than HC for most of the classification methods and CCA thresholds. Classification was performed using the default parameters set by the software toolboxes we used. We did not compute the ROC curve to check if the above result was true over a wide range of threshold. An analysis of how the classification parameters change classification success rate will be performed in future work. Our expectation of obtaining better classification with voxels showing higher DTI–PANSS CCA correlation is not evident from our results. It may be possible to increase classification success rate by tweaking the classification parameters. Such an
alteration would make the result dataset dependent and independent verifications will be required. We hope to perform analyses of such nature in our future work. Our low classification success rates, obtained using default classification parameters, indicate that imaging measures that show high correlation with symptom scores may not be used as good features for efficient classification. Several reasons can be attributed to this result. The primary reason is that the symptom scores are evaluated at a particular point in time and may not be accurate. The manifestations of symptoms in a subject are not consistent over time. Hence an evaluation at a particular time point may not be an accurate representation of the history of symptoms in that subject. Longitudinal evaluation of symptoms and the use of an average symptom score may improve our results. Another reason can be attributed to the subjective nature of the symptom evaluation scheme. Symptom scores are assigned by a trained professional in psychometry but can slightly vary between different administrators. This variation can influence the voxel selection method used in our application. These inconsistencies in symptom scores further emphasizes the need for image based techniques to better understand and better classify schizophrenia.
Data fusion in human subjects is a challenging problem. Unlike fusion in other areas of imaging, such as remote sensing, where the scene is fixed or the degree of variability can be quantified, in human brain imaging the within subject and between subject inconsistencies are more complicated to compute. It is difficult to conceptualize the precise method to fuse data without making many assumptions.

In this thesis we developed novel and straightforward data analytic approaches, with minimal assumptions, to pair wise correlate multi modal brain data. Our analysis included all possible combinations of correlations of voxels from the whole brain. To our knowledge a fusion analysis of this nature is first of its kind. We introduced techniques to reduce and visually present high dimensional data. We also modified existing methods to fuse multi modal data. We showed how these methods can be used to extract features that are different between a group of patients with schizophrenia and healthy controls. The results reported in this work reveal interesting findings that are not possible to derive from conventional one dimensional/unimodal approaches. Techniques to exclude outlier subjects and to verify the robustness of results were also introduced in this work.
The structural–functional analyses we performed indicate that in patients with schizophrenia the correlations between structural data, from gray matter concentration, and functional data, from a sensory motor task activation map, are weaker in patients than in controls. Structural regions in the cerebellum show higher positive correlations with functional data in HC than in SZ. Structural regions in the frontal regions show higher negative correlations with functional data in HC than in SZ.

The functional–functional analyses indicate that in four out of six task combinations we analyzed, patients’ inter–task correlations are more positively correlated than controls’. In one combination the controls show more positive correlations and in the other there are no significant difference. The tasks for which patients show more similarity than do controls occur within the same scanning session and all other task combinations occur in different scanning sessions. This result implies that patients activate more similarly to different tasks that do controls when the tasks occur in different scanning sessions and the vice versa when the tasks are within the same scanning session. Controls are activating unique networks for different tasks and patients are activating less uniquely. Regions in the superior temporal gyrus and the inferior parietal lobule show higher inter–task correlations in patients than in controls.

We used another form of structural data, collected with DTI, to find correlations with behavioral data evaluated through symptom scores. Direct correlation, multiple regression and canonical correlation analyses were performed with these data sets to identify regions in the brain with significant correlation between image based measures and behavioral measures. We find that regions in the superior longitudinal fasciculus have large regions of significant correlations.
Results from all of the above methods support two existing hypotheses of schizophrenia, the disconnection hypothesis and the theory of cognitive dysmetria. Disconnection hypothesis is supported by our finding that structure and function are less correlated in patients than in controls. Patients were found to activate less uniquely for different tasks than did controls and this is consistent with the disconnection hypothesis if, due to disconnections, patients lack specialized networks to activate uniquely to different tasks. The correlation analyses of DTI FA value showed that all three PANSS scores were negatively correlated to FA. This result had been previously reported in other studies and indicates that patients with higher symptom scores are less directionally connected, again a sign of disconnection. The functional–functional analyses indicated aberrant correlation connections in patients in cortical, subcortical and cerebellar regions. Structural–functional analyses indicated less correlation in patients in the cerebellar and frontal regions. Regions we identified with aberrant linkages support the theory of cognitive dysmetria.

