12-31-2008

Data fusion based optimal EEG electrode selection for early diagnosis of Alzheimer's disease

Brian Balut
Rowan University

DATA FUSION BASED OPTIMAL EEG ELECTRODE SELECTION FOR EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE

by
Brian Balut

A Thesis Submitted to the

Graduate Faculty in Partial Fulfillment

of the Requirements for the Degree of

MASTER OF SCIENCE

Department: Electrical and Computer Engineering
Major: Engineering (Electrical Engineering)

Approved:

In Charge of Major Work

For the Major Department

For the College

Members of the Committee

Rowan University
Glassboro, New Jersey
2008

ABSTRACT

Brian Balut
DATA FUSION BASED OPTIMAL EEG ELECTRODE SELECTION FOR EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE
2007/08
Dr Robi Polikar
Master of Science in Electrical Engineering

As medicinal and technological advances lengthen the average human life span, diseases affecting the elderly such as Alzheimer's disease and Parkinson's disease are seeing increasingly growing numbers, especially in developed countries. As a result, the necessity for an accurate, inexpensive, noninvasive means of diagnosis becomes particularly important, since such a method is not readily available to the general population. One biomarker that has recently showed promise is the analysis of the electroencephalogram (EEG).

Over the course of two studies, more than 130 subjects have been recruited providing information from 16-19 EEG electrodes for each subject. These signals have been decomposed using the wavelet transform to be used in a pattern recognition and classification algorithm to serve as a diagnostic tool for Alzheimer's disease. Through the use of multilayer perceptrons and support vector machines, classifiers were generated on different portions of the EEG. These classifiers are then combined using combination methods such as sum rule, product rule, simple and weighted majority voting.

Classification performance for Cohort A was 88.7%, an increase of more than 5% over previous work in this study. Classification performance for Cohort B was 93.6% and the classification performance for both cohorts combined together was 82.7%. These classification performances exceed the diagnostic accuracy of community clinics (75%) and are close to diagnostic accuracy available at research and university hospitals (90%).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Primarily due to technological and medical advances over the course of the past century, the life expectancy of the average American has seen a dramatic increase of 30 years, from 47.3 to 77.8 years [1]. This increase in life expectancy, coupled with the fact that the 78.2 million Americans of the baby boom generation are now moving into their 60's, allows for a higher prevalence of predominantly elderly diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD). With 26 million cases of Alzheimer's disease and 4 million cases of Parkinson's Disease worldwide, these diseases are increasingly becoming a major public health concern [2,3].

1.1 ALZHEIMER'S DISEASE

Of the 26 million cases of AD worldwide, 5.1 million of those cases occur in the United States. AD is predominantly an elderly disease, with only 200,000 of the over 5 million cases occurring in patients younger than the age of 65. After 65, though, the chances of developing AD rise at an alarming rate. While only 2 percent of the population over 65 has AD, that number doubles every five years thereafter, reaching 42 percent of the population being afflicted with the disease after the age of 85. That number is also predicted to grow to 60 percent of the population over 85 by the year 2050 as Alzheimer's disease is estimated to affect somewhere in the range of 11 to 16 million Americans [4]. As a result of this increased prevalence, Alzheimer's disease has become

the seventh leading cause of death in the United States and fifth leading cause of death for Americans over the age of 65, trailing only heart disease, cancer, cerebrovascular disease, and respiratory disease [1]. With the growing population of people affected with AD, comes an increasing economical impact of the disease. A 2005 study showed that the total cost of dementia in the world was estimated to be $315.4 billion annually, with $83.6 billion of that coming from North America at an average of over $24,000 per year per patient. This figure includes direct costs from clinical testing and the care necessary to maintain the health of the individuals [5]. In comparison, the total US government's budget for the Department of Education in 2006 was only $56 billion [6].

The primary concern with Alzheimer's disease, and motivation for this work, is the diagnostic process of the patients. There is very little evidence as to what exactly causes the disease. Some links have been drawn to genetics, general health, and nutritional care, however, there is no known specific cause of the disease [7]. There is also no ideal pre-mortem method for a completely accurate, discernable diagnosis of the disease. Alzheimer's disease is almost always diagnosed through clinical interviews and neurological testing. This diagnosis includes interviews of not only the patient, but also of caregivers of the patient and the family of the patient [8]. These interviews and neurological exams, when conducted by an expert at a research or university hospital can achieve a diagnostic accuracy of up to 90%, which is regarded as the practical 'gold standard' for Alzheimer's disease diagnosis. A considerable problem is that many patients do not have access to or cannot afford the expensive costs of these facilities. Diagnosis is therefore made at local hospitals or community clinics or even by primary care physicians where the overall diagnostic accuracy drops to 75% with a sensitivity of

83% and specificity of 55% [9]. The true gold standard of diagnosis, which guarantees an accurate diagnosis, is brain biopsy, where a small piece of the brain can be examined under a microscope after drilling into the head. This process, however, is incredibly invasive and therefore not a feasible option for diagnosis [10].

The clinical diagnosis typically consists of one or more standardized testing procedures administered to the patient. The simplest and most frequently used of these procedures is the Mini Mental State Exam (MMSE). During this examination the patient is asked simple questions (such as the date, day of the week, present location) and asked to perform simple cognitive and memory tasks (such as repeating back a series of items, counting backwards from one hundred by seven, copying a diagram, or writing a sentence). The tests are scored on a scale of 0 to 30, 30 indicating cognitively normal and 0 being a near vegetative state. The test was developed in 1975 and while very simple in design, it does provide an effective overview of the patient's cognitive ability, verbal and written ability, and orientation. Other exams include the clock test, where the patient is instructed to draw a clock at a given time, the Buschke Selective Reminding Test, which measures the patient's short term verbal memory, the Wisconsin Card Sorting Test, which measures pattern sorting ability, the Trail Making Test, where psychomotor skills are analyzed, or a number of other testing procedures [10].

Apart from neurological testing, other methods such as blood, urine, or spinal fluid tests or neuroimaging can be done. Rather than attempting diagnosis of Alzheimer's disease through these tests, the physician is attempting to eliminate other potential afflictions such as stroke, brain hemorrhaging, or other forms of dementia such

as vascular dementia, or whether the symptoms can be attributed to the natural course of aging.

1.1.1 PATHOLOGY OF ALZHEIMER'S DISEASE

The human brain consists of about one hundred billion neurons. Connected in one hundred trillion different pathways, these neurons allow for the every day function of the brain including all cognitive and motor tasks [10]. When these neurons or the pathways interconnecting the neurons become blocked or damaged, the effect is a reduction in brain activity, such as the symptoms of Alzheimer's disease. The two classic hallmarks of AD resulting in neuronal blockages, as first noted by Alois Alzheimer in 1906, are commonly referred to as plaques and tangles.

The plaques consist of beta-amyloid (Aβ) proteins, which are broken down from larger amyloid protein precursors (APP). The original APP molecules are thought to have a role in nerve growth during development, while the beta-amyloid protein fragments are typically eliminated in a healthy brain. The most common form of the beta-amyloid protein consists of 40 amino acids, however, in Alzheimer's disease, an abundance of longer proteins with 42 to 43 amino acids are produced. These larger amyloid proteins are more prone to aggregate, and when they do, beta-amyloid plaques are formed. The exact role of these insoluble plaques is unknown since there have been cases of Alzheimer's disease reported where these plaques are not present. It is possible, however, that these plaques, when allowed to accumulate, block the connections between neurons eventually leading to neuronal death [11].

Whereas the plaques form outside of the neurons, the tangles are found within the structure of the neuron cell. In fact, the cause of the tangles is what gives the neurons their structure originally. Another protein, called tau, is involved in the creation of microtubules, which in part form the internal support and transfer system for the neuron. In Alzheimer's disease, the tau protein is modified to a point where tau proteins begin to pair together and become tangled within one another. The tangling results in a collapse of the microtubules and eventually the inability to communicate to other neurons, also leading to neuronal death. [12].

As these plaques and tangles develop throughout the brain, more and more neurons are eliminated leading to overall atrophy of the brain itself. Figure 1.1 shows the decline of an Alzheimer's disease patient's brain as the patient moves through the different stages of the disease. There are four main stages of the disease: preclinical AD, mild AD, moderate AD, and severe AD. While all patients may not progress through the disease at the same rate, all patients will experience each of the different stages of the disease. The most distinguishing feature of Alzheimer's disease is that unlike other dementias and conditions with similar symptoms, AD is an incredibly slow, gradual onset disease. Often, its onset is not even perceptible, and it may be years before an identifiable change in the person has developed [10]. It is predicted that the plaques and tangles begin to form one to two decades before any noticeable symptoms become apparent in the patient [12]. The average life expectancy for a patient after being diagnosed with AD is then only 8-10 years, although this varies from patient to patient and is dependent upon such factors as the age at which the patient developed the disease and the stage at which the disease was diagnosed [13].

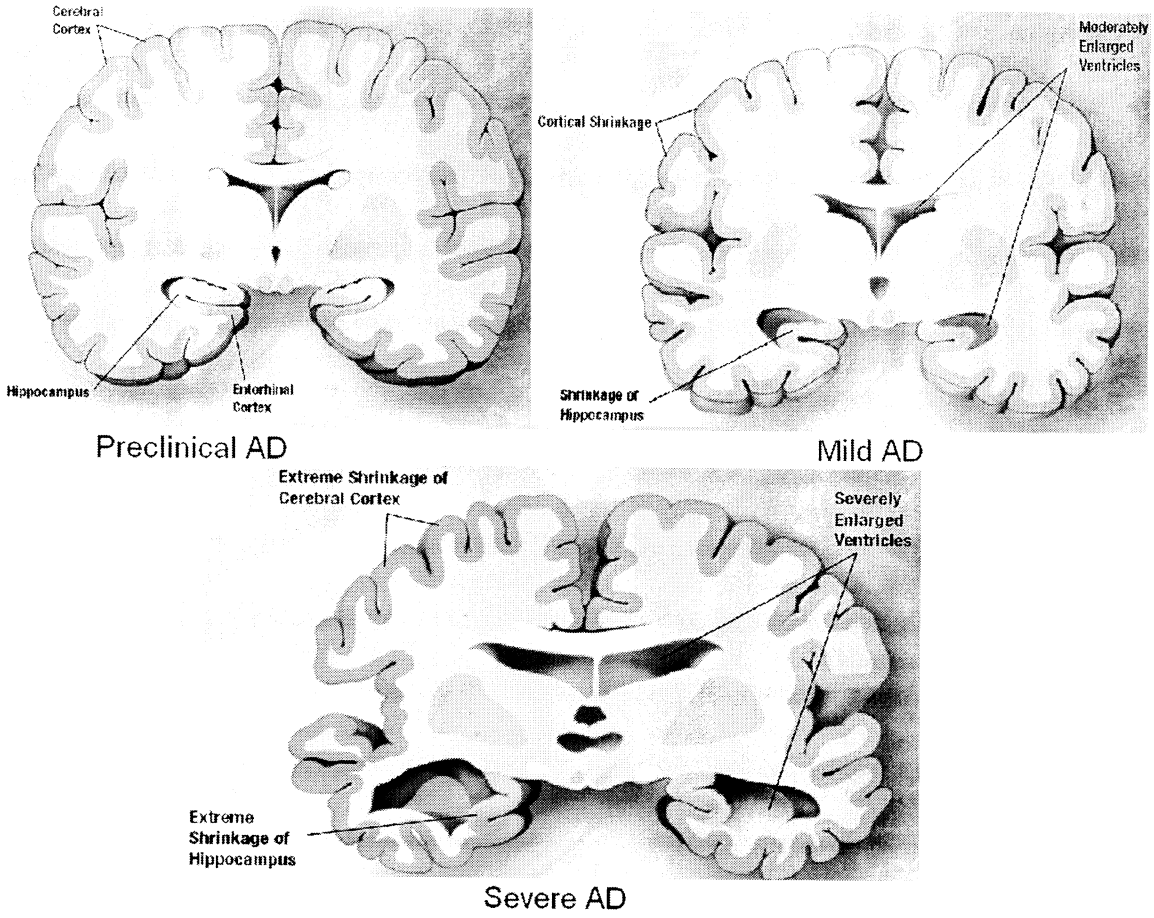

Figure 1.1: Atrophy of a patient's brain through preclinical, mild, and severe stages of Alzheimer's disease [12].

The plaques and tangles first appear in the entorhinal cortex soon followed by the hippocampus, both of which are interconnected. The hippocampus is responsible for short term and long term memory formulation. When this area becomes adequately degenerated, the first signs of AD become visible in a patient. The primary initial symptom is memory loss. This symptom usually is not enough to give a full AD diagnosis, but rather a diagnosis of mild cognitive impairment (MCI). From the hippocampus, the disease then spreads to the temporal lobes of the brain as the patient moves into a state of mild AD. Loss of judgment, reasoning, memory, and language as well as confusion about location and small changes in a patient's ability to perform daily

tasks can be seen at this stage. It is usually at this point that a patient or family of a patient begins to seek medical diagnosis and the diagnosis moves from MCI into Alzheimer's disease. As the patient moves into a moderate stage of AD, the plaques and tangles have extended themselves into regions of the frontal and parietal regions of the brain. Initial symptoms become worse as well as the addition of behavioral problems, hallucinations and motor control issues. The final stage of Alzheimer's disease is severe AD, where the plaques and tangles have spread throughout the entire brain and the patient can no longer function at all by their own means. Weight loss, lack of bladder and bowel control, and failure to even recognize family or loved ones accompanies the final stage of the disease. The brain has also experienced extreme atrophy at this point including shrinking of the cerebral cortex and extreme enlargement of the ventricles in the brain [10,12]. These changes are not immediate and there is no sudden change from stage to stage, but rather a gradual shift over a number of years.

1.2 OBJECTIVES OF THIS STUDY

The goal of this work was to create an automated, non-invasive, readily available for clinical use tool for the early diagnosis of Alzheimer's disease. Two individual studies were conductied, which involved the collection of event related potentials (ERPs), obtained from the electroencephalogram (EEG) from cohorts of normal and Alzheimer's disease patients. The signals were collected for each study in separate locations, using different EEG hardware by different technicians. Both studies used the same oddball paradigm protocol. Analysis of the signals was performed using wavelet analysis that provided a set of wavelet coefficients. The coefficients from the different spectral bands

of the EEG were used as features to train and test an ensemble of neural network classifiers. A detailed analysis of electrodes, frequencies and stimuli was completed for the entire data using two separate neural network architectures, multi layer perceptrons (MLPs) and support vector machines (SVMs). A final analysis was completed to be applied to the general Alzheimer's disease diagnosis problem based on the combined results of both sets of data.

1.3 ORGANIZATION OF THIS THESIS

Chapter 2 provides a literature review on Alzheimer's disease diagnosis research including biochemical analysis, neuroimaging, and electroencephalography. The processes and methods used within this study are outlined in Chapter 3 including signal acquisition, wavelet analysis for feature extraction, neural network architectures for automated classification algorithms, and ensemble based data fusion. Chapter 4 discusses the precise implementations used and the accompanying results. Chapter 5 reviews the analysis and draws conclusions based on this study.

CHAPTER 2

BACKGROUND

In addition to researchers attempting to identify the cause of AD or reliable treatments, the diagnostic process has received considerable attention recently. Looking beyond clinical interview, there are three main areas of research actively seeking a better means of early diagnosis; biochemical analysis, neuroimaging, and physiological analysis.

2.1 BIOCHEMICAL ANALYSIS

The biochemical analysis for the early detection of Alzheimer's disease focuses primarily on the cerebrospinal fluid (CSF), the fluid surrounding the brain and found in the ventricles within the brain. The CSF is extracted through a procedure known as lumbar puncture (or spinal tap), which involves puncturing the spine in the lumbar region of the back. The procedure is highly invasive, painful, and expensive; however, the results from such studies have been promising.

When extracting the CSF, three specific biomarkers are of interest, beta-amyloid$_{42}$ (responsible for plaques), tau protein, and phosphorylated tau (responsible for neurofibrillary tangles). Concentrations of these proteins within the CSF have shown to be somewhat reliable in the predictability that an MCI patient will develop Alzheimer's disease with 95% sensitivity. Differentiation is also possible between Alzheimer's disease patients and cognitively normal patients based on these concentrations [14]. The strength of these biomarkers is increased even more when combined with metrics taken

from neuroimaging techniques as shown in [15,16]. This process is far from becoming standard diagnosis, however. When dealing with CSF, in addition to the invasiveness of the process, there is significant variation in test results from site to site, which is a major drawback of the technique [17].

2.2 NEUROIMAGING

In addition to, or more often in conjunction with, biochemical analysis, various forms of neuroimaging can be used for diagnosis of AD. Virtually all forms of neuroimaging have been applied to the diagnosis or monitoring of Alzheimer's disease in various research applications. The most common forms of neuroimaging used today are MRI, PET, and CT.

The different types of imaging monitor various aspects of the brain and therefore different aspects of the disease progression within an Alzheimer's disease patient's brain. Most common of all is the MRI and fMRI, which are used primarily to identify volume loss and atrophy of the brain, specifically in the hippocampus region. The monitoring is typically done over a longitudinal study, which allows the atrophy to become more visible [18,19]. PET/SPECT scans monitor the glucose metabolism within the brain. Again these are typically longitudinal studies showing a decrease in glucose metabolism in relation to the loss of neurons within the brain [20]. The CT scan is used to monitor blood flow within the brain, also showing the degradation of the brain. The CT scan has also been used in differentiating between Alzheimer's disease and vascular dementia [21]. Although these techniques do provide valuable information regarding brain function over the course of the disease, their ability to serve as an early diagnostic tool is

limited due to the high cost and specialists required to operate the imaging hardware (such imaging hardware may not be available at community clinics), as well as the need for continual follow up visits for the scans to deliver the most helpful diagnostic information.

Fritzsche, *et al.*, 2006, used a dataset of 68 subjects including 27 normal control subjects, 16 MCI subjects and 25 AD subjects. They used MRI images from the subjects and created six features from the images. Principal component analysis (PCA) was used on sets of four features with classification done using a Fisher Linear Discriminant (FLD). They also used neural networks to perform classification. Results provided 80% classification of AD test patients and 85% of control subjects. Sensitivity between controls and MCI subjects was 81% with a specificity of 80%. AD patients were correctly classified 59% of the time when compared to MCI patients [22].

An early study incorporating another form of imaging using the low resolution electromagnetic tomography (LORETA) technique of mapping EEG data to create a three dimensional model of the brain showed some promising results in comparing cognitively normal patients and Alzheimer's disease patients as well as distinguishing between different stages of AD. The technique was able to localize some changes in activity to the left temporo-parietal region of the brain, however, much more work is necessary to create a feasible diagnostic tool using this method [23].

In addition to these well known imaging techniques, some newer techniques have been developed over the last few years, which show some potential for the diagnosis of AD. One relatively new technique being implemented is infrared and near infrared spectroscopy (NIRS) for the potential early diagnosis of AD. These techniques use

infrared light, which penetrates skin and bone tissue to gain a detailed view of the intramolecular bonds and hemoglobin content inside the skull in a completely noninvasive fashion. This technique is still new and much more research is necessary to have any reliable results [24,25].

Another new technique emerging with potential for early diagnosis of Alzheimer's disease is fluorescence anisotropy. This technique works on the biological level through experimentation with the beta amyloid protein. Allsop, *et al.* used time-resolved anisotropy measurements (TRAMS) to study the aggregation of the Aβ protein. The experiments employ small quantities of fluorescently labeled Aβ as well as unlabeled proteins, which were combined with other chemicals and allowed to aggregate. The technique was shown to be able to identify the Aβ aggregation well before it is apparent using other common methods. Once fully developed, this technique could allow for a high-throughput procedure for early detection of the Aβ aggregates, which are a sign of early AD [26].

Research within the realm of genomics is also gaining attention recently. Marcotte, *et al.* have used cDNA microarray and proteomic approaches in their study of the AD brain. From human hippocampal samples from human brains, specific neurons have been identified for use in animal studies. Through the use of rats, they have been able to develop a model of cognitive impairment. This model allows for characterization between impaired and unimpaired subjects. Based on these classifications, they have begun work examining cDNA profiles. By using what they learn from the animal studies, they could potentially apply their findings to AD classification in humans [27].

2.3 ELECTROENCEPHALOGRAM

Unlike biochemical analysis and neuroimaging, the electroencephalogram (EEG) offers a method for diagnosis which is completely noninvasive and inexpensive by comparison. It also is readily available at most community clinics or can be made available. The origins of the technique date back as far as 1875 when Richard Caton recorded electrical signals from dog and rabbit brains, however, it was not applied to humans until 1924 by Hans Berger [28,29]. Since its discovery, the EEG has been used extensively in brain research and disease diagnosis, particularly epilepsy, sleep related disorders, coma, and dementia. The EEG quickly became the premier clinical tool used in research of the human brain until the development of the aforementioned neuroimaging techniques [28].

The scalp EEG is collected through the use of scalp electrodes placed directly on the head with reference electrodes applied to the ear lobes. These electrodes then measure electrical activity typically in the range of microvolts (μV), commonly termed brain waves, from the brain. The recordings show variation when recorded from different areas of the head, which leads to multiple electrodes recording electrical activity concurrently for a complete EEG recording. To ensure a common system of electrode placement on the scalp, a standard method of placement was designed called the International 10/20 System of electrode placement. The numbers in the electrode locations refer to percentage intervals of placement based on the circumference of the head. A measurement is taken from the top of the nose to the bone protrusion at the back of the head, and then electrodes are placed in 20% intervals around the head with 10% intervals added for additional electrode placement. A representation of the placement can be seen in Figure 2.1. In the diagram, the electrodes are labeled by their position on the

head in terms of the lobe of the brain they are placed over. The abbreviations are as follows: F – frontal, C – central, P – parietal, T – temporal, O – occipital, and A – auditory reference. The numbers describe the location of the electrode on that portion of the head with the even numbers on the right side, odd numbers on the left, and z electrodes on the centerline of the head [30]. The electrodes highlighted in black have different names than the standard 10/20 system, however the locations are the same. All highlighted electrodes (both black and grey) are the ones used within one or both studies in this thesis.

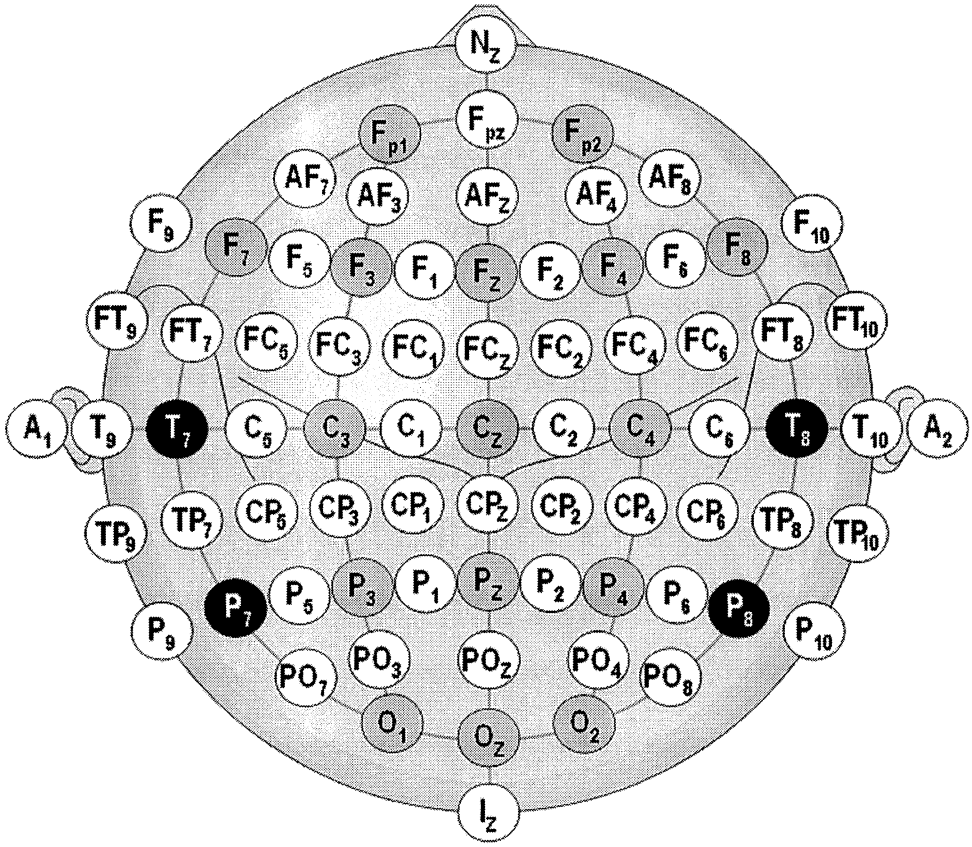

Figure 2.1: International 10/20 System of electrode placement [31].

Despite the benefits of the electroencephalogram, it is far from perfect. Although the time resolution of an EEG is very precise, the spatial resolution suffers greatly due to

so much activity in the brain and a limited number of electrodes on the scalp with bone and skin matter in between them and the brain. Information can typically be localized to a region or hemisphere but not pinpointed exactly. In many seizure related studies, subdural electrodes are placed directly on the brain. While giving an increased spatial resolution, this is highly invasive as it requires cutting into the skull and also runs the risk of infection and hemorrhaging [32]. A third type of electrode is the subdermal needle electrode, which is inserted beneath the skin without penetrating the skull. These electrodes require multiple pinpricks of the patient's scalp, are more costly than scalp electrodes, and more suited for ER and ICU applications.

Artifacts can corrupt the EEG quite readily. Since the electrodes are placed directly on the scalp and recordings are extremely low in amplitude, factors such as eye or muscle movement, electrical noise, and even sweating can cause the recording to become corrupted. These artifacts or the artifactual recordings must be removed prior to any analysis of the data. This removal is usually done by either an EEG technician or in many cases automatically by the data collection equipment.

2.3.1 SPECTRAL CONTENT OF THE EEG

The majority of the information collected through EEG recordings occurs below 50-60 Hz. This range is divided into five separate bands, typically referred to as rhythms, which have been found to relate to different brain functions [28]. The EEG is subdivided as follows:

Delta rhythm (< 4 Hz) – The delta band of EEG activity represents the slowest but the highest amplitude waveforms. These rhythms are most commonly found in deep sleep and pertain mostly to unconscious processes within the mind. In a healthy individual, the activity of the delta band is inversely related to the amount of attention and focus an individual is currently using. When observed in awake adults, it is usually a sign of some sort of cerebral damage or encephalopathy [28,33].

Theta rhythm (4-7 Hz) – The theta band is the second slowest set of rhythms. These rhythms are typically present during drowsy periods such as just before sleep or during daydreaming or fantasizing. They are commonly seen in young children, however, their presence in adults is usually indicative of stress or disease [28,33].

Alpha rhythm (8-13 Hz) – Alpha rhythms are the most common type of activity in a healthy adult brain. Focused around the occipital and frontal regions, these waves are known to be most prevalent in an awake, relaxed state in adults. From this state, a person can easily shift into another state to perform a given task [28,33].

Beta rhythm (14-30 Hz) – The beta range is large and consists of fast activity within the EEG. The band is usually subdivided into two or three separate sub bands, which represent different levels of activity. This rhythm is the most common waveform when a person is awake with eyes open and engaged in some sort of activity such as listening, thinking, or processing information [28,33].

Gamma rhythm (> 30 Hz) – The highest frequency waves are the gamma rhythms, which take place throughout all regions of the brain. These rhythms are thought to occur when information processing in the brain requires information from multiple parts of the brain all at the same time [28,33].

2.4 ACQUISITION PROTOCOLS AND EVENT RELATED POTENTIALS

Over the long course of EEG related research, protocols have been developed to extract particular responses from the EEG. These protocols typically operate by presenting a particular stimulus or sequence of stimuli to the patient, which then elicits a response in the EEG. These stimuli can take one of four forms: auditory, visual, somatosensory, or olfactory. The auditory stimuli are typically a series of different frequency tones or sound clips. Video stimuli are typically conducted with patterns, colors, words, etc. Somatosensory stimuli are common in studies on pain and generally performed by directing an electrical current at some point on the body, for instance the wrist [34]. Olfactory stimuli involve presenting odors to the patient as opposed to sounds or visuals [35]. The resulting response to the stimulus within the EEG is called an event related potential (ERP) or evoked potential. Protocols are then created through arranging a series of stimuli in a particular order to generate a response from the patient. There are many different types of protocols from single stimulus protocols, go/no go experiments, oddball paradigm, etc.

The protocol used in this study is a modified version of the auditory oddball paradigm. In the standard oddball paradigm, there are two distinct stimuli, a standard stimulus and a target or 'oddball' stimulus. The patient is presented the standard stimulus

for a majority of the time. In the uncommon case where the target stimulus is presented, a potential is generated within the EEG. The stimuli are usually presented in a random order and with random interstimulus interval to further increase the novelty of the target stimulus. In most cases, the patient is instructed to either push a button or count each time they are presented with the target stimulus [36]. For the auditory oddball paradigm, the two different stimuli are typically generic sine waves of easily distinguishable frequencies.

The modifications to the oddball paradigm as used in this study came from Yamaguchi, *et al.*, 2000 [37]. In addition to the standard and target tones of the auditory oddball paradigm, a third category was added consisting of novel environmental sounds, such as a dog barking or bell ringing. These sounds were clips taken from Disney movies and edited to a length of 200 ms. A total of 60 unique sounds were presented to each subject. The exact breakdown of stimuli was 65% standard, 20% target, and 15% novel. The subjects again were instructed only to respond to the target tones and were not warned about the novel sounds at all. By using this third stimulus type, another form of evoked potential generated by different regions of the brain can be collected, which may be used to better differentiate between different types of dementia.

The data used in these studies are event related potentials. When a number of these responses to similar stimuli are averaged together, a series of peaks and valleys are generated. The number of individual stimuli necessary for the average has been found to be a minimum of twenty [38], however, a larger number of responses (> 50) is desirable. The need for averaging comes from the nature of the ERP response. The peaks and valleys have such small amplitudes that they are buried within the noise floor of the EEG

itself. By averaging, the features of the ERP become much more prominent. The features of a typical oddball paradigm ERP can be seen in Figure 2.2.

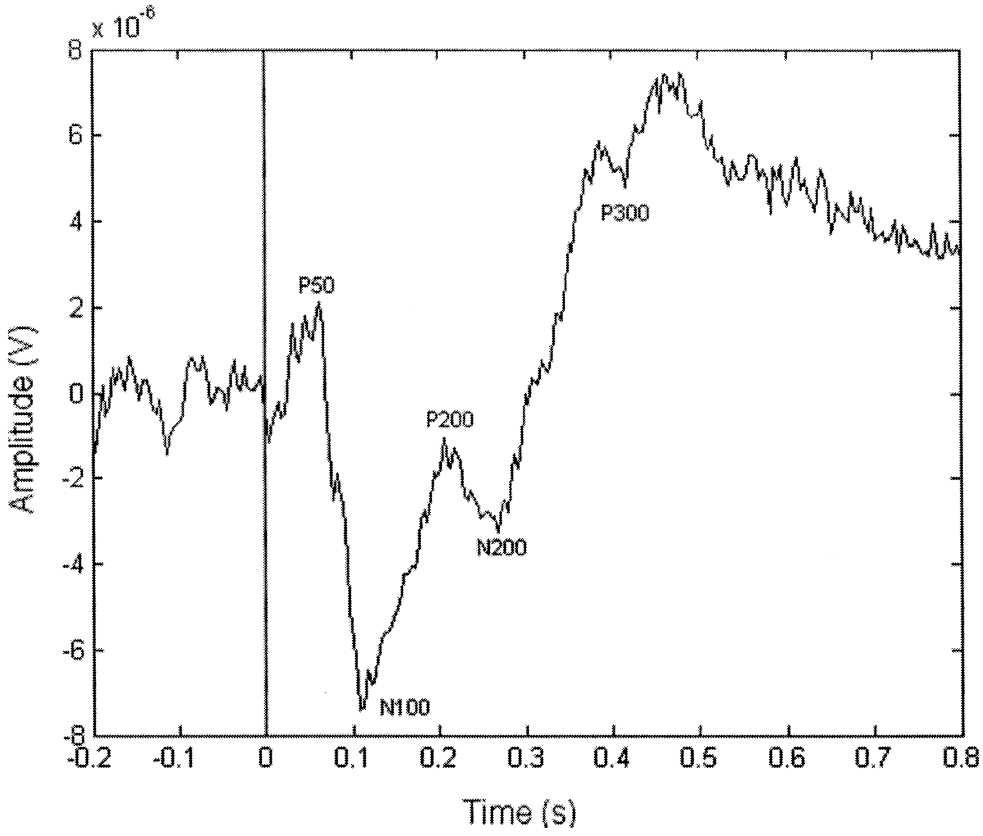

Figure 2.2: Sample ERP signal with components labeled.

The different peaks and valleys of the ERP can be described by the direction of the peak and the latency of the maximum of the peak after the stimulus. The most prominent features of the ERP are the P100 (P1), P200 (P2), P300 (P3), N100 (N1), and N200 (N2). While these may not happen at the exact latency in the name, there are given ranges for each of the different peaks. For example, the P2 occurs from 150-270 ms post-stimulus [37]. The most critical and presumably informative of all these peaks is the P300.

The P300 or P3 is the largest of all the peaks in terms of amplitude, measured at the maximum of the peak occurring around 250-500 ms. First noted in 1965, the P300

has been shown to relate to cognitive ability by numerous studies. It is primarily present in responses to the oddball stimuli in the two-stimulus oddball paradigm and both the oddball and novel responses in the three-stimulus oddball paradigm. The P300s are different in response to the two different stimuli, however, and as such have been given separate names of P3a for novel responses and P3b for target responses. The naming scheme comes from the novel P300 occurring with a lower latency and therefore occurring earlier than the target P300 [39].

The inherent difficulty of using the ERP as a diagnostic tool comes when evaluating a patient's ERP individually. There are known to be distinguishing features when comparing a cohort of AD patients to a cohort of normal patients, however, those features may not be easily differentiated when looking at single patient ERPs. In addition, the P300 is not necessarily prominent in all normal subjects and it may be prominent in some AD patients, which is shown in Figure 2.3. Due to these inconsistencies, further examination is necessary beyond visual analysis.

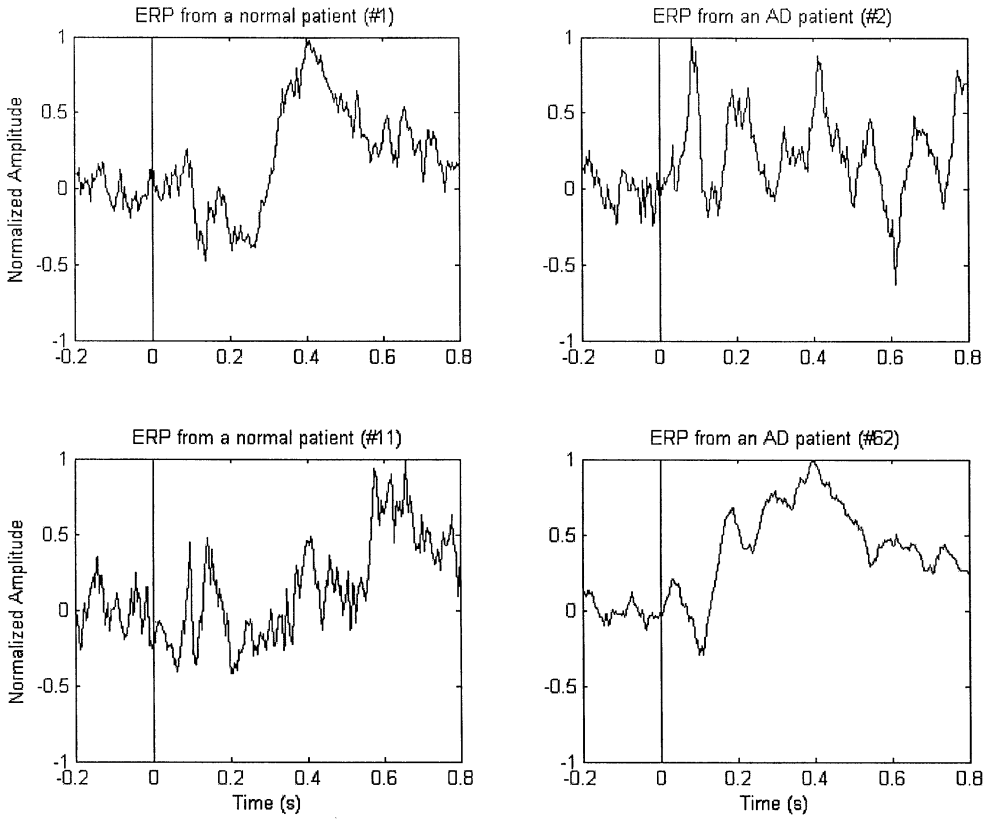

Figure 2.3: Comparison of ERPs of normal and AD subjects. Top left shows an expected P300 from a normal subject. Top right shows an unexpected lack of P300 from a normal subject. Bottom left shows an expected lack of P300 from an AD patient. Bottom right shows an unexpected P300 from an AD patient.

2.5 P300 RESEARCH

The P300, its origin, and its applications to research have been studied extensively over the past 40 years. The original hypotheses about the P300 were that it was generated in the hippocampal region alone. This notion was disproved by Polich and Squire in 1993 as their study showed it was more likely generated from the temporal-parietal junction, although the possibility of multiple sources was also accepted theoretically [39]. In 1997, a direct comparison between two stimulus oddball paradigm and a single stimulus protocol were compared with the only difference in the results being that the oddball

P300 produced a slightly larger P300 while all other aspects were statistically the same. This showed that either type of active protocol, where the patient is required to respond to a stimulus, elicit the same sort of response most likely from the same portion of the brain [36].

In 1999, a study comparing visual and auditory stimuli from both active and passive protocols was performed. Results showed that the active protocols yielded larger P300's having longer latencies. The auditory stimuli also yielded larger P300 components than the visual stimuli [40]. A 2002 study again compared oddball paradigm to a single stimulus paradigm using visual stimuli, with the primary focus on interstimulus interval (ISI). The study again showed that the oddball P300 produced a larger P300 amplitude than the single stimulus paradigm at short ISI of 2.5 seconds, but when the ISI was extended to 30 seconds, the ERPs became a lot more similar with the major difference being the duration of the P300 is longer in the oddball response [41].

It is now believed that the P300 is composed from components from different areas of the brain simultaneously, however, the exact contributing factors are unknown. It has also been shown that the P300 varies in relation to the difficulty of the task and attention required to accurately distinguish between standard, target, and novel stimuli. In addition, the P3a and P3b have been shown to originate from similar areas with some variation in the generation, which is thought to be indicative of the neurophysiology of memory storage. Due to the variation, the two different P300 responses are considered to carry different information from one another. The P3a has also been shown to be found in only 20% of cognitively normal patients [42].

2.5.1 WAVELET P300 ANALYSIS

The EEG is decomposed into frequency sub bands, which are analyzed during EEG analysis. The frequency content of ERPs is commonly analyzed according to these sub bands. After averaging the ERP signals together to generate the visible peaks and valleys, the resulting signal is quite short in length (around 1 second). In this averaged signal, the latencies are just as critical as the amplitudes to the informational value of the signal, therefore a time-frequency representation is necessary for analysis. The most widely used method is the wavelet transform, which allows for a frequency analysis of the signal while keeping the transformed signal time dependent.

A 2007 study by van Vugt, *et al.* gave a comparison between three different spectral analysis techniques, wavelets, multitapers, and $P_{episode}$. Conclusions showed that the $P_{episode}$ was most useful for longer signals with multiple cycles of information while multitapers are most helpful at localizing higher frequencies in the signal. Wavelets were most helpful for lower frequencies and can trade off between frequency and time localization based on the type and length of wavelet [43]. Since this study deals with ERPs, which are short, non-periodic signals with emphasis of interest on low frequency bands, the wavelet appears to be the best suited method.

Başar-Eroglu, *et al.* 2001, compared three different paradigms of stimulus presentation: oddball paradigm, single stimulus with variable ISI, and single stimulus with constant ISI with every fourth stimulus removed. Using wavelets, the signals were then decomposed and viewed topographically to visualize the distribution of the P300 throughout the head. While the differences in frequency bands between paradigms is not

of all that much interest, the oddball paradigm P300 was visible in all recording sites and accompanied by delta peaks around 2 Hz [44].

Demiralp *et al.*, 1999a, used the oddball paradigm to generate ERPs from subjects. They then applied wavelets to the signals in the form of the quadratic B-spline wavelet to acquire a time-frequency representation of the signal. The goal was to show differences in cognitive processing between the different stimuli responses. Ultimately, differences were shown in the theta, alpha, and specifically delta frequency bands. The N1 and P2 components showed little to no difference between the stimuli responses at any frequency, while the N2 showed larger amplitudes in the time domain and higher latency in the delta frequency range for target responses versus non-target. Overall, they observed significant benefits from the decomposition helping to separate simultaneous events related to cognitive information processing, which are not possible in the time domain alone [45].

In a companion study, Demiralp *et al.* 1999b, also applied the use of wavelets to single trials from the oddball paradigm. In this study they used a 5 octave quadratic B-spline wavelet on each stimulus response. They detected that the 4^{th} delta coefficient was the main feature indicating a P300 response. They then used this as a criterion for selecting which responses to average together. After averaging, they were able to compare the target responses using the enhanced P300s from the criterion rejection method and the remaining P300s that were not included in the averaging. As expected, the target responses yielded larger P300s with significantly different latencies [46].

Başar, *et al.*, 2001, applied wavelets to EEG oddball paradigm signals and compared them with digital filtering methods. The focus of the spectral analysis is from

the delta to gamma bands. The wavelets were shown to extra delta band information and check the occurrence of alpha band information not visible through a Fourier transform. They summarized the advantages of using wavelets over standard filtering techniques into three categories: time localization of frequency components, no need for a fixed time window, and significant data reduction and compression [47].

Demiralp, *et al.*, 2001a, again used a five octave wavelet for analysis of the P300 for both averaged and single trial ERPs. Essential results of this study involved the ability to compare P300s from different groups of subjects with different cognitive behavior, ages, or pathologies. It also provided more confirmation that the key distinguishing characteristics of the P300 lie in the delta band after wavelet decomposition [48].

Demiralp, *et al.*, 2001b, examined the differences between P3a and P3b under a variety of different difficulty visual stimuli. The difficulty was based on how easily distinguishable the target and novel stimuli were from the standard. Like in other studies, when the difference was easily detectable, the target P300 had a higher amplitude at all electrode sites. When the difference became smaller, the novel P300 became more pronounced. Through the use of wavelets, it was also determined that theta activity increased in the P3a as the difficulty increased and the delta coefficients were affected in the P3b [49].

2.6 ERPS IN ALZHEIMER'S DISEASE RESEARCH

The ERP has been used extensively in AD research over the past few decades. To list every study involving the use of EEG and AD would extend well beyond this thesis, so

only some basic points will covered. For a more complete overview of the EEG in AD, there are two overview papers exploring the subject by Hillert, Jeong and Olichney, [50,51].

A majority of the studies involve the amplitude and latency of the ERP components, in particular the P300. It has been shown that there is an overall linear slowing of the ERP in relation to age in cognitively normal patients, found to be 1-2 ms in latency per year [52,53]. In almost all studies, the P300 in Alzheimer's disease patients has been shown to have a longer latency than normal subjects, however, this is not unique to AD since conditions such as vascular dementia and schizophrenia exhibit similar symptoms [54,55,56,57]. This trend is also only evident when comparing cohorts of patients. On a patient to patient basis, the difference is usually not statistically significant. In addition to the latency, an amplitude decrease has been in debate for a number of years. Studies from Polich, *et al* in 1989 showed a definitive decrease in amplitude of the P300, however, a 2000 study by Golob and Starr showed there is little statistical significance of the amplitude [58,59]

Coherence analysis has also been insightful in the determination of ERP differences between AD and normal patients. A study of 14 AD patients and 10 normal patients showed an increase in the alpha 2 sub band of the ERP in AD patients and a decrease in control subjects after task completion as compared to the EEG before the task. Even larger differences were found in the alpha 1 sub band where there was a marked increase in control subjects and no change in AD patients [60].

Most of the aforementioned studies were carried out using auditory stimuli in the oddball paradigm, which represents the predominant testing method in ERP studies.

Visual stimuli have also been used quite extensively. Like the auditory stimuli, amplitude reduction in the P300 has been noted in visual stimuli studies [61]. While there have been some similarities between results from visual and auditory stimuli, the results are not exactly identical. There were similar yet differentiable features in the visual and audio ERPs, which suggests slight differences in the way the brain processes the two different stimuli. In addition to the differences between stimuli, relations were also made to the testing procedures showing auditory stimuli correlate better with the MMSE while visual stimuli correlate better with the Raven's Coloured Progressive Matrices test [62].

A major swing in research now is in the direction of neuroimaging, particularly in conjunction with EEG and ERP. It has been shown that diagnosis performance is increased when combining the outcomes of EEG and MRI tests than either testing procedure can perform on its own. This increase is due to the tests monitoring different aspects of cognition and therefore complementing the information contained in one another [63]. Additionally, correlations have been found between the severity of dementia and the metabolism of glucose through the combination of EEG and PET scans. Szelies, *et al.* showed that the latency, but not amplitude, of the P300 in EEG correlates to the glucose metabolism during a PET scan in Alzheimer's disease patients [64].