The approaches introduced in this work can be modified and applied to fuse data from other modalities, for example to fuse EEG or genetic data with structural or functional brain data. There are several advantages of the methods we have introduced. One advantage is that the two modalities that are to be fused can have different resolutions and thereby the loss of finer resolutions acquired in one modality can be avoided. The approaches are simple, easy to implement and efficient. They are simple, since they do not use complex mathematical equations, but rather the fundamental correlation equation to make all computations. They are easy to apply since the same simple computation is performed iteratively. The same methods can be used to pair wise
combine data of many different modalities. Cross correlation matrices of higher dimensions can also be computed with the same fundamentals introduced.

In our work we attempt to identify linear relationships between brain data. In future work we intend to extend the methods to identify non–linear relationships, for example using mutual information where higher order statistics can be investigated. Our methods make minimal assumptions and use minimal \textit{a priori} knowledge. The features obtained are results from purely data driven methods. The methods introduced can be further improved to investigate more precisely how structure and function are connected within a certain group of subjects. What we demonstrate here is that even using simple techniques, it is possible to extract features that are significantly different between groups of patients diagnosed with schizophrenia and healthy controls.

To compute the activation maps we used the canonical hemodynamic response function (HRF) and assumed that it was identical across all voxels, for all subjects, for all different tasks at all times. This assumption is commonly applied in a vast number of studies and precise measurement of the HRF for each voxel and time point is not easy. There are methods, for example independent component analysis (ICA), where analyses are performed without this assumption. Future work can be done where ICA and our methods are combined. The HRF assumption was applied to collapse the time domain of the activation maps. In this work no analyses were made in the time domain. Future work is needed to find differential features between the groups that may exist in the time domain.
Further studies are also needed to confirm our results. For structural–functional analyses we analyzed data collected from a large number of subjects from different sites and increases the statistical significance of our result. For functional–functional analyses a smaller number of subjects were used but we verified our results through a Monte Carlo method. It may be useful to apply our methods on theoretically related tasks to check if the same results can be replicated. Subject demographics were not incorporated in our analyses. We tried to match the mean for most demographics but they had a wide range. For example, subject age had a range of about forty years. It is reasonable to expect younger subjects to perform differently than older subjects. The groups included a mix of different handedness and medication. Subsets with these different attributes need to be formed and analyzed. In order to assess specific demographics, a larger number of subjects would be required. For demographics that had different mean values, we did separate analyses to show that they did not change the results.

The development of a classification algorithm to efficiently classify patients with schizophrenia remains a challenging task. This is due to the diversity of the data collected, especially that of patients. An initial step towards better classification algorithms is to identify features that discriminate groups of schizophrenic patients and healthy controls. Features that separate the two groups maximally are used to develop classifiers. The main focus of this work was to find new features of schizophrenia. We did not make a significant effort to develop better classifiers. In the histogram shift methods, we used a few features with a low level algorithm to show that additional information or data from different modalities can increase the classification rates. A classification success rate of about eighty percent for patients and about seventy percent
for controls was achieved with the leave–one–out scheme. Future work in classification includes using other statistical measure of the histograms with a more sophisticated classification algorithm. Investigations of different combinations of inter–task correlations with a weighting systems and the effect of the number of voxels are of interest.

We used advance classification schemes that used features from DTI–PANSS fusion, but the results did not show high classification rates. Patient classification rates improved slightly when more significantly correlated (DTI–PANSS) voxels were used as features. PANSS scores are evaluated at a particular time point and may not reflect the severity of the symptoms over a long time period. Longitudinal evaluations of the symptoms, if used, may result in better classification rates.

To summarize, this thesis introduced and investigated methods to combine multi modal data while incorporating the whole brain. To our knowledge, an analysis of this nature is first of its kind. Our results indicate several interesting findings that can be applied to better understand schizophrenia.
8 References

Andreasen NC. 1981. Scale for the Assessment of Negative Symptoms (SANS): University of Iowa.


Michael AM, Baum SA, Calhoun VD. 2009a. A Technique to Detect Outliers Automatically in Multi-Site fMRI Data. ISMRM. Honolulu, HI.


Michael AM, Baum SA, White T, Andreasen NC, Jung RE, Clark V, Gollub RL, Ho BC, Calhoun VD. 2009b. Fusion of Structural-Functional Brain Images Reveals Differences in Schizophrenia in a Multi Site Study. ISMRM. Honolulu, HI.


Michael AM, Calhoun VD, Pearlson GD, Baum SA, Caprihan A. 2009e. Application of Canonical Correlation Analysis to Identify Regions of Significant Correlation between Symptom Scores and DTI Measures in Schizophrenia ISMRM. Honolulu, HI.

Michael AM, Fries JF, Baum SA, Ho BC, Andreasen NC, Calhoun VD. 2008c. A Method to Analyze Correlations between Multiple Brain Imaging Tasks to Characterize Schizophrenia. IEEE SSIAI. Santa Fe, NM p125-128.