Another area receiving much research attention presently is familial AD (FAD). Although only a small percentage of AD cases are linked to genetics, they do encompass the majority of early onset AD (prior to 65 years of age). A 1995 study revealed links to the P50, N100 and P300 showing potential risk to acquire AD in subjects with a first-degree relative afflicted with the disease [65]. Another study from 2006 recorded ERPs

from Alzheimer's disease patients and their children to examine any potential evidence for development of the disease in the children. This particular study not only replicated results showing that AD patients have longer latencies in their P300 components, but more significantly that their adult children also showed a slowed P300 as well as smaller amplitude P300. They also reference other studies in which no such correlations are found and relate the differences to experimental methods [66].

2.7 ALZHEIMER'S DISEASE CLASSIFICATION

With some knowledge of what features are capable of distinguishing an Alzheimer's disease patient from a cognitively normal subject, classification can be done, eventually leading to an automatic diagnosis method.

Polikar, *et al.*, 1997, proposed an approach of using multilayer perceptron neural networks as a means of diagnosing AD through the use of ERPs. A 28 patient cohort, 14 AD, and 14 controls was analyzed. A two stimulus oddball paradigm was used with 86% 1 kHz standard tones and 14% 2 kHz target tones. Training and testing cohorts each consisted of 14 subjects, 7 AD and 7 normal for each. The raw time domain signals were the original input data for the neural network. Wavelets were then used to extract features from the ERPs to be used as the input into an additional neural network. Classification performance on the raw data was only 50-64%, while the DWT coefficients provided 79% performance. Through additional training methods, performance was increased to 93%. While this experiment demonstrated that a neural network based approach to the diagnosis of AD is possible, a larger cohort would be necessary to validate the procedure [67].

Petrosian, *et al.*, 1999, used a cohort of 3 probable AD subjects and 3 age matched controls. A recurrent neural network (RNN) was used. Nine channels of the EEG were recorded and the resulting signal was divided into 2 minute segments, which were used as inputs to the neural network. The RNN used an Extended Kalman Filter based algorithm that adapts the network weights on an instance by instance basis. In addition to the raw data, wavelets were used to decompose the data to provide additional training features with the Daubechies 4 (db4) wavelet. The network was trained on 1 AD patient and 1 normal subject and tested on the remaining four. Although the sample size was quite small, the performance was consistent throughout all testing [68].

Petrosian, *et al.*, 2001, expanded their cohort to include 10 AD patients and 10 controls. Again a nine channel EEG was taken from this patient, this time while the subject was at rest with no acquisition protocol. The db4 wavelet was used once again to extract input features for the RNN. For the training, 3 AD patients and 3 normal controls were used and the remaining 14 subjects were used for testing. The network was able to classify all of the remaining subjects correctly (7 of 7) and 5 of the 7 remaining AD patients [69].

Park, *et al.*, 2001, used a cohort of 25 mild AD, 12 severe AD, 17 age matched normal controls, and 7 young controls in P300 analysis of the EEG. A two stimulus oddball paradigm was used with 75% 1 kHz standard tones and 25% 1.5 kHz target tones. Five electrodes were used from the 10/20 International system. Rather than pressing a button for the stimuli, the subjects were asked to count the number of times the target stimuli was presented. There was a large gap in counting accuracy between the controls and the AD patients: age matched control (88%), young control (100%), mild

AD (20%) and severe AD (0%). Analysis of the ERP signals was done using the N200 peak as well as the P3a and P3b. Results showed that the P3 component in the mild AD patients had a higher latency and the severe AD patients high still, while the amplitude showed no significant differences between the groups [70].

Yagneswaran, *et al.*, 2002, used a cohort of 9 probable AD patients and 10 controls. From these patients, 37 recordings were used: 18 from AD patients and 20 from controls. A total of nine electrode positions were recorded for each subject. Diagnosis was done by looking at power frequency and wavelet coefficients of the recordings. The overall power frequencies of the signals yielded no distinguishing differences between the two groups, so bandpass FIR filters were used with a Hamming window to split the signals into four subbands: delta, theta, alpha, and beta. These subbands along with relative power of the signals and slower wave ratio were used as the inputs into a neural network. They also used the averages of wavelet coefficients extracted using the Daubechies 5 wavelet as features for an additional neural network. The classifier used was learning vector quantization. The resulting classification on the test data provided 94.7% (18 of 19) classification accuracy using the power spectra features and 89.4% (17 of 19) classification accuracy using the wavelet coefficients [71].

De Thad, 2002, looked at the physiological markers of AD by using a resonant recognition model (RRM) to analyze the amyloid plaques, precursor proteins (APP) and amino acids found in the creation of these proteins. The RRM is a physico-mathematical model used for analysis in the interactions of protein and its target using digital signal processing methods. The RRM determines common frequency components in the spectra of proteins, which then allows for identification of particular biological functions or

interactions. A variety of wavelets were used for frequency analysis (including Morlet, Coiflets, Daubechies, Symlets, and Meyer). Results showed that the resulting energies from the wavelet decomposition were highly wavelet dependent and a specific wavelet for this task would possibly yield better results [72].

Cho, *et al.,* 2003, used a cohort of 16 probable AD patients and 16 age matched controls. Their method used a two stimulus oddball paradigm with 75% 1 kHz standard sounds and 25% 1.5 kHz target tones. Only a single electrode at location P4 was used for recording. The patients were asked to count the number of times the target tone was presented. The EEG data was divided into 30 second segments and then analyzed to compute 118 features with the following breakdown: 88 power spectral measurements, 28 statistic measurements, 2 chaotic features, and 10 ERP features. Through the use of genetic algorithms, a 35 feature long "chromosome" was created and optimized to include a minimal amount of the most dominant features for use in training a neural network. The final selection of features included 24 spectral features, 8 statistical features, 1 chaotic feature, and 2 ERP features. From the original segmentation, 137 segments from 11 AD patients and 10 normal subjects were used for network training while 72 segments from the remaining 5 AD patients and 6 normal subjects were used for testing. The resulting classification accuracy of the neural network was 73% for AD, 88% for normal with a total classification performance of 81.9% [73].

Abasolo, *et al.*, 2003, used a cohort of 7 AD patients and 7 normal controls. Five minutes of data was recorded using the EEG, which was then broken down into 5 second segments. Analysis was done using the P3 electrode by suggestion of an electroencephalographer. For the analysis, approximate entropy (ApEn) was used, which

quantifies regularity in sequences and time series data by assigning a non-negative number to sequences with more instances of recognizable features or patterns. Results showed that the ApEn values of normal subjects were higher than those of AD patients. This experimentation provides a potential application of ApEn to the diagnosis of AD through the use of the EEG [74].

Abasolo, *et al.*, 2006, continued their 2003 study by expanding their cohort to 11 AD patients and 11 normal controls, as well as adding auto mutual information (AMI) analysis. The AMI provides a measure of both linear and non-linear statistical dependencies between two time series. Their analysis also expanded to include 19 electrode locations. They used this method by looking at the AMI between a given point in the data and a given amount of time later in the data. The results showed that the ApEn values increased in the temporal electrodes and occipital electrodes with the frontal electrodes providing the lowest ApEn values. For the AMI, the rate of decrease was also higher in those regions, which agrees with the higher complexity in the time series as shown by the ApEn. Like the first study, the ApEn was lower in AD patients than control subjects as well as slower AMI decreases in the AD patients [75].

Tao and Tian, 2005, used a cohort of 12 probable AD patients, and 18 MCI patients for a coherence analysis of gamma-band EEG. The EEGs were taken during rest using a protocol of presenting the subject with a random series of odd and even numbers less than 30. The subjects were instructed to count the number of odd rnumbers only. Each recording was performed at 21 different electrode sites. A Mexican hat wavelet was used to extract the gamma band information from the EEGs. Coherence analysis was then done between pairs of electrodes. Results showed a reduction in

coherence in both rest state and during cognitive task. In the MCI patients, there was a significant difference between rest state and cognitive state. This difference showed that the coherence during cognitive processing could be a helpful measure in AD diagnosis [76].

Chapman, *et al.*, 2007, used a discriminant function as their method of classification between early stage AD and normal controls. The cohort included a total of 12 AD patients and 12 normal controls. For the ERP extraction, a number-letter paradigm was used, which is a visual stimuli task using ordering as the active component of the task. Principal component analysis (PCA) was then used to extract eight features from the ERP used in their discriminant analysis. Two different methods of validation were used: a 50/50 split of the data and leave one out cross validation. Performance was high for the 50/50 split with 92% classification accuracy, sensitivity of 100%, and specificity of 83%. For the leave one out cross validation, classification dropped to 79%, sensitivity of 83%, and specificity of 75% [77].

Gomez, *et al.*, 2006, used a cohort of 12 probable AD patients and 12 normal controls. They used the magnetoencephalogram (MEG), which measures the magnetic fields produced by the electrical activity of the brain, using a full 148-channel whole head magnetometer. Again, the AMI was calculated to locate time series dependencies. The AMI of the MEG decreased at a lower rate in AD patients in all channels compared to the control subjects [78].

Henderson, *et al.,* 2006, used two different approaches for classification. With their cohort of 39 AD patients, six vascular dementia patients, and 42 cognitively normal patients, they used both a fractal-dimension based method and a method based on a

probability density function based on zero-crossing intervals for classification. The analysis was done on the EEG itself rather than the ERPs. Promising results showed a specificity of 99.9% and sensitivity of 67% for the fractal-dimension method and 99.9% specificity and 78.8% sensitivity for the probability density function method [79].

CHAPTER 3

METHODS

3.1 PREVIOUS RESEARCH IN THIS STUDY

The previous work on this project included a single cohort of Alzheimer's disease patients and normal control subjects. The finalized cohort totaled 71 patients. The process involved conducting a discrete wavelet transform on the ERPs of each patient and then using the coefficients from the transform as inputs into a classification algorithm. The classification algorithm then provided a diagnosis for the signal being analyzed based on those features.

The initial research stage of the project, as conducted by Genevieve Jacques, included only 32 subjects. The primary objective was to test the effectiveness of an automated classification system based on wavelet features to distinguish between Alzheimer's disease and normal subjects. This stage of the project focused on wavelet selection, specifically the Daubechies wavelet (db4) and quadratic B-spline wavelets. Analysis was also only carried out on the Pz electrode. Overall performance from the different wavelets was in the low to mid 80% range [80].

The project was continued by Apostolos Topalis, who expanded the analysis to include the Cz and Fz electrode responses commonly used in ERP analysis. A Learn++ algorithm was used to create ensembles for classification using Daubechies wavelet coefficients as inputs, eliminating the B-splines. Single trial performance reached 83.1%

with a 5-trial average of 79.2%. The cohort had also been expanded to include all 71 patients [81].

Nicholas Stepenosky further expanded the scope of analysis. Again, the full cohort of 71 patients was used with a subset of 66 patients also analyzed due to troublesome patients. This part of the study began preliminary investigation over all 19 electrode responses, as in Figure 2.1, with a focus on those electrodes from the parietal region. Feature sets were again obtained using the Daubechies wavelet, with 4 vanishing moments, and only the lowest three frequency bands were analyzed, 1-2 Hz, 2-4 Hz, and 4-8 Hz. Again ensemble techniques were employed, this time using decision level fusion using combination rules such as sum rule, product rule, weighted majority voting, and decision templates. Performance reached as high as 84.85% for the 66 patient cohort and 83.1% for the 71 patient cohort [82].

Hardik Gandhi then used a stacked generalization method as his classification algorithm. Again, all 71 patients were used with early analysis using the Fz, Pz, and Cz electrodes, and then expanding to include a second set of electrodes from this work. Ghandi also subdivided the Alzheimer's disease cohort into two separate cohorts, mild and mild-moderate AD, and then conducted a three class classification analysis. For the two class problem, the best performance achieved was 85.65%, while 71.34% was the best classification performance for the three class problem [83].

3.2 CURRENT RESEARCH

This stage of the research includes a continuation of the work done by Stepenosky. The full cohort of 71 patients including all electrode locations were analyzed. First, the

individual electrode responses for each of the three lowest frequency bands were analyzed individually. Based on this analysis, the most informative features were selected for ensemble construction using decision level fusion using sum rule, product rule, weighted majority voting, and Dempster Shafer rule. An exhaustive search through all possible combinations of classifiers also provided the most ideal ensembles and gave insight as to which electrode responses provided the most complementary information when selecting a final set of electrode responses to use for diagnosis.

Additionally for the current work, a second cohort from a new study has been added including Alzheimer's disease patients, Parkinson's patients, MCI patients, Parkinson's disease with dementia patients, and cognitively normal controls. Aside from distinguishing between Alzheimer's disease and normal subjects, the new classes will eventually be used for diagnosis between all classes as well as testing the sensitivity of the algorithm for future prediction of dementia development in individual subjects. A flowchart showing the overview of the process is shown in Figure 3.1. The two cohorts will be referred to as Cohort A for the original 71 AD and normal subjects and Cohort B for the cohort from the new study including 62 AD and normal subjects for the remainder of this thesis.

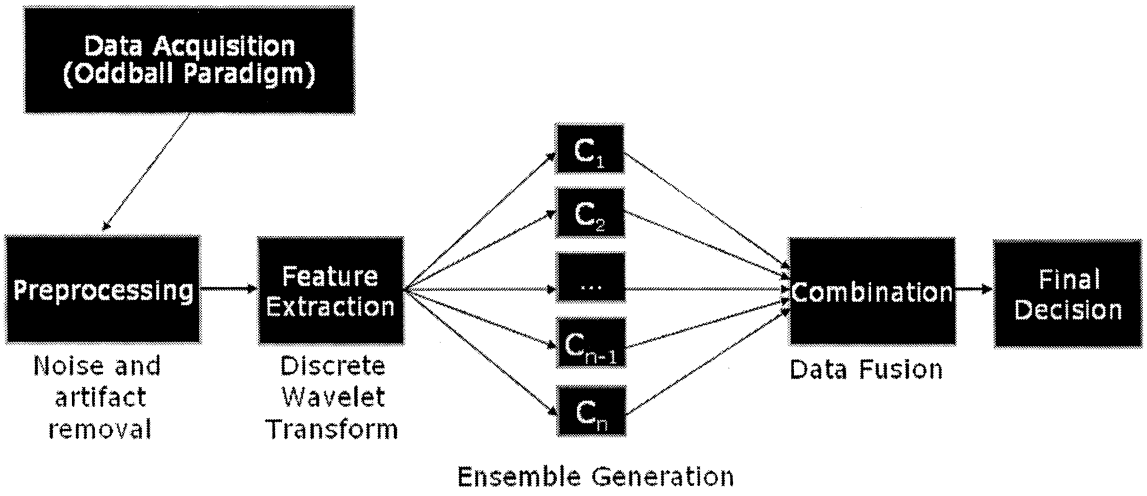

Figure 3.1: Flowchart showing the overview of the approach.

3.3 RESEARCH SUBJECTS

The data used for this work has been received as a collaboration among three institutions. The subjects from Cohort A were first given a neuropsychological examination at the University of Pennsylvania where an initial diagnosis is made by an expert neurologist. This diagnosis is considered the gold standard for our research. The patients were then sent to Drexel University, where their EEGs were collected, cleaned of artifacts, and averaged together to provide ERP signals. The resulting ERP signals were then transferred to us at Dr. Polikar's Signal Processing and Pattern Recognition Laboratory (SPPRL) at Rowan University where the data analysis and automated classification is performed.

The subjects from Cohort B were again given a neurological examination and diagnosis at the University of Pennsylvania. The EEGs were then also acquired at the University of Pennsylvania and the raw EEG data was sent to us at SPPRL at Rowan University. Because there is no intermediary processing at Drexel University, artifact

removal and averaging of the signals was done at SPPRL. The motivation for this is to keep as much of the process as automated as possible to ensure minimal expertise required for implementation on a large scale at community clinics, the only processing necessary for the patients is collection of raw EEG signals.

For Cohort A, a total of 71 subjects provided sufficient usable data. Of these 71 subjects, 34 were diagnosed with Alzheimer's disease and 37 were cognitively normal. For Cohort B, a total of 62 subjects were collected in the study up to this point. The breakdown of diagnoses is 31 AD, 31 normal. All patients in both cohorts were verified to be free of any evidence of other neurological disorders by history or by exam.

During the neurological assessment of the subjects, the Mini-mental State Exam (MMSE) was administered. As mentioned previously, the MMSE tests for memory, language, and praxis skills and is one of the commonly used diagnostic tests for dementia. The test is scored from 0 to 30 with 30 being cognitively normal. Any score lower than 19 indicates obvious impairment. Patients were also tested on the Clinical Dementia Rating (CDR) Scale, which is scored as 0, .5, 1, 2, or 3 depending on the severity of the dementia, with 3 being severely impaired and 0 being cognitively normal. Both of these tests are part of the NINCDS-ADRDA (National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association) criteria for probable AD [84]. For Cohort A, only the AD subjects with the highest MMSE scores were selected, since the motivation of the project was to be able to diagnose AD in its earliest stages. For Cohort B, patients were not excluded based on MMSE scores alone, but rather their ability (or inability) to successfully complete the data acquisition protocol.

For Cohort A, the following inclusion and exclusion criteria were used:

Inclusion criteria for cognitively normal cohort: (i) age > 60; (ii) Clinical Dementia Rating score = 0; (iii) Mini-mental State Exam Score > 26; (iv) no indication of functional or cognitive decline during the two years prior to enrollment based on a detailed interview with the subject's knowledgeable informant.

Exclusion criteria for cognitively normal cohort: (i) evidence of any central nervous system neurological disease (e.g. stroke, multiple sclerosis, Parkinson's disease, etc.) by history or exam; (ii) use of sedative, anxiolytic or anti-depressant medications within 48 hours of ERP acquisition.

Inclusion criteria for AD cohort: (i) age > 60; (ii) Clinical Dementia Rating score $\geq .5$; (iii) Mini-mental State Exam score ≤ 26; (iv) presence of functional and cognitive decline over the previous 12 months based on a detailed interview with a knowledgeable informant; (v) satisfies the NINCDS-ADRDA criteria for probable AD.

Exclusion criteria for AD cohort: Same as those for the cognitively normal controls.

The criteria were modified slightly for Cohort B. There was no restriction on the MMSE score and Parkinson's disease served as its own category, rather than an exclusion condition. Table 3.1 summarizes the demographics from each of the cohorts.

Table 3.1: Cohort details including the number of patients, average ages and standard deviations, and average MMSE scores and standard deviations.

Cohort A (71 Subjects)					
	Number of Subjects	Average Age	Standard Deviation	Average MMSE Score	Standard Deviation
AD	34	74.97	7.09	24.68	2.95
Normal	37	76.14	7.28	29.24	1.19
Cohort B (62 Subjects)					
AD	31	76.36	8.23	19.29	5.85
Normal	31	70.44	7.99	28.59	1.82

3.4 DATA ACQUISITION

Data for this project came in the form of ERP signals acquired through an auditory oddball paradigm. Aside from the artifact removal, each cohort followed the same data acquisition procedure, which is a slightly modified version of the protocol described in [37]. Subjects were equipped with a set of headphones while seated. Before beginning the protocol a hearing threshold was set unique to each patient by presenting them with a 1000 Hz tone and adjusting the volume appropriately to a comfortable level for the subject.

The subject was then presented with a series of tones to acquaint them with the protocol. The tones were 100 ms each including a standard 1 kHz tone, target (oddball) 2 kHz tone, and novel sounds, which included 60 unique environmental sounds digitally recorded and edited to 200 ms duration. Once the subject understood the process, the actual data acquisition began.

A total of approximately 1000 stimuli were presented to the subject in a pseudo-random order. The frequency of the different tones was 65% standard, 20% target, and 15% novel. The inter-stimulus interval was varied from 1.0-1.3 seconds. The patient was instructed to press a button each time they heard only the target tone. Approximately

20 minutes of total EEG was recorded in intervals of 3-5 minutes with 3 minutes of rest in between each set. The entire process lasted roughly half an hour.

The EEG signals were recorded from 19 tin electrodes either embedded in an elastic cap for Cohort A, or applied directly to the patients' scalp for Cohort B. The impedances on the electrodes were limited to 20 kΩ. Signals were amplified and digitized at 256 Hz/channel and stored. All signals were notch filtered at 59-61 Hz to remove electrical noise and baselined with respect to the pre-stimulus interval.

For Cohort A, the technician at Drexel University removed all artifactual recordings from the EEG before averaging the signal. The averaged signal was then calculated based on the stimulus so that there were 3 different mean EEG recordings from each patient. In cases where a large number of stimuli were artifact free, multiple averages were created with 30-85 recordings per average. Averaging was done starting at 200 ms pre-stimulus and ending 800 ms post-stimulus.

For Cohort B, an automated derivative-based artifact removal technique was devised. Each of the raw signals was split into 3 signals based on stimulus. A 20^{th} order derivative was then taken for each point in the signal. A threshold was set to mark any change greater than the threshold as being an artifact, and subsequently removing that recording from the signal. The remaining artifact free signals were then averaged together in the same manner as Cohort A yielding three averaged ERPs, one for each stimuli. An example of the artifact removal is shown in Figure 3.2.. This artifact removal technique is shown by Equation (3.1), which is applied to the entire signal; however, the full recordings are removed, not just the points exceeding the threshold.

$$f(x) = \begin{cases} f(x), & where \quad f(x+20) - f(x) < threshold \\ remove, & where \quad f(x+20) - f(x) \geq threshold \end{cases} \tag{3.1}$$

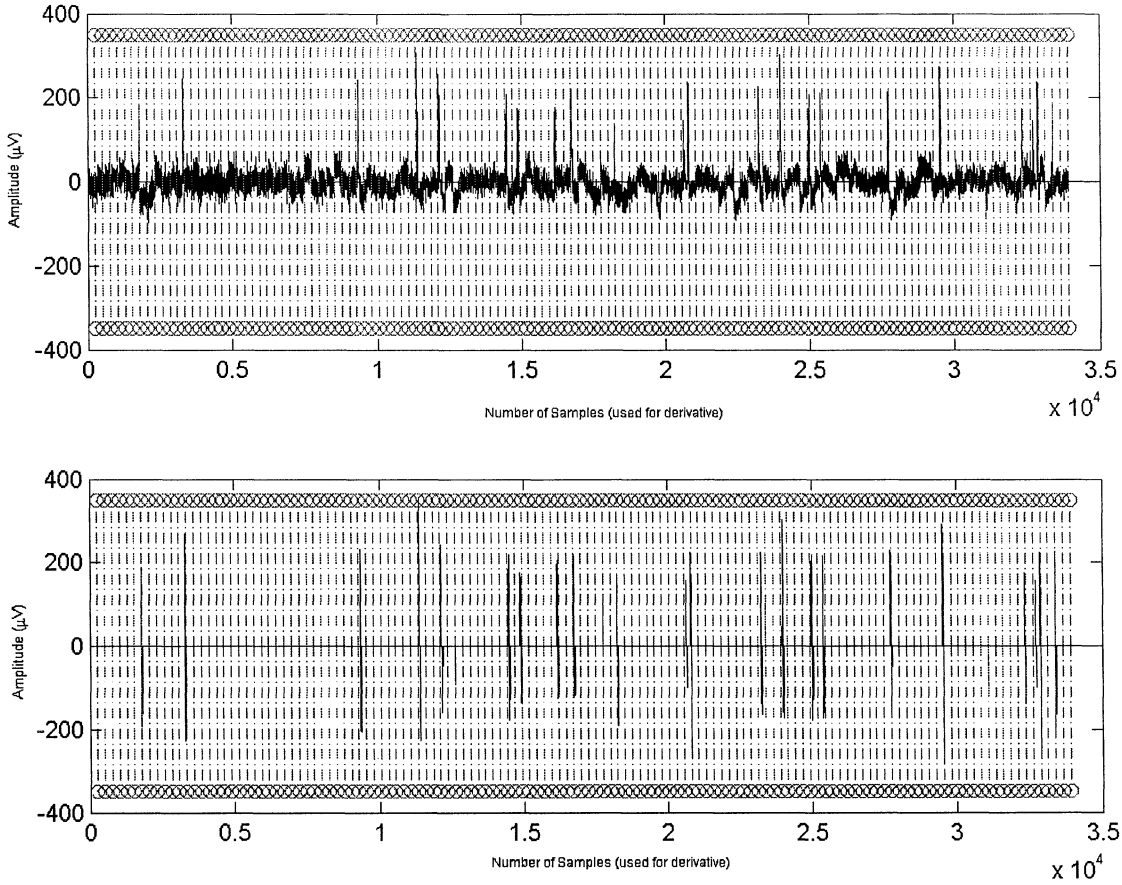

Figure 3.2: Example of artifact removal. The vertical lines break up by signal into 1 second intervals for each stimulus. Any recording exceeding a set threshold results in removal of that entire interval. The top plot shows the signal with artifacts. The bottom plot shows the location of artifacts as identified by the derivative. These recordings were removed..

To view the differences of the signals between the two cohorts, grand averages were created for each cohort. The grand average is an average from every patient in the cohort on each of the electrodes. Figure 3.3 shows a comparison of the Pz electrode from both Cohort A and Cohort B on both novel and target stimuli responses. The overall look of the two different cohorts shows some strong similarities in the shape of the P300 and the differences between normal subjects and AD patients. There are also some minor differences such as the amount of noise present in the averages. For a full comparison,

grand averages from every electrode are shown in Appendix A for Cohort A and Appendix B for Cohort B.

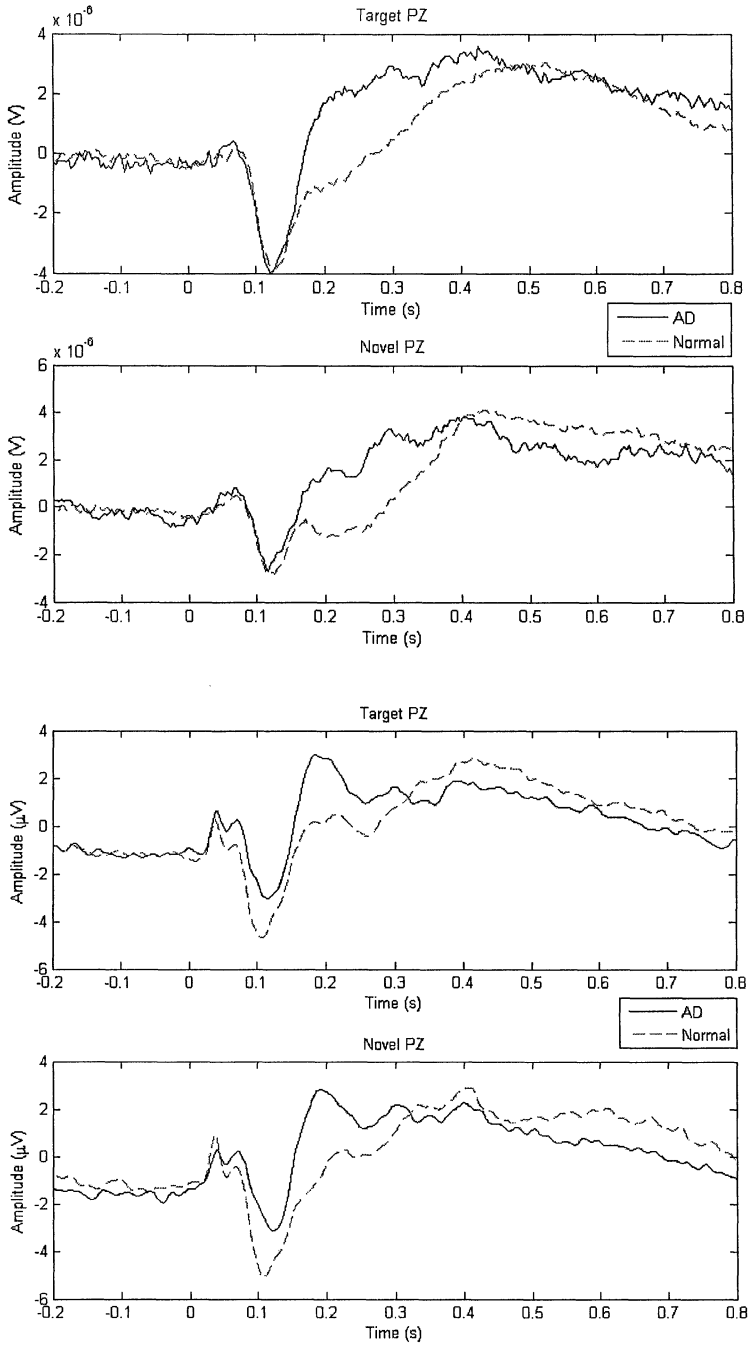

Figure 3.3: Comparison between the data from Cohort A after artifact removal (top) and the data from Cohort B after artifact removal (bottom). Comparison is shown as grand averages from each cohort on the Pz electrode.

3.5 FEATURE EXTRACTION

The EEG, even as an averaged ERP, contains a significant amount information in its original time-dependent form. It can be hard to tell exactly what is useful in the signal as well as how much of the information is redundant or noninformative in the classification process. Based on this, the frequency domain representation of the signal can be quite helpful in extracting discriminating information from the ERP. The frequency domain representation allows for a visualization of the frequency content of the signal, which as described throughout this thesis, is of utmost relevance to EEG signals.

The most extensively used method of creating a frequency domain representation of a signal is through the use of the Fourier transform (FT). The Fourier transform works by decomposing the signal into complex exponential functions at different frequencies, which can then be integrated over an infinite interval to give the frequency representation of the signal. The equations for the FT and inverse FT are shown in equations (3.2) and (3.3), where *x(t)* is the time representation of the signal and *X(f)* is the frequency representation of the signal. The inherent problem with the FT is that when the signal is transformed from time domain to frequency domain, all temporal information is completely lost in the magnitude spectrum. The time information does still remain in the phase response of the FT, however, it does not lend itself in a practical form to use in a time-frequency analysis. This lack of temporal information poses a significant problem when dealing with nonstationary ERPs since the latencies of the ERP component play a significant role in the analysis. Therefore, the FT is insufficient for feature extraction and a time-frequency analysis method is required.

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2j\pi ft} dt \tag{3.2}$$

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{2j\pi ft} dt \tag{3.3}$$

The short time Fourier transform (STFT) attempts to overcome the temporal limitations of the FT through use of windowing. By shifting a window through the signal and multiplying only the section of the signal in the window by the exponential function, a piecewise time-frequency representation can be generated. By adjusting the width of the window, a tradeoff is created between the amount of resolution in the time and frequency domains. This tradeoff is in effect the Heisenberg Uncertainty Principle stating that in a time-frequency representation of a signal, the exact time and frequency cannot be known simultaneously. The equation for the STFT is shown in equation (3.4) where $x(t)$ is again the original time signal and $w(t-\tau)$ is the windowing function shifted by a time τ. A narrow window allows for excellent time resolution, but fairly poor frequency resolution, whereas a wide window allows for excellent frequency resolution but poor time resolution. This effect poses a problem in a very nonstationary signal such as the EEG, since the size of the window remains constant for all calculations of a given STFT [85]. This drawback leads to the need for a different time-frequency analysis tool, the wavelet.

$$STFT_x^{(\omega)}(\tau, f) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-j2\pi t} dt \tag{3.4}$$

3.6 WAVELET TRANSFORMS

The wavelet transform (WT) was first developed in the 1980's by Grossmann and Morlet to analyze geological data, and in over two decades since its development, it has been researched and applied extensively, especially in many areas of pattern recognition. WT is able to address the major shortcoming of the STFT and create a more informative time-frequency representation of a signal through the use of multiresolution analysis (MRA). This shortcoming is overcome by varying the size of the window according to the particular frequency band under current examination. The WT has a few different forms: the continuous wavelet transform (CWT), the wavelet series, which is a discretized version of the CWT, and the discrete wavelet transform (DWT), the latter of which is used for feature extraction in this work [86].

3.6.1 CONTINUOUS WAVELET TRANSFORM

The CWT is similar to the STFT in the way it calculates the time-frequency representation of a signal. Like the STFT, which uses (exponential) basis functions, the CWT also uses basis functions that are multiplied with the function using a window that is shifted through time. The bases account for the first major difference between the two transforms. The STFT applies the FT each time the window is shifted, whereas the CWT does not use the FT, but rather calculates the wavelet transform. The second major difference between the two is the windowing function. As shown in equation (3.4), in the STFT the size of the window remains constant regardless of which frequency is being analyzed.

For the CWT, two new parameters are introduced, scale (which relates to the inverse of frequency) and translation (which relates to time). Using these two parameters, the window function can change in width according to changes in frequency. The general form of the CWT is as follows:

$$CWT_x^{\psi}(\tau,s) = \Psi_x^{\psi}(\tau,s) = \int x(t)\psi^*(t)dt \tag{3.5}$$

where τ is the variable for translation, s is the variable for scale, $x(t)$ is the signal being transformed, and $\psi(t)$ is the transforming function called the mother wavelet, similar to the exponential in the Fourier transform, and * denotes the conjugate.

The mother wavelet refers to the original wavelet function before any alterations in scale or translation. The term wavelet refers to a "short" wave, or a wave where the function is of very limited duration before decaying to zero. Figure 3.4 shows some examples of different wavelets.

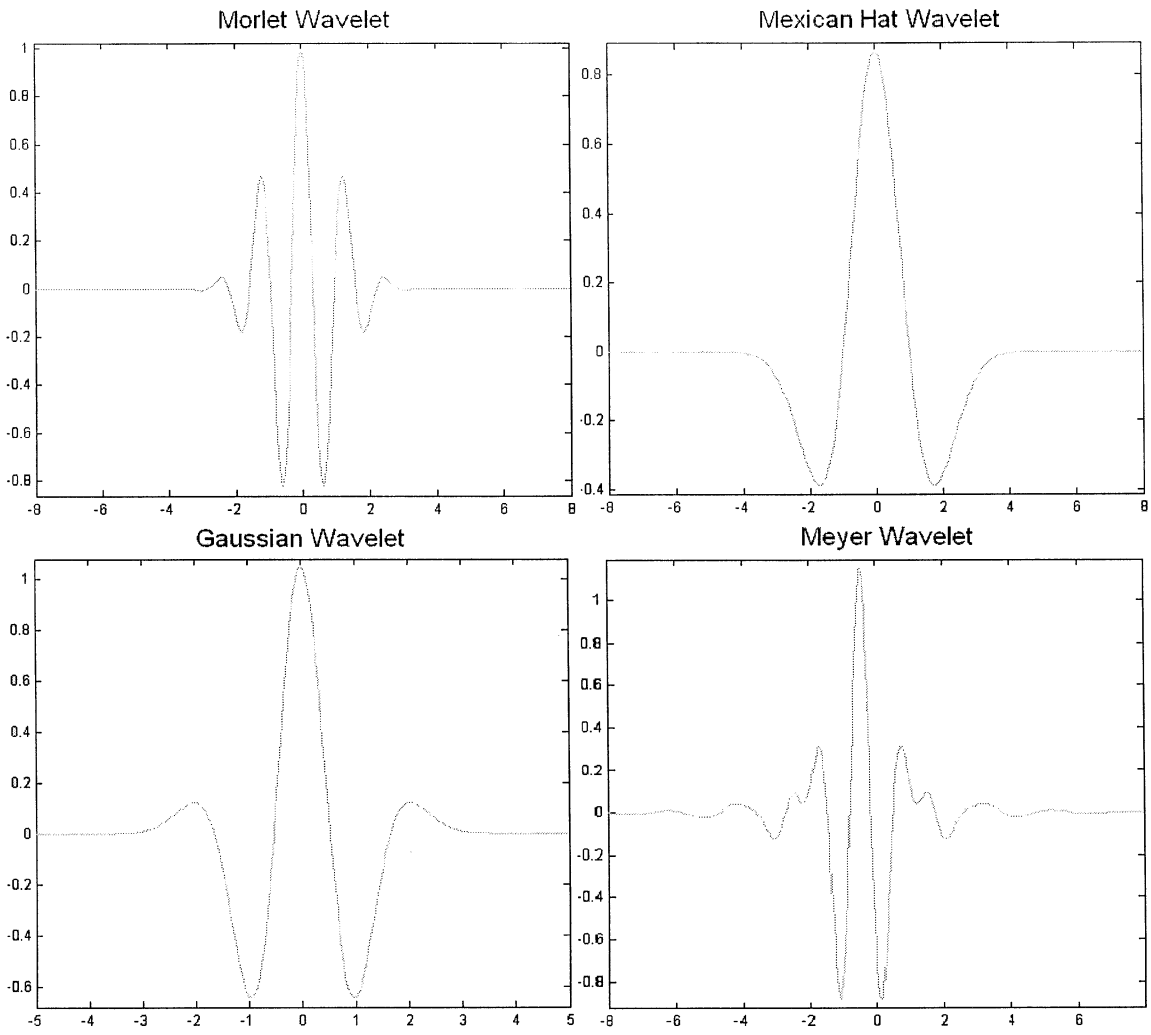

Figure 3.4: Various wavelets.

The wavelet must satisfy two conditions. First, the wavelet must in fact be a wave. To qualify for this, it must be oscillatory. Second, it must be of finite duration by having a finite energy. This second requirement is achieved by nonzero amplitudes for only a short period [86]. The restrictions are shown mathematically as follows:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{3.6}$$

and

$$\int |\psi(t)|^2 \, dt < \infty \tag{3.7}$$

The wavelet function is defined as:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \tag{3.8}$$

where the s and τ variables allow for different dilations (or compressions) and translations of the original mother wavelet. The square root term allows for energy normalization to ensure that the transformed signal has the same energy at any given scale. By varying τ, the location of the wavelet varies along the time axis, while increasing s compresses the width of the wavelet and decreasing s dilates the width of the wavelet [85,87]. An example of compression and dilation, along with translational changes to a wavelet are represented in Figure 3.5.

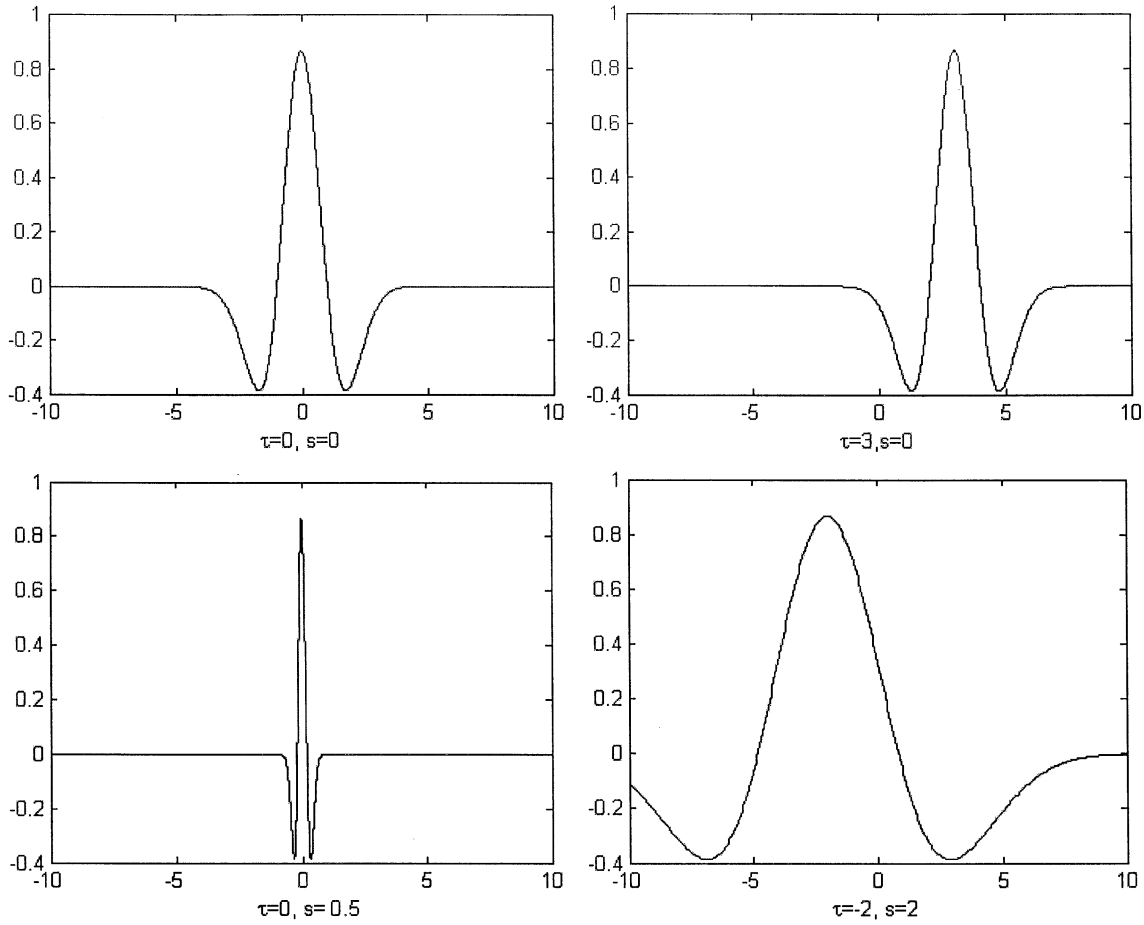

Figure 3.5: Mexican hat wavelet showing the effects of different scales and translations.

By inserting the wavelet function into the equation for the CWT,

$$CWT_x^{\psi}(\tau,s) = \Psi_x^{\psi}(\tau,s) = \frac{1}{\sqrt{s}} \int x(t)\psi^* \left(\frac{t-\tau}{s} \right) dt \qquad (3.9)$$

Equation (3.9) shows how the CWT is computed. The equation is evaluated for each scale s and translation τ. It can be thought of as the inner product of the original time signal $x(t)$ and the wavelet function $\psi(t)$.

The exact process starts with a scale of 1 and translation of 0. The better the wavelet is correlated with the signal at that point, the higher the value obtained from the transform. If the wavelet and the signal do not correlate at that given time, then a low value is given by the transform. To compute the entire CWT, the scale is kept constant

while the translation is incremented until the shifted wavelet reaches the end of the signal. The scale is then increased or decreased and the process is repeated. This process is repeated until all values for s and τ have been calculated. Since this is a continuous transform, all results are integrated over a continuous range rather than summed as will be seen in the wavelet series and discrete wavelet transform. The exact range of the scale values necessary for computation is typically governed by the range of frequencies present in the signal.

3.6.2 WAVELET SERIES

Since almost all computations are done on computers, it is impractical to use a continuous transform. To circumvent this limitation, some form of sampling is necessary for the transform to be calculated. Sampling is done for the wavelet series through dyadic grid sampling of the translation-scale plane. Since each scale represents a frequency band, as the scales become larger, fewer samples of the data are needed to meet the Nyquist criterion; *number of samples > twice the highest frequency*. The exact sampling rate necessary can be found as follows:

$$N_2 = \frac{s_1}{s_2} N_1 = \frac{f_2}{f_1} N_1 \tag{3.10}$$

where N_1 is the sampling rate at scale s_1 and N_2 is the necessary sampling rate for s_2 given that $s_1<s_2$, which relates to frequencies as $f_1>f_2$ [85].

The simplest way to achieve the dyadic grid sampling is by using logarithmic discretization of the scale component and then adjusting step sizes for the translation accordingly[87]. As the scale is increased logarithmically, the sampling rate is simultaneously reduced for the translation resulting in the wavelet equation in (3.11):

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} \psi \left(\frac{t - n b_0 a_0^m}{a_0^m} \right) \tag{3.11}$$

where m is an integer that controls the dilation of the wavelet and n is an integer that controls the translation. It can be seen that the translation is directly related to the scale where the size of the time steps are $b_0 a_0^m$. The most convenient values for a_0 and b_0 are $a_0 = 2$ and $b_0 = 1$. When inserted into equation (3.11), this results in:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{2^m}} \psi \left(\frac{t - n2^m}{2^m} \right) \tag{3.12}$$

For the wavelet series, the wavelet function is required to be either orthonormal, biorthogonal or frame. To satisfy the orthonormal case, the wavelets must be orthogonal to each other and normalized to have unit energy:

$$\int_{-\infty}^{\infty} \psi_{m,n}(t) \psi_{m',n'}(t) dt = \begin{cases} 1, & \text{if } m = m' \text{ and } n = n' \\ 0, & \text{otherwise} \end{cases} \tag{3.13}$$

The analysis and synthesis wavelets then become the same as shown:

$$\Psi_x^{\psi_{j,k}} = \int x(t) \psi^*_{j,k}(t) dt \tag{3.14}$$

or

$$x(t) = c_\psi \sum_j \sum_k \Psi_x^{\psi_{j,k}} \psi_{j,k}(t) \tag{3.15}$$

where c_ψ is a constant that depends on the wave used and * denoted the conjugate. For the biorthogonal and frame cases, equations (3.14) becomes:

$$\Psi_x^{\psi_{j,k}} = \int x(t) \widehat{\psi^*_{j,k}}(t) dt \tag{3.16}$$

[85,87].

3.6.3 DISCRETE WAVELET TRANSFORM

While the wavelet series transform does compute the CWT in a discrete manner that can be executed on a computer, it has one inherent drawback. The wavelet series is not a true discrete transform and leaves significant redundant information in the signal after being transformed. The solution for redundancy removal is the discrete wavelet transform (DWT), which provides non-redundant information for decomposition and reconstruction of the original signal.

The origins of the DWT stem from research completed in 1976. Crosier, Eseban, and Garland were working on a method to decompose discrete time signals, while in the same year Crochiere, Webber, and Flanagan were doing similar work on the coding of speech signals, which was called subband coding [88,89]. Later in 1983, Burt created a similar technique and called it pyramidal coding, now known as multiresolution analysis [90]. In 1989, Vetterli and Le Gall improved the subband coding technique by removing some redundancy issues found in the pyramidal coding technique [91]. These techniques allow for a completely discrete version of the wavelet transform.

3.6.4 MULTIRESOLUTION ANALYSIS

The idea behind multiresolution analysis is to analyze a function at various levels of approximation or resolution. This analysis allows a complicated function to be divided into several simpler functions that can then be analyzed individually. Specifically, time localization of spectral components is possible using this technique. During the discretization process, there is also the ability to fluctuate sampling rates allowing for significant reduction in the number of data points necessary to represent the signal [92].

The process begins by sampling a function. The function is then sampled again with the step size between samples doubled. This process is continued for a set number of samplings with the final sampling of the data being the coarse approximation of the function and information removed from each of the previous samplings being the details at each level of the function.

Mathematically, each time the function is sampled, the approximation of the signal is assigned to the approximation subspace $\mathbf{A_s}$, while the information that is removed is added to the wavelet subspace, $\mathbf{W_s}$. $\mathbf{A_s}$ is generated by the base:

$$\phi_{k,s} : 2^{s/2}\phi(2^s t - k); k \in Z \tag{3.17}$$

while $\mathbf{W_s}$ is generated by the base:

$$\psi_{k,s} : 2^{s/2}\psi(2^s t - k); k \in Z \tag{3.18}$$

Any function $x_s(t)$ and $y_s(t)$ can then be represented as linear combinations of $\phi_{k,s}(t)$ and $\psi_{k,s}(t)$.

A scaling function with finite energy is used, as shown in (3.19), that generates a nested sequence $\{\mathbf{A_j}\}$ as shown in (3.20).

$$\phi(t) \in L^2(\mathbb{R}) \tag{3.19}$$

$$\{0\} \leftarrow \ldots \subset A_{-1} \subset A_0 \subset A_1 \subset \ldots \rightarrow L^2 \tag{3.20}$$

This scaling function must also satisfy a dilation equation:

$$\phi(t) = \sum_k g_0[k]\phi(at - k) \tag{3.21}$$

where $a > 0$, and typically chosen to be 2. The scaling function, $\phi(t)$, is a superposition of scaled and translated version of itself. From this dilation function, it can be seen that:

$$x(t) \in A_s \Leftrightarrow x(2t) \in A_{s+1} \tag{3.22}$$

$$x(t) \in A_s \Leftrightarrow x(t+2^{-s}) \in A_s \tag{3.23}$$

For each s, $\mathbf{A_s}$ is a subspace of $\mathbf{A_{s+1}}$. $\mathbf{W_s}$ is complementary to $\mathbf{A_s}$ and when combined with $\mathbf{A_s}$, it forms $\mathbf{A_{s+1}}$, meaning:

$$A_s \cap W_s = \{0\} \tag{3.24}$$

$$A_s \oplus W_s = A_{s+1} \tag{3.25}$$

where $\oplus$ represents the direct summation.

The method which these subspaces are created can be related to $x(t)$ and $y(t)$ such that $x_s(t) \in \mathbf{A_s}$ and $y_s(t) \in \mathbf{W_s}$. These functions can be calculated explicitly as:

$$x_s(t) = \sum_k a_{k,s} \phi(2^s t - k) \tag{3.26}$$

$$y_s(t) = \sum_k w_{k,s} \psi(2^s t - k) \tag{3.27}$$

To apply multiresolution analysis within the DWT, the subband coding technique is implemented.

3.6.5 SUBBAND CODING

The method of implementing MRA discretely for the DWT is through subband coding using digital filters. The filters used are quadrature mirror filters, which are half-band lowpass and half-band highpass filters that are odd indexed, alternated, reversed versions of each other. First, a lowpass filter with an impulse response of $h[n]$ is selected, which is determined by the wavelet being used. To create the corresponding highpass version of the filter, the following conversion is performed on the impulse response of the lowpass filter:

$$g[L-1-n] = (-1)^n h[n] \tag{3.28}$$

where $g[n]$ is the highpass filter, $h[n]$ is the lowpass filter, and L is the length of the filter coefficients.

The transformation process begins with the original signal being filtered by the lowpass filter. This filtering is done in discrete time by convolving the signal with the impulse response $h[n]$:

$$x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n-k] \tag{3.29}$$

Since this process is done in discrete time, angular frequency will be used and related to Hertz later. According to the Nyquist rule, the signal must be sampled at 2π if the highest frequency found in the signal is π, that is, twice the highest frequency present in the signal. After applying the half band filter, the highest frequency left in the signal is now $\pi/2$, yet the sampling frequency is still 2π, four times the highest frequency. Half the points in the signal remain as redundant information that can be removed. The redundant information removal is achieved by subsampling the signal by two, or simply removing every other sample in the signal. This subsampling leaves the sampling rate at π, which still meets the Nyquist criteria. Additionally, by reducing the sampling frequency of the signal, the scale of the signal is increased by a factor of 2.

After the lowpass filter is applied, the resulting frequency band of the signal is 0-$\pi/2$. The frequency range from $\pi/2$-π is obtained using the same procedure with the highpass filter. Like the result of the lowpass convolution, the highpassed signal is also subsampled by 2. The results from each filtering operation are shown below:

$$y_{low} = \sum_{n} x[n] \cdot h[2k-n] \tag{3.30}$$

$$y_{high} = \sum_{n} x[n] \cdot g[2k - n] \tag{3.31}$$

These equations relate to MRA in the form of the two-scale equation, shown in (3.21). The outputs of the highpass filter become level 1 detail coefficients. The outputs of the lowpass filter are the level 1 course approximation coefficients of the signal with the higher frequencies removed.

This process is then repeated on the output of the lowpass filter using the same high and lowpass filters. The second set of filters results in the level 2 detail coefficients and an even courser representation of the original signal. Filtering is continued until there is only one sample left after subsampling. All detail coefficients and the final level of approximation coefficients represent the DWT of the original signal. For each of the levels of the decomposition, the time resolution is reduced by half due to the subsampling and the frequency resolution is doubled since the bandwidth is halved. A flowchart of the process is shown in Figure 3.6.

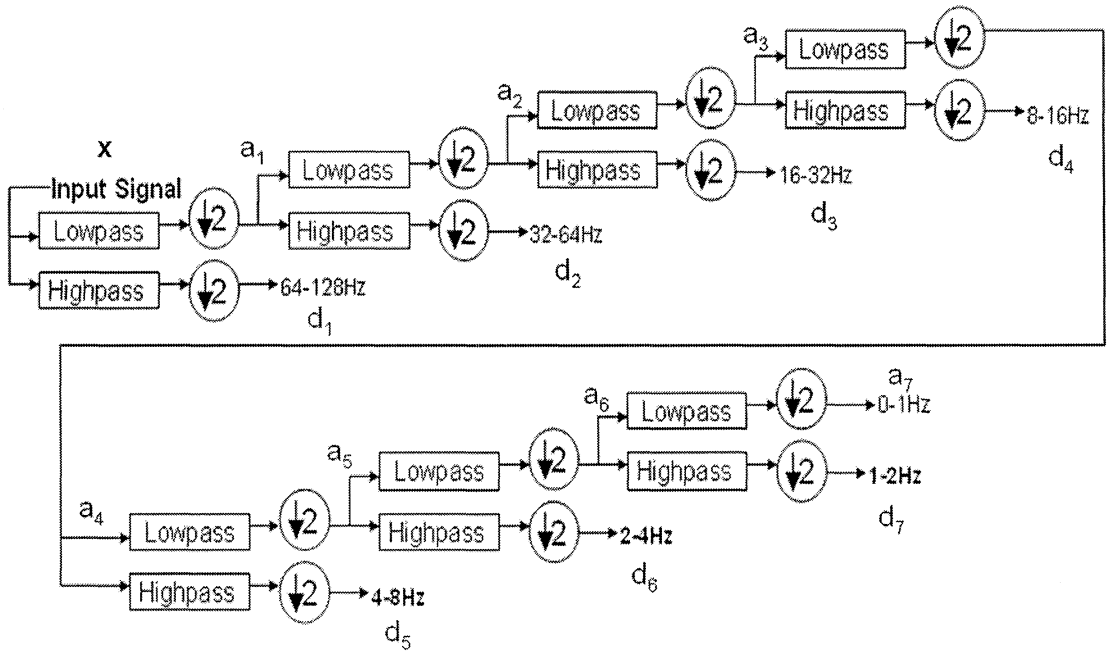

Figure 3.6: Flow of an ERP signal through the discrete wavelet process.

3.6.7 RECONSTRUCTION OF THE SIGNAL

To create a discrete version of the wavelet transform, an inverse transform must exist that allows perfect reconstruction of the original signal. An approximation of the signal at any level can be reconstructed after the entire DWT has been taken. At any intermediate level of reconstruction, two signals will be involved: the course approximation at that level and a summation of the detail coefficients up to that level. These two signals can be combined together to create the approximate reconstruction for that level. This reconstruction can be written as:

$$x_s(t) + y_s(t) = \sum_k a_{k,s}\phi_{k,s}(t) + \sum_k w_{k,s}\psi_{k,s}(t) = x_{s+1}(t) \tag{3.32}$$

where s is the scale level of reconstruction, a are the approximation coefficients, w are the detail coefficients and ϕ and ψ are the scale and wavelet functions used.

By using the filter implementation, the resulting discrete conversion leads to:

$$x[n] = \sum_{k=-\infty}^{\infty} (y_{high}[k] \cdot g[-n+2k]) + (y_{low}[k] \cdot h[-n+2k]) \tag{3.33}$$

where $g[n]$ is the highpass filter, $h[n]$ is the low pass filter, and y_{high} and y_{low} are the respective outputs from those filters. A complete reconstruction including all levels of detail coefficients would require an additional summation over s. Supposing ideal half band filters are used, perfect reconstruction can be obtained.

3.6.8 DAUBECHIES WAVELET

In 1987, Ingrid Daubechies created a wavelet that allows for perfect reconstruction of the original signal. These wavelets are referred to as compactly supported orthonormal

wavelets, and allow for the discrete transformation process. To accurately break down a signal into its frequency components, the wavelet must match the frequency characteristics of the signal. The better the wavelet matches the signal, the better the structures of interest can be localized. The Daubechies family of wavelets provides a smooth wavelet function through the use of vanishing moments, which relate to the order of the wavelet. Increasing the number of vanishing moments in the wavelet allows a smoother wavelet function capable of decomposing more complicated signals. The Daubechies wavelets have one vanishing moment for every order of wavelet (ie. 4^{th} order wavelet has 4 vanishing moments). The wavelet order is abbreviated in the Daubechies family as dbN, where *N* is the order of the wavelet [93]. A comparison of how the vanishing moments affect the wavelet function is shown in Figure 3.7.

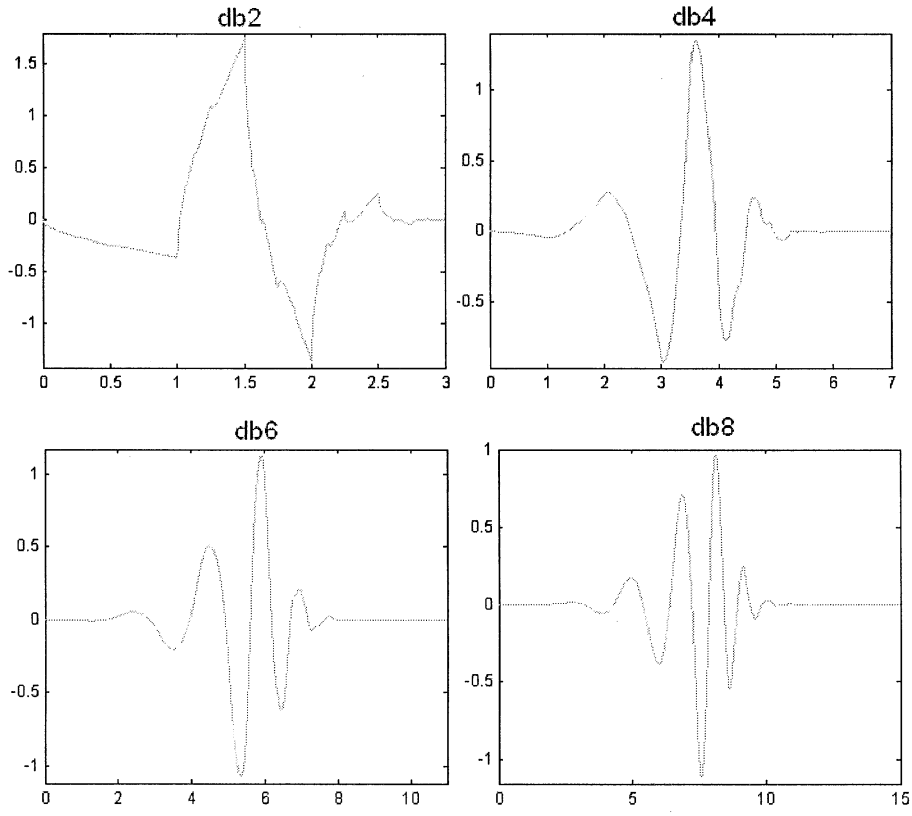

Figure 3.7: Daubechies wavelets with 2, 4, 6, and 8 vanishing moments.

The Daubechies wavelets also work well for ERP analysis, as can be found in numerous studies implementing this particular wavelet. For each of the vanishing moments in the wavelet, two filter coefficients are required for both the high and low pass filters to generate the smoothness and resolution necessary. For this research, the db4 wavelet has been the wavelet of choice primarily due to its length in comparison to the length of the signals being analyzed (8 sample filter to analyze a 256 sample signal). The wavelet and scaling functions are shown in Figure 3.8, and an example of the wavelet decomposition is shown in Figure 3.9 comparing the grand averages from normal subjects and AD patients from Cohort B (Figure 3.3).

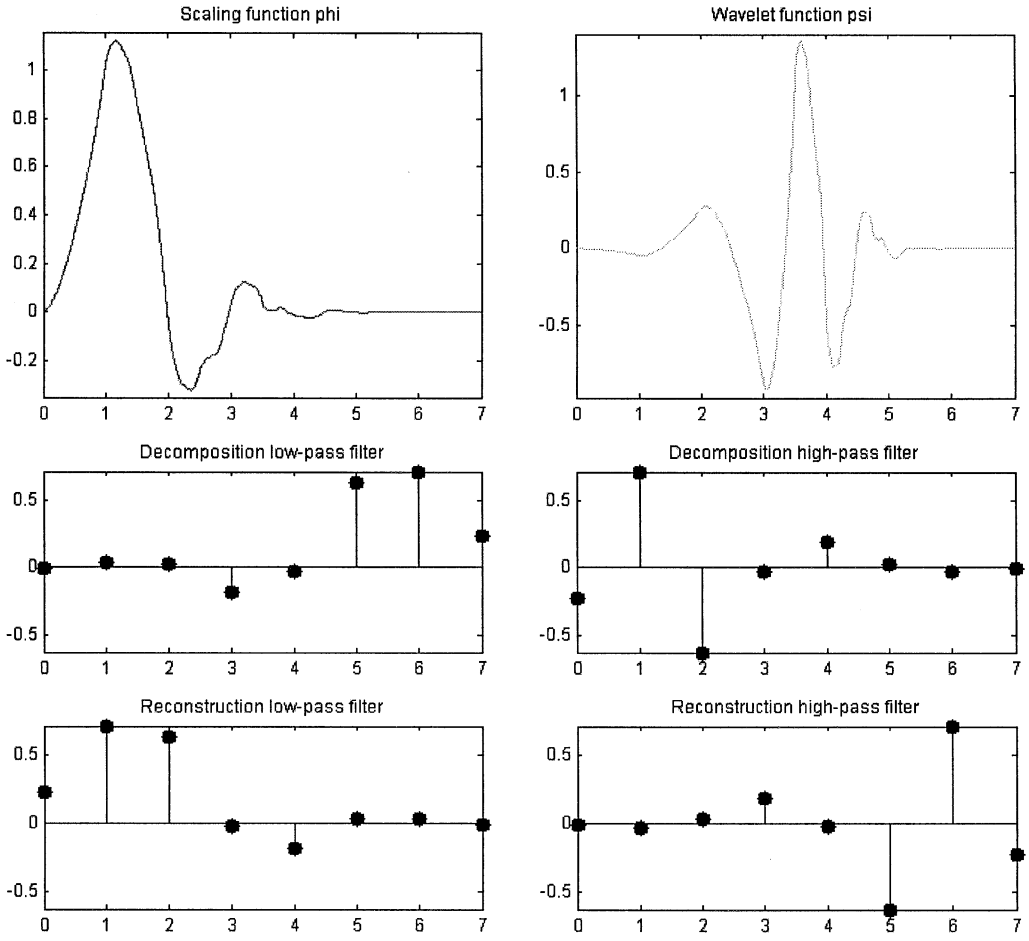

Figure 3.8: db4 scaling and wavelet functions along with decomposition and reconstruction filters.

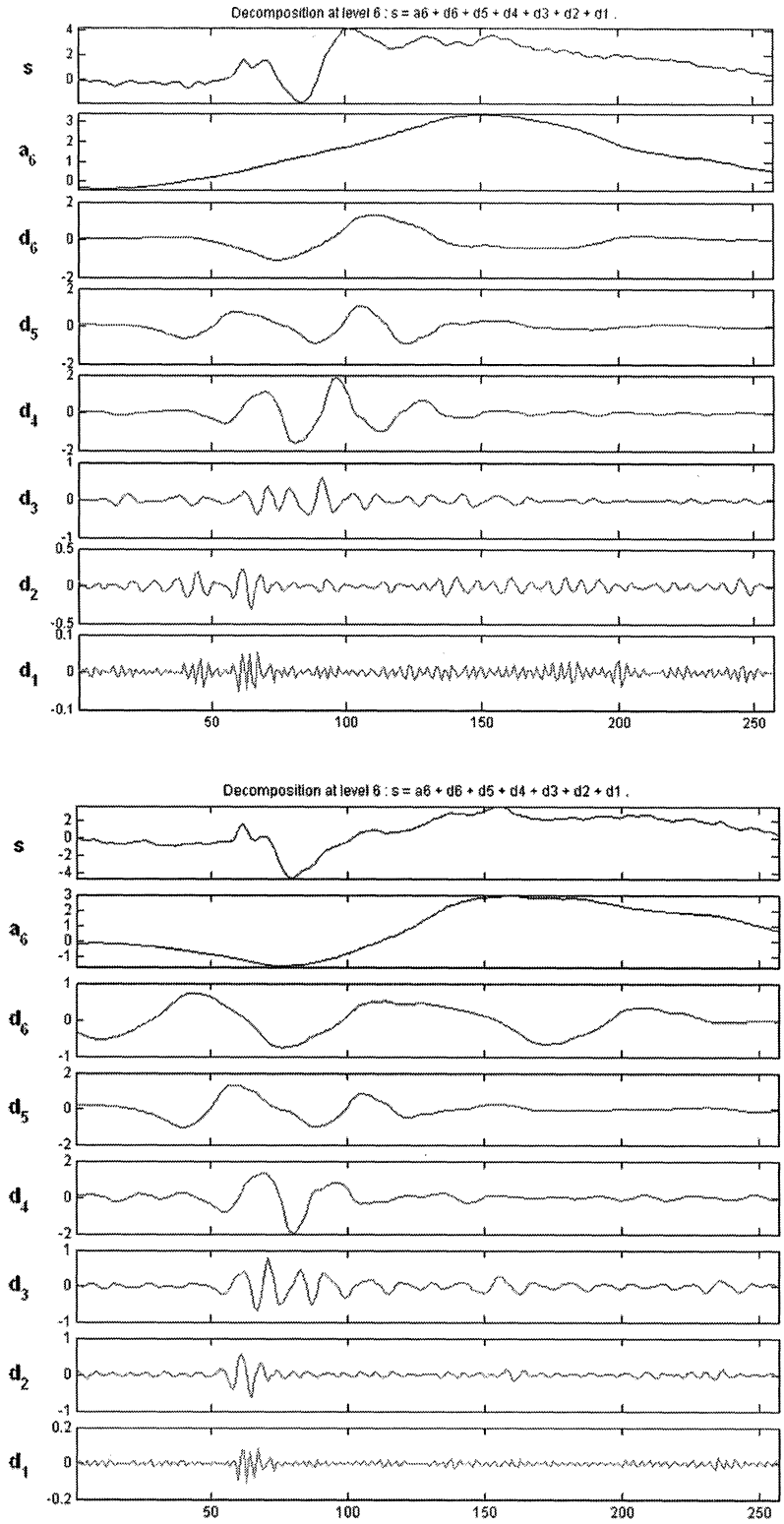

Figure 3.9: Wavelet decomposition of a subject's ERP signal. Decomposition shows the original signal (top), the approximation coefficients a6 (2nd from top) and all levels of detail (d1 through d6).

3.7 CLASSIFICATION AND PATTERN RECOGNITION

To create an automated diagnostic tool, a classification system must be implemented. The general idea behind classification is extraction of key distinctive features, which are then used as inputs to a pattern recognition algorithm used to detect patterns in the features and make a classification. There are numerous methods of feature extraction including filtering, Fourier transform, MRA, and wavelets and just as many different pattern recognition techniques. One of the most popular families of pattern recognition tools is the neural networks, which have many different subsets. Neural networks take the features extracted as input data into the algorithm, which then trains a set of adjustable parameters of the neural network such as the weights of a multilayer perceptron. The trained network, referred to as a classifier, can then generate a class decision for previously unseen data used as input.

3.7.1 MULTILAYER PERCEPTRONS

The MLP is a feedforward artificial neural network consisting of a series of layers, where each layer consists of a series of nodes. The general structure is shown in Figure 3.10. The first layer is a group of input nodes on the left side and the signal then propagates to the right (forward) through one or more hidden layers of the network and finally to the output layer. The initial input layer performs no calculations on the data, but rather serves as input to the following layers. The remaining nodes are used to perform calculations to establish a set of synaptic weights, which can then be applied to any data passing through the network. These weights are updated using the error back propagation

algorithm. This algorithm consists of four general steps, initialization, presentation of training data, forward computation, and backward computation, where the forward and backward computation are iterated many times with each iteration representing a single epoch. These steps are described in detail below [94]:

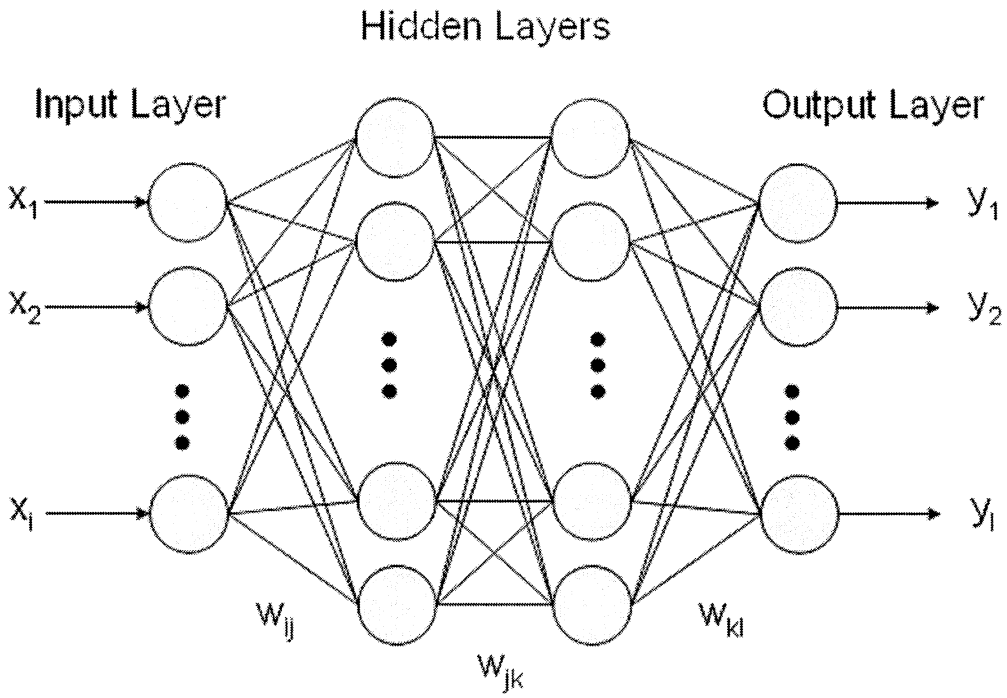

Figure 3.10: General model of a multilayer perceptron network.

- **Initialization** – Before any data is input into the MLP, the synaptic weights are initialized from a uniform distribution whose mean is zero. The variance is chosen to make the standard deviation of the induced local fields of neurons lie at the transition part of the sigmoid activation function, between the linear and saturated parts. For this work, the logarithmic sigmoid was used for the activation function.

- **Presentation of Training Data** – A single epoch of training data is first input into the network. The network then performs a forward and backward computation for each example in the training set.
- **Forward Computation** – For the computation, a single training example is denoted as $(\mathbf{x}(n),\mathbf{d}(n))$, where $\mathbf{x}(n)$ is the input vector at the first layer of nodes and $\mathbf{d}(n)$ is the correct class presented to the output layer of nodes. The induced local fields and function signals of the MLP are then computed on the forward pass through the network, one layer at a time. The induced local field represented as $v_j^{(l)}(n)$ for neuron j in layer l is:

$$v_j^{(l)}(n) = \sum_{i=0}^{m_l} w_{ij}^{(l)}(n) y_i^{(l-1)}(n) \tag{3.34}$$

where m_l is the size of layer l and ij means from the output of neuron I to the input of neuron j, and $y_i^{(l-1)}(n)$ is the output signal of neuron i in the previous layer l-1 at iteration n. For $i=0$, $y_0^{(l-1)}(n)=+1$ and $w_{j0}^{(l)}(n)=b_j^{(l)}(n)$ is the bias applied to neuron j in layer l. Based on the use of the sigmoid function, the output signal from neuron j in layer l is:

$$y_j^{(l)} = \varphi_j(v_j(n)) \tag{3.35}$$

where the sigmoid function, represented by φ is:

$$\varphi(n) = \frac{1}{(1+e^{-n})} \tag{3.36}$$

In the case where $l=1$, or the input layer:

$$y_l^{(0)}(n) = x_j(n) \tag{3.37}$$

where $x_j(n)$ is the jth element of the input vector $x(n)$.

For the case where l is the output layer:

$$y_j(n) = o_j(n) \tag{3.38}$$

where L is the depth of the network.

The error signal can then be computed based on the difference of ***d(n)*** and ***o(n)***:

$$e_j(n) = d_j(n) - o_j(n) \tag{3.39}$$

Backward Computation – Once the error signal is generated, the backward computation begins. The error can then be used to find the instantaneous error energy in Equation (3.40) when C is the set of all possible output nodes.

$$\varepsilon(n) = \tfrac{1}{2}\sum_{j \in c} e_j^2(n) \tag{3.40}$$

The objective then becomes to minimize this error by adjusting the weights of the nodes throughout the system using equation (3.41).

$$v_j(n) = \sum_{i=1}^{m} w_{ji}(n) y_j(n) \tag{3.41}$$

The signal appearing at the output of neuron j is then computed using the activation function as shown in Equation (3.42) where Equation (3.43) shows the activation function.

$$y_j(n) = \varphi(v_j(n)) \tag{3.42}$$

$$\varphi(v_j) = \frac{1}{1 + e^{-av_j}} \tag{3.43}$$

To be able to minimize this error, the gradient of the error with respect to the weights must be calculated. This calculation is done using the steepest decent

algorithm in which steps are taken of length of η in a direction negative to the gradient as shown in equation (3.44).

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} = \frac{\partial \varepsilon(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} = -e_j(n)\varphi_j{}'(v_j(n))y_j(n) \qquad (3.44)$$

The weight updates are then done as shown in Equation (3.45) and (3.46) for an output node j assuming a logistic sigmoid activation function where gradient δ(n) is the derivative of Equation (3.44) for Equation (3.45).

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \qquad (3.45)$$

$$\delta_j(n) = a\left[d_j(n) - y_j(n)\right] y_j(n)\left[1 - y_j(n)\right] \qquad (3.46)$$

For a hidden layer, the gradient function takes a slightly different form, that of Equation (3.47).

$$\delta_j(n) = -\frac{\partial \varepsilon(n)}{\partial v_j(n)} = -\frac{\partial \varepsilon(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \qquad (3.47)$$

where

$$\frac{\partial \varepsilon(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} \qquad (3.48)$$

and

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n) \qquad (3.49)$$

The gradient for each layer is calculated based on the gradient for the subsequent layer:

$$\delta_j(n) = \varphi_j{}'(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \qquad (3.50)$$

The synaptic weights for each layer are then adjusted according to the generalized delta rule:

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha[w_{ji}^{(l)}(n-1)] + \eta\delta_j^{(l)}(n)y_i^{(l-1)}(n) \qquad (3.51)$$

where η is the learning rate parameter and α is the momentum constant.

The forward and backward computation are then iterated by presenting new epochs of training examples to the network until either the maximum number of epochs or the error goal is reached.

3.7.2 SUPPORT VECTOR MACHINES

Despite the fact that MLP's are the most popular classifiers used in many pattern recognition and classification applications, they do have some drawbacks. The training and testing phases can be extremely computationally expensive. If a large number of epochs are used for the back propagation scheme during the training phase, complex classification problems can take hours or even days to complete. The weights of the hidden layer nodes are randomly initialized each time a new MLP is created. As a result, two MLP's being trained on the same set of data will each formulate distinct decision boundaries leading to possible conflicts in classification on test data. For this reason, multiple trials are necessary to be averaged together to gain an overall performance for a given training and test set. The initial randomization could also lead to the algorithm settling on a local minimum, rather than the correct global minimum. These drawbacks are not meant to be an exhaustive list.

The solution to these inherent problems of the MLP comes in the form of the support vector machine (SVM). SVMs are intended to provide classification in a

computationally efficient manner by using high dimensional space. In this space, hyperplanes are created and optimized to allow for maximal margins between classes. As a result, the same decision boundaries are created each time on a given set of training data in significantly less time than it takes to create an MLP.

3.7.2a VC DIMENSION AND RISK

In a typical classification algorithm, the goal of the training phase is to minimize the error on the training data. The best possible classifier is the one that minimizes the expected risk (or error) of the training function:

$$R(\alpha) = \int \frac{1}{2} | y - f(\mathbf{x}, \alpha) | \, dP(\mathbf{x}, y) \tag{3.52}$$

where $f(\mathbf{x}, \alpha)$ is a set of functions for a given input **x** and choice of α, where α is the corresponding weights or biases of the neural network, y is the corresponding true class of **x**, and $P(\mathbf{x},y)$ is the probability distribution, which is not necessarily known *a priori*. Since the probability distribution is not usually known, the empirical risk is instead used:

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^{l} | y_i - f(\mathbf{x}_i, \alpha) | \tag{3.53}$$

where ρ is the number of instances in the training data. As ρ grows towards infinity, $\mathbf{R}_{emp}$ approaches the expected risk, **R**. If ρ remains relatively small, as in this study, the empirical risk can fail to approach **R.**

Rather than focusing on the error of the classifier, focus is placed on the Vapnik-Chervonenkis (VC) dimension, which is a combination of reducing parameters and retaining maximum classification of the classifier. The VC dimension describes the maximum number of points that can be shattered by a given $f(\alpha)$, where shattering is the

ability to assign all possible class labels for a given number of data points. Shattering is limited by the number of dimensions such that in d dimensional space, maximally $d+1$ points can be shattered through the use of hyperplanes.

The VC dimension can then be used to create a relationship between the expected error and the empirical error:

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\log(2\rho/h)+1) - \log(\eta/4)}{\rho}} \tag{3.54}$$

where η is chosen between 0 and 1, and h is the VC dimension. This relation can then be used to minimize the upper bound of the expected risk with probability of $1-\eta$:

$$\frac{m}{4l} \leq \ln(2\rho/m) + 1 - (1/m)\ln(\eta/4) \tag{3.55}$$

where m is a number of training instances such that $m \leq l$.

The VC dimension can then be used along with the empirical risk to choose the best possible function to serve as a classifier. Structural risk minimization (SRM) states that the best choice of classifiers is the one that minimizes the sum of the empirical risk and VC confidence. SRM allows the classifier to accurately describe the training data, but not be overfitted to the data, which leads to the main strength of SVMs, in that the margin between data classes are maximized.

3.7.2b SVM CLASSIFIERS

The SVM classifier works by maximizing the boundaries between classes. In the simplest, linearly separable case, a hyperplane is used satisfying:

$$\mathbf{w}^T\mathbf{x} + b = 0 \tag{3.56}$$

where **w** is normal to the hyperplane, ||**w**|| is the Euclidean norm of **w** and $|b|/||\mathbf{w}||$ is the perpendicular distance from the hyperplane to the origin as shown in Figure 3.12. Class labels are encoded such that $y_i \in [-1,1]$. If d_+ and d_- are the distances from the hyperplane to the closest positive and negative data points, the maximal margin between the classes is $d_+ + d_-$. All points must then satisfy the following conditions:

$$\begin{aligned} \mathbf{x}_i \cdot \mathbf{w} + b \geq +1 \quad & for \quad y_i = +1 \\ \mathbf{x}_i \cdot \mathbf{w} + b \leq -1 \quad & for \quad y_i = -1 \end{aligned} \tag{3.57}$$

which can be combined to be:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \quad \forall i \tag{3.58}$$

The margin is then defined as:

$$m = 2 / || \mathbf{w} || \tag{3.59}$$

as shown in Figure 3.11.

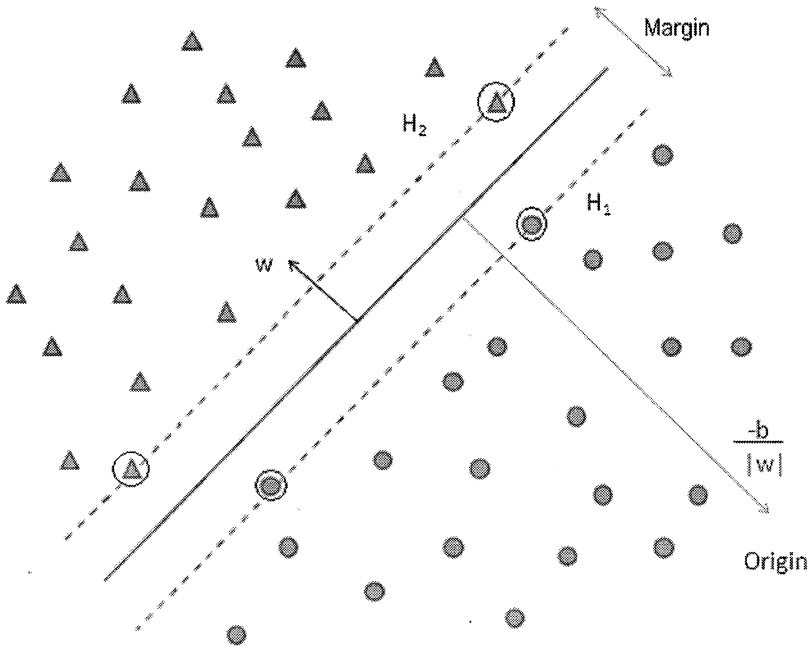

Figure 3.11: Example of two linearly separable classes and the maximal margin between them [95].

The instances that are circled are those that lie on the two hyperplanes separating the classes. These instances are referred to as the support vectors, since changing them will completely change the decision boundary. To optimize the inequalities, Lagrangian multipliers, α_i, are introduced, one for each constraint, such that the primal Lagrangian is:

$$L_P \equiv \frac{1}{2} \| \mathbf{w} \|^2 - \sum_{i=1}^{l} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{l} \alpha_i \tag{3.60}$$

L_P is then minimized with respect to **w**, b and simultaneously the derivatives of L_P must vanish with respect to α, which results in:

$$\begin{aligned} \mathbf{w} &= \sum_i \alpha_i y_i \mathbf{x}_i \\ \sum_i \alpha_i y_i &= 0 \end{aligned} \tag{3.61}$$

By substituting (3.61) into (3.60), the dual formula if te Lagrangian is obtained:

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \tag{3.62}$$

The support vectors lying on either of the defining hyperplanes have Lagrangians where α_i>0, whereas all other points will have α_i=0 and lie on either side of the margin.

For any optimization problem with constraints, the Karush-Kuhn-Tucker (KKT) conditions must be met. For the Langrangian, the conditions apply as follows:

$$\begin{aligned} &\frac{\partial}{\partial w_v} L_P = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \;\; v = 1, \ldots, d \\ &\frac{\partial}{\partial b} L_P = -\sum_i \alpha_i y_i = 0 \\ &y_i (\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \;\; i = 1, \ldots, l \\ &\alpha_i \geq 0 \;\; \forall i \\ &\alpha_i (y_i (\mathbf{x}_i \cdot \mathbf{w} + b) - 1) = 0 \;\; \forall i \end{aligned} \tag{3.63}$$

The above definitions ensure that the classes are linearly separable. Most real life problems, however, are not completely linearly separable, or even separable at all. To allow for non-separability in the data, slack variables are introduced. The constraints from (3.57) become:

$$\begin{aligned} &\mathbf{x}_i \cdot \mathbf{w} + b \geq +1 - \xi \quad &for \;\; y_i = +1 \\ &\mathbf{x}_i \cdot \mathbf{w} + b \leq -1 + \xi \quad &for \;\; y_i = -1 \\ &\xi_i \geq 0 \;\; \forall i \end{aligned} \tag{3.64}$$

The effects of the slack variables are shown in Figure 3.12.

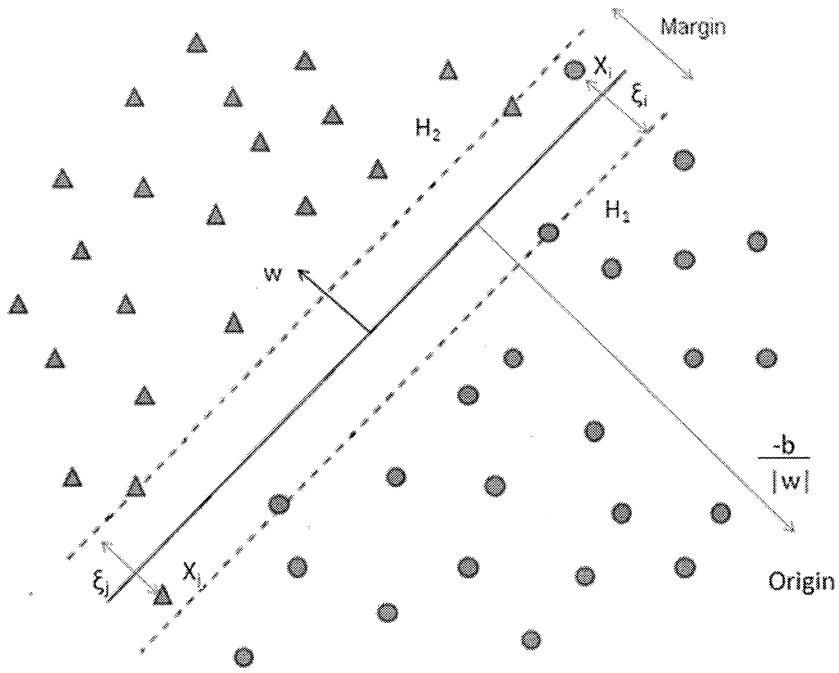

Figure 3.12: Effects of slack variables for non-linearly separable data.

The goal now is to allow the margin to be as large as possible, while keeping the number of data points with positive slack variables to a minimum. The constraints of the Lagrangian in (3.62) then become:

$$\begin{aligned} &0 \leq \alpha_i \leq C \\ &\sum_i \alpha_i y_i = 0 \end{aligned} \tag{3.65}$$

with the only difference from the optimal hyperplane being the addition of the upper bound of C on α_i, which essentially controls the width of the margin. The new primal Lagrangian becomes:

$$L_P \equiv \frac{1}{2}\| \mathbf{w} \|^2 + C\sum_i \xi_i - \sum_{i=1}^{l} \alpha_i \{y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 + \xi_i\} - \sum_{i=1}^{l} \mu_i \xi_i \qquad (3.66)$$

where μ are the new Lagrange multipliers introduced to enforce positivity of the ξ_i. The KKT conditions must once more be satisfied as follows:

$$\begin{aligned}
& v = 1,\ldots,d \qquad i = 1,\ldots,l \\
& \frac{\partial}{\partial w_v} L_P = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \\
& \frac{\partial}{\partial b} L_P = -\sum_i \alpha_i y_i = 0 \\
& \frac{\partial}{\partial \xi_i} L_P = C - \alpha_i - \mu_i \\
& y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 + \xi_i \geq 0 \\
& \xi_i \geq 0 \\
& \alpha_i \geq 0 \\
& \mu_i \geq 0 \\
& \alpha_i(y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 + \xi_i) = 0 \\
& \xi_i \mu_i = 0
\end{aligned} \qquad (3.67)$$

Satisfying these conditions allows for a linear SVM, capable of dealing with non-separable data, however, it is still not capable of dealing with nonlinear decision boundaries. To overcome this drawback, the data are mapped to a higher dimension, where they are much more likely to be linearly separable by a higher dimensional hyperplane. Since the data points of the Lagrangian only appear as dot products, as long as those inner products can be calculated in the higher dimensional space, it is not

necessary to perform any higher dimensional computation. Instead, computation is done through the kernel trick. By replacing the $\mathbf{x}_i \cdot \mathbf{x}_j$ in the Lagrangian with the kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \tag{3.68}$$

the calculations are still completed in approximately the same amount of time. The kernels chosen for the kernel trick must satisfy Mercer's Theorem:

$$K(\mathbf{x}, \mathbf{y}) = \sum_i \Phi(\mathbf{x})_i \Phi(\mathbf{y})_i \tag{3.69}$$

if and only if, for any *g(x)*:

$$\int g(\mathbf{x})^2 d\mathbf{x} < \infty \tag{3.70}$$

then:

$$\int K(\mathbf{x}, \mathbf{y}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0 \tag{3.71}$$

with the final decision boundary:

$$g(x) = \sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b = \sum_{\mathbf{x}_i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b = 0 \tag{3.72}$$

The new decision boundary is then a weighted sum of the kernel function with respect to the support vectors, and the original Lagrangian from (3.62) turns into:

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{3.73}$$

still subject to the constraints of (3.65).

The kernel used in this study is the Gaussian kernel:

$$K(\mathbf{x}, \mathbf{y}) = \exp(-\| \mathbf{x} - \mathbf{y} \|^2 / 2\sigma^2) \tag{3.74}$$

where σ is the width of the Gaussian [95,96].

3.7.3 NEURAL NETWORK TRAINING

The data input to the neural network are the detail coefficients from the discrete wavelet transform of the ERP signal. The Daubechies 4 wavelet is used throughout all tests. With a sampling frequency of 256 Hz, a total of 7 detail levels are possible in addition to 1 course approximation level. Table 3.2 outlines the numbers of coefficients in each of the levels. The number of coefficients in each level is enlarged slightly due to the filter of length 8 for the db4 wavelet used to perform the wavelet transform.

Table 3.2: Breakdown of the number of coefficients in the different levels of detail and approximation on the 256 Hz signal

	d1	d2	d3	d4	d5	d6	d7	a7
Frequency band	64-128Hz	32-64Hz	16-32Hz	8-16Hz	4-8Hz	2-4Hz	1-2Hz	0-1Hz
# of coefficients	132	69	38	22	14	10	8	8

In this analysis, only the final three levels of detail coefficients are used, d5, d6, and d7. This decision is based on the majority of the information from the P300 occurring in these frequency bands, as well as the previous work on this project. For convenience, a modified naming scheme for the different levels is used:

Level 1 – 1-2 Hz (d7)

Level 2 – 2-4 Hz (d6)

Level 3 – 4-8 Hz (d5)

The wavelet transform retaining temporal information coupled with the fact that the ERP signals used here are 1 second recordings from -200 ms to 800 ms allows for further reduction in the number of features. Since the distinguishing features of the ERP occur within 200-600 ms latency post-stimulus, the information from the first and last 200 ms can be effectively removed. This information removal reduces the coefficients

such that level 1 contains 4 coefficients, level 2 contains 4 coefficients, and level 3 contains 6 coefficients. A comparison between the classification performance of the different numbers of coefficients is presented later. These coefficients are used as the input to the neural network during training and testing. Each level of coefficients for each electrode for each type of stimulus is used to train a single classifier. As a result, there are a total of 114 (19 electrodes x 2 stimuli x 3 frequency bands) possible classifiers for Cohort A and 96 (16 electrodes x 2 stimuli x 3 frequency bands) possible classifiers for Cohort B.

3.7.4 K-FOLD CROSS-VALIDATION

Training is done using a cross-validation procedure. During the training process, the set of data is divided into K blocks, where each block contains $1/K$ of the data. The network is then trained on all but the Kth block of data and tested on block K. The process is repeated K times so that each block is used for testing once. The performances on each test block are then averaged to obtain an overall performance for the dataset [97,98]. This process prevents a particularly hard or easy to classify set of data from biasing the performance in a particular direction. The general process is shown in Figure 3.13.

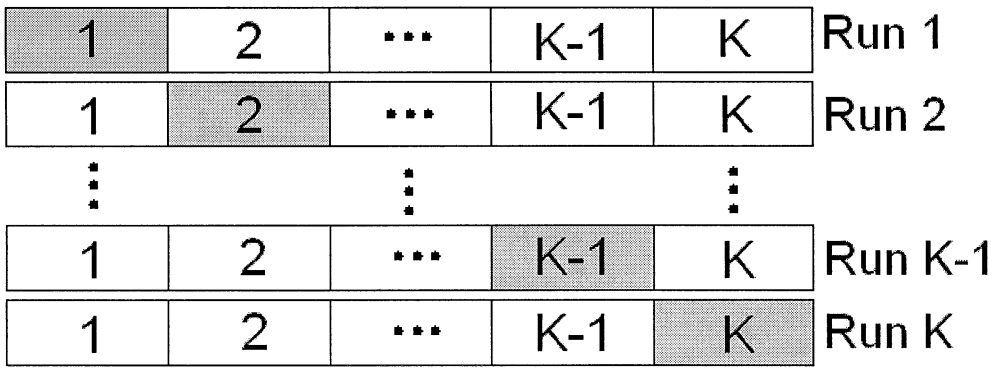

Figure 3.13: K-fold cross-validation. The highlighted block is the test data for each run with all other blocks used for training.

By setting K equal to the number of instances, or subjects, a leave-one-out cross-validation is obtained. For leave-one-out, each subject is removed one at a time while the rest of the cohort is used as training data. The removed patient is then used as testing data. Each subject is used for testing once during the process. As a result, the performance is either 0 or 1 for each run depending on whether the subject was classified correctly or not. These 0's and 1's can then be averaged to create a final performance of the classifier. This type of cross-validation is ideal for this dataset since the size of the dataset is not large. It also ensures that the classifier is trained on the maximum amount of data so that the classifier can make the best possible classification on the test subject. The drawback is that the computational time is significantly higher than using a smaller K.

3.8 DATA FUSION

Data fusion is the process of combining data or information from multiple sources to achieve a more reliable solution. In pattern recognition, there are many possible ways of performing fusion. Most commonly, the fusion occurs either at the feature level or the decision level. Based on empirical results from [82] as well as additional experiments, decision level fusion was determined to be the most effective approach in this study.

3.8.1 ENSEMBLE OF CLASSIFIERS

The decision level data fusion approach is performed through an ensemble of classifiers; also known as a committee of learners, mixture of experts, classifier ensemble, multiple classifier system, or consensus theory. The idea behind the ensemble of classifiers is

similar to consulting multiple doctors for additional opinions. By utilizing multiple classifiers, all trained on different training sets or different regions of the same feature set, and then combining their decisions together, a more robust and reliable classifier can be created. This better decision is achieved through each of the classifiers generating different decision boundaries and thereby producing different errors. When the decisions from the classifiers are combined together, the errors can be eliminated. To ensure the most benefit from an ensemble system, a collection of adequately diverse classifiers trained on complementary information is necessary. Within this study, the ensembles are created from classifiers trained on different electrode responses or different frequency bands of the same electrode response. One of the key components of this thesis then is to determine what electrode responses should be added to the ensemble to provide the most robust decision. A model of how an ensemble system works is shown in Figure 3.14. The classifier fusion is executed using various combination rules and is dependent upon the type of classifier being used.

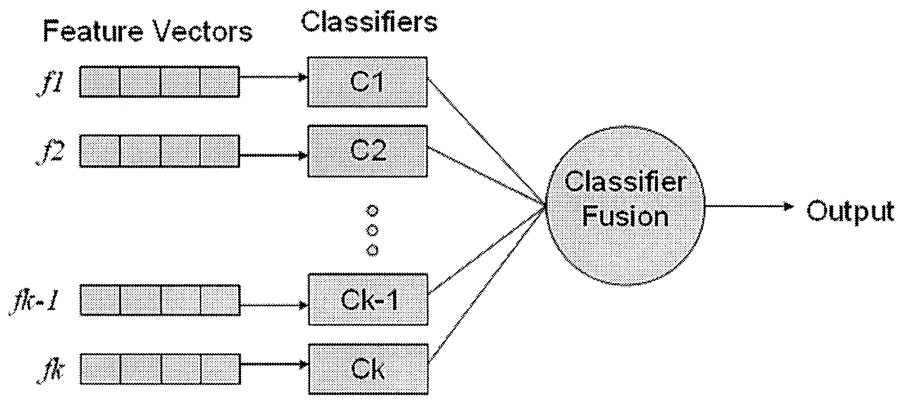

Figure 3.14. Model of an ensemble of classifiers system.

3.8.2 COMBINATION RULES

The combination rules used for the data fusion were sum rule, product rule, and weighted majority voting for MLP ensembles and simple majority voting for SVM ensembles. For the MLP ensembles, a continuous valued $d_{i,j} \in [0,1]$ is used to represent the degree of support given by the classifier C_i to class j, where $i=1,\ldots,N$; $j=1,\ldots,c$; N is the total number of classifiers, and c is the total number of classes.

Using these supports, the sum and product rules can be calculated. For these two rules, the support for each of the j classes from each of the c classifiers are either summed or multiplied together to yield a single support for each class. The class with the highest support is then selected as the chosen class for that particular ensemble. The rules can be shown mathematically as:

$$\begin{aligned} \mu_j(\mathbf{x}) &= \frac{1}{N}\sum_{j=1}^{N} d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in [0,1] \\ \mu_j(\mathbf{x}) &= \frac{1}{N}\prod_{i=1}^{N} d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in [0,1] \end{aligned} \tag{3.75}$$

For the weighted majority vote rule (WMV), the supports themselves are not needed, but rather only the accuracy of the selected class labels, $d_{i,j} \in \{0,1\}$. If the given classifier selects the correct class, a "1" is assigned and "0" otherwise. Next, $d_{i,j}$ is multiplied by a weighting factor based on the error of the training process:

$$W(T_i) = \log\left(1/\beta_{T_i}\right) \tag{3.76}$$

where β_{Ti} is the normalized training error for classifier T_i.

The class decision for each instance can then be found as class j that provides maximum support:

$$\mu_j(\mathbf{x}) = \sum_{j=1}^{N} W(T_i) d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in \{0,1\} \tag{3.77}$$

For SVMs, there are no separate class supports output by the classifier, therefore only the true and assigned class labels are used. By removing the weighting term from the WMV rule, a simple majority voting scheme can be used for SVMs. In this case $d_{i,j}$ follows the same form as for WMV, with selection of the appropriate class done as follows:

$$\mu_j(\mathbf{x}) = \sum_{j=1}^{N} d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in \{0,1\} \tag{3.78}$$

In this case, the ensemble is simply selecting the class that the majority of independent classifiers originally selected.

3.8.3 DEMPSTER-SHAFER COMBINATION RULE

Based on Dempster-Shafer (DS) theory, DS combination uses decision templates (DT) and the method shown below to combine classifiers and choose a class. A decision profile (DP) is a *TxC* matrix that displays the output of each classifier, where *T* is the number of classifiers and *C* is the number of classes. Using the training data, decision templates ($DT_1, \ldots, DT_C$) are created based on typical decision profiles for each class. Next, decision profiles are created for every new instance *x*. Each *DP(x)* is compared to the *C* decision templates using Dempster-Shafer combination in order to determine which class the instance belongs to. The advantage to this method is that classifiers that consistently give support to the wrong class will not decrease the overall performance of the ensemble.

Equation (3.79) is used to calculate the "proximity" between the ith row of DT_j and the output of the ith classifier for instance x, where $\| \cdot \|$ represents Euclidean distance between these two vectors:

$$\Phi_{j,i} = \frac{(1+\| DT_j^i - D_i(x) \|^2)^{-1}}{\sum_{k=1}^{c}(1+\| DT_k^i - D_i(x) \|^2)^{-1}} \tag{3.79}$$

Belief degrees are then calculated for every class j and classifier i as:

$$b_j(D_i(x)) = \frac{\Phi_{j,i}(x)\prod_{k\neq j}(1-\Phi_{k,i}(x))}{1-\Phi_{j,i}(x)[1-\prod_{k\neq j}(1-\Phi_{k,i}(x))]} \tag{3.80}$$

Finally, the class of x is selected based on the degrees of support for each class j, which are

$$\mu_j = K\prod_{i=1}^{T} b_j(D_i(x)), \qquad j=1,...,C \tag{3.81}$$

3.9 MEDICAL DIAGNOSTIC MEASURES

In medical studies, several measures of the effectiveness of the test are generally included along with classification or diagnostic performance numbers. These metrics are based on binary classification where the diagnosis is either positive or negative for a particular condition. The metrics used here are sensitivity, specificity and positive predictive values. Table 3.3 outlines how the metrics are defined.

Table 3.3: Chart of how test performance metrics are derived.

		Condition	
		Present	Not Present
Test Outcome	Positive	True Positive (A)	False Positive (B)
	Negative	False Negative (C)	True Negative (D)

Sensitivity refers to the probability of a positive diagnosis when the condition is present in the subject. In this study, sensitivity is defined as the ratio of AD subjects correctly identified by the algorithm. Specificity refers to the probability of a negative diagnosis when the condition is not present in the subjects. In this study, specificity is defined as the ratio of cognitively normal subjects correctly identified by the algorithm. Positive predictive value (PPV) refers to the probability that the subject has the disease when tested positive by the algorithm. In this study, PPV is defined as the proportion of subjects identified as AD by the algorithm who actually have AD. Negative predictive value (NPV) refers to the probability that the subject does not have the disease when tested negative by the algorithm. In this study, NPV is defined as the proportion of subjects identified as cognitively normal by the algorithm who actually are cognitively normal. These metrics can be calculated according to Table 3.3 as follows:

$$Sensitivity = \frac{A}{A+C} \tag{3.82}$$

$$Specificity = \frac{D}{D+B} \tag{3.83}$$

$$PPV = \frac{A}{A+B} \tag{3.84}$$

$$NPV = \frac{D}{C+D} \tag{3.85}$$

CHAPTER 4

RESULTS

The description of results from both studies follow a similar procedure. The ERPs are first decomposed using a discrete wavelet transform. The coefficients are used to train and test an MLP or SVM in a leave-one-out scenario. Decision level fusion is then used to combine classifiers, each trained on different electrode locations and stimulus types, to create an ensemble decision, which becomes the final classification or diagnosis from the algorithm. The presentation of results is as follows:

- Individual classifier results from Cohort A using MLPs
- Individual classifier results from Cohort B using MLPs
- Individual classifier results from Cohort A using SVMs
- Individual classifier results from Cohort B using SVMs
- Data fusion results from Cohort A using MLPs
- Data fusion results from Cohort A using SVMs
- Data fusion results from Cohort B using MLPs
- Data fusion results from Cohort B using SVMs
- 20% test results from Cohort A
- 20% test results from Cohort B
- Mixed cohort test results
- Cross cohort training/testing results

4.1 SINGLE CLASSIFIER RESULTS FROM COHORT A USING MLPS

Before applying data fusion, all 114 three-tuple (electrode location, frequency band, stimulus) combinations were analyzed individually to determine the best performing responses to use in an ensemble. Also shown is a comparison between using the full set of wavelet coefficients at each level and using only the middle 60% of coefficients focusing on the prominent features of the ERP such as the P200, N100, P300, etc. Only

the three lowest frequency bands were used as they have been shown to be the most informative [82]. Results are presented in Table 4.1 for ERPs obtained in response to target stimuli and Table 4.2 for ERPs obtained in response to novel stimuli as averages of 10 independent trials with a 95% confidence interval. Classifier performances over 60% are bolded and highlighted.

Table 4.1: Results for classifiers trained on target stimuli responses from all 19 electrodes at all 3 feature levels for the subjects in Cohort A using MLPs. Also shown is a comparison between using the full set of wavelet coefficients and only the middle coefficients.

Electrode Response	1-2 Hz, all	1-2 Hz, middle	2-4 Hz, all	2-4 Hz, middle	4-8 Hz, all	4-8 Hz, middle
C3	49.43 ± 2.19%	45.21 ± 2.83%	48.03 ± 1.82%	56.62 ± 2.51%	51.41 ± 2.86%	48.17 ± 1.91%
C4	51.54 ± 2.37%	55.21 ± 2.39%	54.23 ± 1.93%	51.69 ± 1.33%	49.30 ± 2.24%	52.25 ± 2.67%
CZ	46.47 ± 1.88%	50.42 ± 2.60%	52.25 ± 2.12%	54.08 ± 2.28%	50.28 ± 2.90%	47.32 ± 2.64%
F3	44.65 ± 2.43%	47.89 ± 1.92%	49.72 ± 2.24%	50.99 ± 2.57%	50.28 ± 3.35%	53.38 ± 1.56%
F4	58.59 ± 2.25%	58.03 ± 1.95%	47.04 ± 3.01%	44.65 ± 2.31%	52.54 ± 2.31%	48.73 ± 1.97%
F7	56.33 ± 3.05%	51.55 ± 1.97%	48.87 ± 2.00%	53.24 ± 3.04%	44.23 ± 1.55%	50.00 ± 1.81%
F8	56.90 ± 1.28%	**65.21 ± 1.88%**	49.44 ± 1.60%	52.54 ± 2.34%	51.83 ± 2.60%	52.34 ± 3.12%
FP1	57.60 ± 1.86%	57.18 ± 2.47%	58.45 ± 1.85%	57.32 ± 2.98%	44.23 ± 1.89%	42.67 ± 3.07%
FP2	**62.39 ± 1.96%**	**62.39 ± 2.21%**	50.99 ± 2.60%	48.59 ± 1.69%	49.72 ± 2.14%	44.51 ± 3.17%
FZ	54.51 ± 3.54%	58.87 ± 1.32%	53.52 ± 3.54%	52.25 ± 2.41%	48.59 ± 1.89%	48.45 ± 2.78%
O1	42.26 ± 2.00%	42.82 ± 1.89%	50.99 ± 2.39%	51.83 ± 2.60%	45.35 ± 1.99%	47.46 ± 2.88%
O2	57.89 ± 1.65%	57.47 ± 1.83%	47.18 ± 1.93%	47.47 ± 2.61%	47.18 ± 1.60%	49.16 ± 1.30%
P3	59.86 ± 2.31%	**63.24 ± 1.94%**	**60.28 ± 2.84%**	**60.56 ± 2.40%**	53.24 ± 2.17%	50.56 ± 2.50%
P4	**61.27 ± 1.97%**	**60.14 ± 1.84%**	**61.97 ± 1.49%**	**60.56 ± 2.24%**	40.56 ± 1.95%	40.99 ± 3.67%
P7	**62.68 ± 2.11%**	54.79 ± 2.32%	57.61 ± 0.60%	51.83 ± 3.14%	58.03 ± 2.95%	59.72 ± 2.11%
P8	56.48 ± 2.56%	54.37 ± 2.37%	52.25 ± 2.22%	55.63 ± 2.70%	45.35 ± 2.45%	35.92 ± 2.70%
PZ	**62.25 ± 1.71%**	52.53 ± 1.96%	57.75 ± 2.95%	58.45 ± 2.28%	45.78 ± 2.61%	44.65 ± 2.61%
T7	48.73 ± 1.459%	53.66 ± 2.32%	56.76 ± 2.40%	54.79 ± 2.44%	58.17 ± 1.80%	**62.82 ± 2.44%**
T8	57.04 ± 2.11%	59.43 ± 2.13%	51.13 ± 2.37%	51.83 ± 3.37%	42.68 ± 1.96%	42.26 ± 2.52%

Table 4.2: Results for classifiers trained on novel stimuli responses from all 19 electrodes at all 3 feature levels for the subjects in Cohort A using MLPs. Also shown is a comparison between using the full set of wavelet coefficients and only the middle coefficients.

Electrode Response	1-2 Hz, all	1-2 Hz, middle	2-4 Hz, all	2-4 Hz, middle	4-8 Hz, all	4-8 Hz, middle
C3	49.58 ± 2.33%	55.21 ± 2.36%	**62.95 ± 2.49%**	61.97 ± 2.07%	50.42 ± 1.87%	53.66 ± 2.70%
C4	51.13 ± 2.66%	54.79 ± 2.01%	**60.70 ± 2.47%**	57.60 ± 1.12%	45.07 ± 1.72%	48.87 ± 1.92%
CZ	55.77 ± 2.91%	55.35 ± 1.92%	59.01 ± 1.51%	**65.49 ± 2.22%**	50.28 ± 2.63%	51.83 ± 1.53%
F3	51.83 ± 3.02%	53.52 ± 3.35%	49.71 ± 3.33%	50.98 ± 2.23%	49.43 ± 2.56%	40.28 ± 1.77%
F4	57.47 ± 2.84%	56.48 ± 2.05%	53.80 ± 1.79%	52.53 ± 2.49%	50.00 ± 1.64%	48.02 ± 1.56%
F7	50.70 ± 3.72%	54.93 ± 1.84%	53.38 ± 2.86%	49.43 ± 2.05%	38.03 ± 1.63%	42.67 ± 3.44%
F8	**60.99 ± 2.61%**	**61.41 ± 1.45%**	48.87 ± 2.43%	54.64 ± 2.51%	44.78 ± 3.70%	43.10 ± 2.67%
FP1	**65.35 ± 2.08%**	**66.90 ± 2.75%**	58.02 ± 1.87%	**61.26 ± 2.01%**	49.29 ± 2.21%	53.24 ± 3.91%
FP2	**62.40 ± 1.63%**	57.61 ± 2.09%	55.49 ± 2.15%	59.01 ± 2.99%	49.29 ± 3.03%	42.81 ± 1.93%
FZ	45.49 ± 3.17%	50.42 ± 1.95%	48.73 ± 2.28%	57.60 ± 2.44%	53.52 ± 2.30%	43.52 ± 3.61%
O1	53.10 ± 1.22%	54.64 ± 1.66%	50.42 ± 1.57%	54.08 ± 2.83%	44.08 ± 2.19%	47.88 ± 2.63%
O2	**62.39 ± 2.07%**	**66.75 ± 2.47%**	56.90 ± 2.18%	59.71 ± 2.41%	43.09 ± 2.58%	44.22 ± 3.22%
P3	**61.97 ± 2.34%**	**60.70 ± 1.98%**	**60.98 ± 1.54%**	**67.74 ± 1.82%**	54.93 ± 3.14%	**64.93 ± 1.56%**
P4	**60.99 ± 2.14%**	**61.69 ± 2.23%**	**65.07 ± 1.43%**	**62.67 ± 1.35%**	52.11 ± 3.12%	53.94 ± 2.27%
P7	57.32 ± 2.24%	51.54 ± 1.81%	58.59 ± 2.70%	**60.98 ± 2.52%**	50.70 ± 2.77%	56.05 ± 1.48%
P8	**62.25 ± 2.23%**	**60.70 ± 1.94%**	**62.39 ± 2.66%**	**66.33 ± 1.74%**	59.71 ± 1.97%	51.54 ± 1.55%
PZ	59.72 ± 2.50%	**67.46 ± 2.26%**	**65.07 ± 2.06%**	**72.25 ± 1.49%**	**63.09 ± 3.04%**	**70.14 ± 2.30%**
T7	48.17 ± 4.13%	48.73 ± 1.77%	47.46 ± 2.63%	47.32 ± 3.17%	50.98 ± 2.03%	46.33 ± 2.19%
T8	**61.97 ± 1.22%**	**65.91 ± 1.62%**	53.66 ± 2.41%	55.06 ± 2.26%	38.45 ± 2.58%	52.25 ± 2.32%

Significant information is presented in these two tables. The comparison between using the full set of wavelet coefficients and only the middle coefficients shows that both sets perform similarly. For the target set, the middle coefficients performed better in 30 of the 57 three-tuples and for the novel set the middle coefficients performed better in 38 of the 57 three-tuples. Based on this analysis, the majority of the information within the wavelet coefficients that distinguishes between AD and normal is contained within the middle 60%. For the duration of this thesis, all training and testing is done using only those middle coefficients. By using only the middle coefficients, the focus is on the prominent features of the ERP. In addition, a reduced number of features are used to

train and test the MLP by nearly half, effectively reducing computation time, which is very appealing when considering this as a real world application.

Looking at only the middle coefficients, it can be seen that the novel responses perform slightly better than the target responses. The peak performance for the target responses was 65.21% using the F8 electrode response from 1-2 Hz. The peak performance for novel responses was 72.25% using the Pz electrode response from 2-4 Hz. In addition, the novel responses had many more electrode responses yielding performance figures in the 60-65% range. For each set, the performance figures for the entire scope of the head were generally in the 45-55% range with the higher numbers coming from electrode responses in the parietal region. It is clear that these numbers are still too low to provide a confident diagnosis between AD and normal, which leads to the implementation of ensemble systems.

4.2 SINGLE CLASSIFIER RESULTS FROM COHORT B USING MLPS

Similar to the analysis from Cohort A, the data from Cohort B were first analyzed on the individual three-tuple levels. Since the data were recorded by different technicians using different equipment and preprocessed using an automated system rather than by an EEG expert, it is necessary to compare the individual results. Like Cohort A, the data are presented in two tables, one for target and one for novel, as averages of 10 independent leave-one-out trials along with a 95% confidence interval. For Cohort B, only the middle coefficients were analyzed based on the results of Cohort A. Also important to note is a slight change in the electrode placements. The total number of electrodes has been reduced from 19 to 16 due to equipment limitations. The F7 and F8 locations have been

removed completely and the O1 and O2 locations have been replaced by a single Oz electrode. The results are in Table 4.3 for target responses and Table 4.4 for novel responses. Again, classifiers performing above 60% are highlighted and bolded.

Table 4.3: Results for classifiers trained on target stimuli responses from all 16 electrodes at all 3 feature levels for the subjects of Cohort B using MLPs.

Electrode Response	1-2 Hz	2-4 Hz	1-2 Hz
C3	**67.58 ± 2.39%**	50.00 ± 1.59%	58.06 ± 2.89%
C4	**65.81 ± 1.69%**	54.52 ± 3.11%	56.77 ± 2.05%
CZ	50.48 ± 2.89%	52.74 ± 1.39%	56.94 ± 2.25%
F3	52.42 ± 3.39%	50.97 ± 2.34%	**63.23 ± 2.40%**
F4	56.77 ± 2.28%	47.90 ± 2.16%	**67.58 ± 2.26%**
FZ	58.55 ± 2.20%	54.52 ± 2.10%	**68.23 ± 2.25%**
FP1	52.26 ± 1.26%	**62.10 ± 2.34%**	46.94 ± 2.26%
FP2	49.19 ± 1.72%	**65.81 ± 2.05%**	47.26 ± 2.16%
Oz	**62.58 ± 2.19%**	57.90 ± 1.84%	51.13 ± 2.75%
PZ	**68.39 ± 1.98%**	56.61 ± 2.18%	44.19 ± 2.54%
P3	57.10 ± 3.50%	55.16 ± 2.28%	49.68 ± 2.36%
P4	58.87 ± 2.79%	59.19 ± 2.53%	59.84 ± 4.50%
P7	56.13 ± 2.05%	47.26 ± 1.82%	54.35 ± 1.87%
P8	**62.58 ± 1.45%**	**62.10 ± 3.21%**	58.06 ± 2.85%
T7	51.94 ± 2.91%	50.81 ± 2.69%	50.48 ± 2.75%
T8	**61.77 ± 2.20%**	41.45 ± 2.02%	57.90 ± 3.13%

Table 4.4: Results for classifiers trained on novel stimuli responses from all 16 electrodes at all 3 feature levels for the subjects of Cohort B using MLPs.

Electrode Response	1-2 Hz	2-4 Hz	4-8 Hz
C3	44.84 ± 2.44%	53.55 ± 1.31%	**69.84 ± 2.02%**
C4	46.94 ± 2.35%	55.00 ± 1.55%	58.39 ± 2.71%
CZ	43.39 ± 1.73%	53.87 ± 2.03%	59.84 ± 2.18%
F3	48.06 ± 2.01%	55.97 ± 3.47%	**66.94 ± 3.06%**
F4	55.48 ± 2.25%	**68.06 ± 1.63%**	**62.58 ± 2.63%**
FZ	44.84 ± 2.10%	51.77 ± 2.35%	**62.58 ± 1.45%**
FP1	45.00 ± 2.08%	52.26 ± 1.54%	52.10 ± 2.33%
FP2	42.74 ± 3.02%	48.71 ± 2.84%	**60.00 ± 1.69%**
Oz	**66.61 ± 1.87%**	**72.74 ± 2.69%**	47.26 ± 2.57%
PZ	50.00 ± 3.05%	**60.16 ± 2.60%**	**60.00 ± 1.96%**
P3	**62.90 ± 2.37%**	**60.16 ± 2.16%**	**60.81 ± 1.65%**
P4	**68.71 ± 1.26%**	59.52 ± 1.67%	57.90 ± 2.06%
P7	**68.71 ± 2.25%**	57.90 ± 2.08%	54.52 ± 1.80%
P8	**67.42 ± 2.63%**	**66.45 ± 2.36%**	59.84 ± 2.26%
T7	52.90 ± 2.23%	47.90 ± 2.71%	59.35 ± 2.98%
T8	53.71 ± 1.92%	**66.61 ± 2.16%**	**60.48 ± 2.69%**

Similar to the results from Cohort A, the responses to novel tones yielded many more classifiers performing over the 60% level. The majority of these classifiers also occur from responses in the parietal region of the head, as expected based on the results from Cohort A. The two cohorts differ in that Cohort B provides some high performing classifiers in the frontal region, predominantly in the 4-8 Hz range. The top performing classifier for the target responses was Pz on 1-2 Hz at 68.39%. For the novel responses, however, the Pz electrode responses did not perform well at all. Oz was the top performing electrode response on 2-4 Hz at 72.74%. This change could be due to the relatively close location of the Oz electrode to the Pz electrode, although still surprising.

4.3 SINGLE CLASSIFIER RESULTS FROM COHORT A USING SVMS

To compare the performance of SVMs against MLPs, the single classifiers were looked at individually, this time using SVMs as the base classifier. Since SVM provides the optimal boundary each time, it is not necessary to average 10 trials together and therefore results are presented as single trials. Additional trials would yield identical results. Target responses are presented in Table 4.5 and novel responses in Table 4.6.

Table 4.5: Results for classifiers trained on target stimuli responses from all 19 electrodes at all 3 feature levels for the subjects of Cohort A using SVMs.

Electrode Response	1-2 Hz	2-4 Hz	4-8 Hz
C3	45.07%	54.93%	50.70%
C4	50.70%	47.89%	33.80%
CZ	46.48%	38.03%	45.07%
F3	50.70%	46.48%	53.52%
F4	50.70%	49.30%	50.70%
F7	56.34%	43.66%	46.48%
F8	59.15%	50.70%	53.52%
FZ	50.70%	52.11%	50.70%
FP1	**63.38%**	**64.79%**	42.25%
FP2	**69.01%**	53.52%	43.66%
O1	40.85%	53.52%	53.52%
O2	49.30%	46.48%	52.11%
PZ	49.30%	59.15%	42.25%
P3	57.75%	56.34%	56.34%
P4	54.93%	**66.20%**	53.52%
P7	59.15%	52.11%	59.15%
P8	**63.38%**	**66.20%**	45.07%
TZ	53.52%	52.11%	50.70%
T8	50.70%	47.89%	47.89%

Table 4.6: Results for classifiers trained on novel stimuli responses from all 19 electrodes at all 3 feature levels for the subjects of Cohort A using SVMs.

Electrode Response	1-2 Hz	2-4 Hz	4-8 Hz
C3	53.52%	**70.42%**	57.75%
C4	40.85%	**63.38%**	47.89%
CZ	54.93%	**69.01%**	50.70%
F3	54.93%	45.07%	53.52%
F4	46.48%	57.75%	46.48%
F7	56.34%	**60.56%**	45.07%
F8	**60.56%**	50.70%	45.07%
FZ	46.48%	**61.97%**	52.11%
FP1	**61.97%**	**60.56%**	**61.97%**
FP2	49.30%	49.30%	36.62%
O1	**60.56%**	47.89%	47.89%
O2	52.11%	59.15%	47.89%
PZ	**61.97%**	**70.42%**	**63.38%**
P3	**60.56%**	**63.38%**	**69.01%**
P4	56.34%	54.93%	52.11%
P7	50.70%	**61.97%**	54.93%
P8	**63.38%**	**64.79%**	52.11%
TZ	42.25%	49.30%	42.25%
T8	47.89%	45.07%	53.52%

Results in both tables are similar to the results from the MLPs, although generally the performance is slightly lower for the SVMs. FP2 yielded the highest performance for the targets at 69.02% from 1-2 Hz and Pz and C3 tied for the highest for novel responses at 70.42%, each from 2-4 Hz. The majority of high performing classifiers again originated from the parietal region. Based on these early results, the MLP seems to be the superior classifier in terms of performance, although the SVM was able to run at a fraction of the time. For the MLP to generate 10 independent leave-one-out trials for all electrode responses at each frequency band, computation time was roughly 65 hours, or approximately 6.5 hours per trial. The single trial from the SVM training completed in

just under 6 minutes with the added benefit of achieving identical results each time therefore eliminating the need for multiple trials.

4.4 SINGLE CLASSIFIER RESULTS FROM COHORT B USING SVMS

Cohort B was tested in the same fashion using SVMs. Results are presented in Table 4.7 for target responses and Table 4.8 for novel responses as single trial performances.

Table 4.7: Results for classifiers trained on target stimuli responses from all 16 electrodes at all 3 feature levels for the subjects of Cohort B using SVMs.

Electrode Response	1-2 Hz	2-4 Hz	4-8 Hz
C3	**70.97%**	43.55%	58.06%
C4	**67.74%**	46.77%	56.45%
CZ	54.84%	54.84%	54.84%
F3	56.45%	46.77%	54.84%
F4	51.61%	50.00%	53.23%
FZ	50.00%	53.23%	59.68%
FP1	**62.90%**	**62.90%**	53.23%
FP2	51.61%	**69.35%**	38.71%
OZ	**62.90%**	**64.52%**	51.61%
PZ	**67.74%**	**61.29%**	46.77%
P3	54.84%	51.61%	50.00%
P4	59.68%	58.06%	**66.13%**
P7	48.39%	56.45%	51.61%
P8	59.68%	**70.97%**	**64.52%**
TZ	56.45%	50.00%	43.55%
T8	59.68%	46.77%	51.61%

Table 4.8: Results for classifiers trained on novel stimuli responses from all 16 electrodes at all 3 feature levels for the subjects of Cohort B using SVMs.

Electrode Response	1-2 Hz	2-4 Hz	4-8 Hz
C3	45.16%	50.00%	**62.90%**
C4	48.39%	53.23%	56.45%
CZ	53.23%	51.61%	59.68%
F3	30.65%	51.61%	**70.97%**
F4	56.45%	**74.19%**	**64.52%**
FZ	50.00%	54.84%	53.23%
FP1	45.16%	59.68%	53.23%
FP2	41.94%	45.16%	54.84%
OZ	**74.19%**	**72.58%**	41.94%
PZ	53.23%	56.45%	58.06%
P3	**64.52%**	**61.29%**	59.68%
P4	**67.74%**	58.06%	54.84%
P7	**66.13%**	53.23%	**64.52%**
P8	**62.90%**	**70.97%**	58.06%
TZ	58.06%	40.32%	58.06%
T8	53.23%	**72.58%**	59.68%

For Cohort B, results from the MLP training versus SVM training were again similar, with some variance on an electrode by electrode basis. The maximum target response performance was 70.97% by C3 1-2 Hz and P8 2-4 Hz while the highest novel response was 74.19% by F4 2-4 Hz and Oz 1-2 Hz. Each of these maximums is higher for the SVMs than the MLPs. This trend is true for most of the electrode/frequency bands, but not all. For example, novel F3 1-2 Hz went from 48% using MLPs to 30% using SVMs. From the individual results for this cohort, it would seem that the ideal choice for a classifier would be the SVM since it provides better performance in a much shorter time.

Aside from the intricacies of the different classifiers, some other points can be observed from this comparison. In general, over the scope of all the different experiments, the parietal region is consistently the best performing region individually.

This characteristic is expected since the P300 is suspected to originate from this region of the head. The results also give some validity to this process, showing that the EEG was able to extract that information and through the use of neural networks distinguish between the Alzheimer's disease patients and the controls at a reasonable confidence. There were also significant performances from frontal electrodes, specifically for Cohort B, which could suggest the possibility of complementary information for use in diagnosis. To further investigate, data fusion was explored using these individual classifiers.

4.5 DATA FUSION RESULTS FROM COHORT A USING MLPS

In order to achieve a higher diagnostic performance, decision level data fusion was implemented. The decisions from the individual classifiers were combined through combination rules such a product rule, sum rule, simple and weighted majority voting, and Dempster-Shafer rule. Early attempts at data fusion on this project focused on the on Fz, Cz, and Pz electrode responses. Based on the individual classifier performances, it would make the most sense to try and combine the best performing electrodes, which are found predominantly in the parietal region. This region was evaluated exclusively in [82], although the combinations of responses were selected seemingly at random. The first step here was to look at all possible combinations of electrode responses in the region at all different frequency bands.

With a total of 5 electrodes in the region (Pz, P3, P4, P7, and P8), two stimuli (novel and target), and three frequency bands (1-2 Hz, 2-4 Hz, and 4-8 Hz), there are 30 different three-tuple combinations to be combined. The responses were first combined

into ensembles of 3, 5, and 7 classifiers separately for each stimulus type. Electrode responses were combined exhaustively resulting in 455 unique combinations of 3 electrode responses (15 choose 3), 3003 combinations of 5 classifiers, and 6435 combinations of 7 classifiers. Training was done using MLPs and results are an average of 10 independent leave-one-out trials. The diagnostic performances for the top 5 ensembles for combinations of 3, 5, and 7 classifier responses are shown in Table 4.9 along with the 95% confidence interval and combination rule that was used.

Table 4.9: Best performing ensembles for target and novel responses for ensembles of 3,5, and 7 classifiers trained using MLPs for Cohort A.

Target - Combinations of 3			
Electrodes/ Levels	**Mean (%)**	**CI (%)**	**Comb Rule**
$P3_1,P3_2,P7_3$	70.0	2.8	Sum
$P3_1,P4_1,P4_2$	69.6	3.0	Sum
$P3_1,P4_2,P7_3$	69.3	2.3	Sum
$Pz_3,P3_1,P4_2$	69.2	3.2	Sum
$P3_1,P3_2,P4_2$	69.2	2.4	Sum
Target - Combinations of 5			
$Pz_3,P3_1,P3_2,P4_2,P7_3$	70.1	3.2	WMV
$P3_1,P3_2,P4_1,P4_2,P7_3$	70.1	2.2	Sum
$Pz_2,Pz_3,P3_2,P3_3,P7_3$	69.9	2.9	Sum
$Pz_2,P3_1,P3_2,P4_1,P7_3$	69.7	2.6	Sum
$Pz_2,Pz_3,P3_1,P4_1,P4_2$	69.2	2.4	Sum
Target - Combinations of 7			
$Pz_2,P3_1,P3_2,$ $P4_2,P7_2,P7_3,P8_1$	70.0	2.1	Prod
$Pz_1,Pz_2,P3_1,$ $P3_2,P3_3,P7_3,P8_1$	69.4	3.6	Prod
$Pz_3,P3_1,P3_2,$ $P4_1,P4_2,P7_3,P8_2$	69.4	2.6	Sum
$Pz_2,Pz_3,P3_1,$ $P3_2,P4_2,P7_3,P8_1$	69.2	2.5	Prod
$Pz_3,P3_1,P3_2,$ $P4_1,P4_2,P7_3,P8_1$	69.0	2.6	Sum

Novel - Combinations of 3			
Electrodes/ Levels	**Mean (%)**	**CI (%)**	**Comb Rule**
$Pz_2,Pz_3,P8_2$	75.4	2.4	Sum
$Pz_1,Pz_3,P8_2$	74.8	2.2	Sum
$Pz_2,Pz_3,P3_2$	74.5	2.4	Sum
$Pz_2,Pz_3,P4_3$	74.2	2.2	Sum
$Pz_2,Pz_3,P3_3$	74.2	2.2	Sum
Novel – Combinations of 5			
$Pz_1,Pz_3,P3_2,P3_3,P8_1$	77.3	2.0	Sum
$Pz_2,Pz_3,P3_1,P7_2,P8_2$	77.3	1.5	Sum
$Pz_1,Pz_2,Pz_3,P3_3,P8_2$	77.0	1.1	Sum
$Pz_1,Pz_2,Pz_3,P3_2,P3_3$	76.9	1.7	Sum
$Pz_2,Pz_3,P3_3,P7_3,P8_2$	76.5	2.5	Sum
Novel - Combinations of 7			
$Pz_1,Pz_2,Pz_3,$ $P3_2,P3_3,P8_1,P8_2$	**79.0**	**2.2**	**Sum**
$Pz_1,Pz_2,Pz_3,$ $P3_3,P7_2,P8_1,P8_2$	78.7	2.0	Sum
$Pz_1,Pz_2,Pz_3,$ $P3_3,P7_1,P7_2,P8_1$	78.7	1.6	Sum
$Pz_1,Pz_2,Pz_3,$ $P3_1,P3_3,P7_2,P8_1$	78.6	1.5	Sum
$Pz_1,Pz_2,Pz_3,$ $P3_2,P3_3,P7_2,P8_1$	78.5	2.2	Sum

The results shown in this table support the feasibility of data fusion in increasing performance. As shown by the novel combinations of 7 classifiers, the best performance has increased from 72% to 79%, a statistically significant increase. This performance is also attainable by multiple different combinations of responses since the confidence

intervals of all of the top 5 combinations are overlapping. Also, the sum rule is clearly the best performing of all combination rules, providing 26 of the 30 best performance figures shown in the table. Not shown here are larger ensembles containing 9 or more responses, which start to show a drop off in classification performance. For this reason, future ensembles will be limited to 7 electrodes or less. Most importantly from this analysis, it is quite evident that there is complementary information shared between electrodes that allows for the increase of performance when combined together in an ensemble.

Knowing that there is complementary information shared between electrodes, it was necessary to determine whether there was also complementary information shared between the responses to different stimuli types. To accomplish this, the three-tuples from each of the stimuli types were combined together in ensembles. Since the number of three-tuple responses is now 30 (15 novel, 15 target), a full exhaustive analysis is not practical since there would be over 2 million unique ensembles of 7, $\binom{30}{7}$. To keep the results manageable, the 25 best performing ensembles of 7 classifiers from each stimulus type, target and novel, were examined and the 6 most frequently occurring three-tuples from each stimulus were selected for inclusion in the mixed stimuli ensembles. The selection method is shown in Figure 4.1 for clarification. N is equal to 6435, n is equal to 25, and M is equal to 57 in this case. An example of how the selection process works is shown in Figure 4.2.

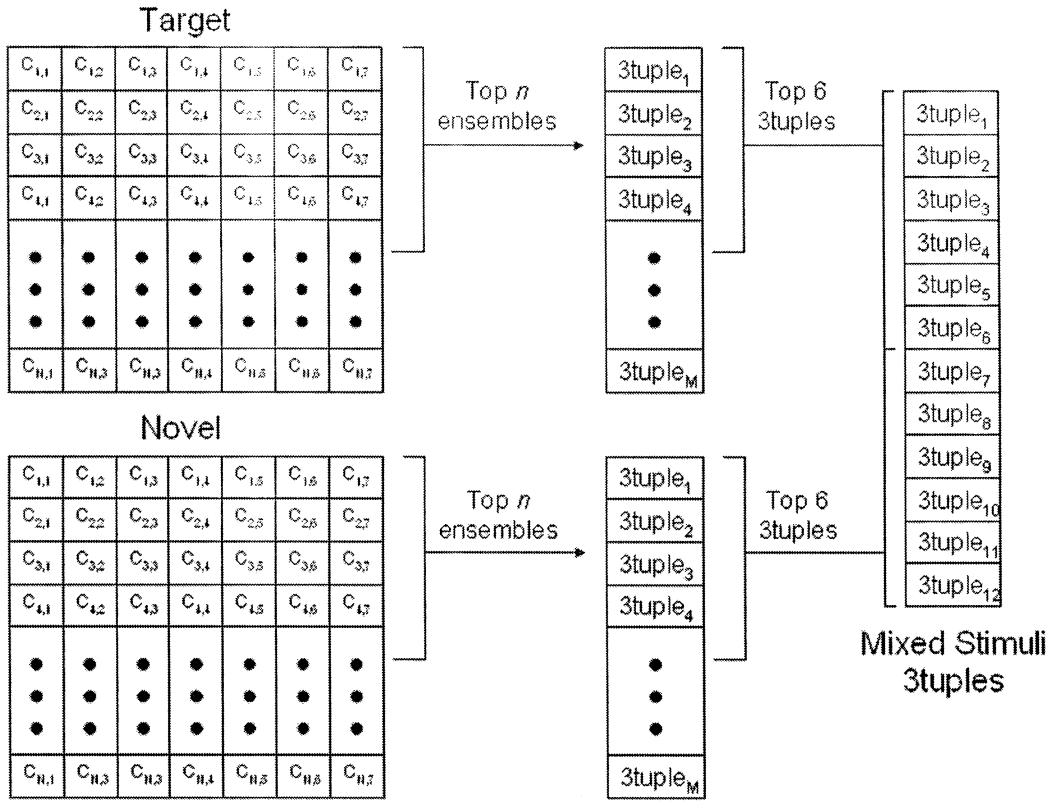

Figure 4.1: Method of selection of three-tuples from target and novel for use in mixed stimuli ensembles. Ensembles have been sorted by performance in decreasing order. The middle lists of three-tuples have been sorted by number of appearances in top n *ensembles in decreasing order.*

The 12 three-tuples were then combined exhaustively in combinations of 3, 5, and 7 once again. The resulting performance figures are shown in Table 4.10 as averages of 10 trials alone with a 95% confidence interval as well as sensitivity, specificity, and positive predicted value measures. This time the Dempster-Shafer rule was also included in the combinations. The electrode labels are appended with a 't' if the response came from the target stimulus.

Table 4.10: Best performing ensembles for mixed stimuli ensembles of 3,5, and 7 classifiers trained using MLPs for Cohort A.

Target/Novel - Combinations of 3						
Electrodes/Levels	**Mean (%)**	**CI (%)**	**Sens (%)**	**Spec (%)**	**PPV (%)**	**Comb Rule**
Pz_3, $P3_2$, $P3t_1$	78.2	2.2	77.1	78.4	70.8	Sum
Pz_3, $P3_2$, $P3_3$	78.0	3.1	82.1	75.6	68.1	Sum
Pz_2, Pz_3,$P3_3$	77.5	3.1	76.8	77.3	66.9	Sum
Pz_1, Pz_3, $P3_2$	77.3	2.0	77.9	71.9	66.6	Sum
Pz_3, $P3_3$, $P3t_1$	76.5	3.3	76.5	79.2	70.3	Sum
Target/Novel - Combinations of 5						
Pz_2, Pz_3,$P3_2$, $P3_3$,$P3t_1$	80.9	2.3	78.2	80.3	73.9	Sum
Pz_1, Pz_3,$P3_2$, $P3_3$,$P3t_2$	80.9	1.9	79.7	79.2	66.1	Sum
Pz_1, Pz_2,Pz_3, $P3_2$,$P7t_3$	80.8	1.2	80.3	81.4	71.8	DS
Pz_3,$P3_2$,$P3_3$,$P3t_1$,$P7t_3$	80.4	2.2	80.0	80.8	68.0	Sum
Pz_2, Pz_3,$P3_3$, $P3t_1$,$P7t_3$	80.1	2.6	76.7	79.5	68.7	Sum
Target/Novel - Combinations of 7						
Pz_1, Pz_3,$P3_2$,$P3_3$, $P3t_1$, $P3t_2$, $P7t_3$	**81.7**	**3.8**	**79.7**	**84.1**	**71.3**	**Sum**
Pz_3,P3-2,$P3_3$,$P3t_1$, Pzt_3, $P3t_2$, $P7t_3$	81.5	3.4	76.8	82.7	74.6	Sum
Pz_3,P3-2,$P3_3$,$P3t_1$, $P3t_2$, $P4t_2$, $P7t_3$	81.4	6.1	78.8	80.3	72.4	DS
Pz_3,$P3_2$,P8-2,$P3_3$, $P3t_1$,$P4t_2$, $P7t_3$	81.2	2.6	80.6	81.9	74.1	DS
Pz_3,$P3_2$,$P3_3$,Pz_2t, $P3t_1$, $P3t_2$,$P7t_3$	81.1	3.6	81.5	80.8	71.0	DS

From this analysis, it is obvious that there is additional complementary information provided between the two different stimuli types. In each case the performance from single stimuli ensemble to mixed stimuli ensembles is increased, 75% to 78% for ensembles of 3 classifiers, 77% to 80.9% for ensembles of 5 classifiers, and 79% to 81.7% for ensembles of 7 classifiers. An increasing performance trend is shown as additional electrodes are added to the ensemble from 3 to 7, also observed in the single stimulus ensembles. There is also both target and novel responses included in all top performing ensembles, with the exception of a few ensembles of 3. Also notable is the fact that the sum rule still provided the most top performances despite the increased computational cost of the Dempster-Shafer rule.

Only the parietal region electrodes have been fully examined so far. To fully analyze all data available, the remaining 14 electrodes should also be included in the analysis. Including all electrodes will help to show whether other regions of the head

contribute any further complementary information helpful in diagnosing AD, outside of the parietal region where the P300 is believed to originate. A full exhaustive examination is still not possible due to the fact that between the two different stimuli, there is a total of 114 different three-tuples, which would yield over 240,000 ensembles of 3 and over 40 billion ensembles of 7 classifiers.

To make the magnitude of the data more manageable, only a portion of the 114 three-tuples were used. By first looking at the responses to novel and target stimili individually, a more manageable 29,620 unique ensembles of 3 exist for each stimulus. To select exactly which three-tuples were to be used, the top 25 performing combinations for each response were examined. Again following Figure 4.1 (with 3-classifier ensembles, N=29,260), only the top 6 three-tuples most frequently occurring in those 25 top combinations were selected for the full analysis. The selected three-tuples were as follows: novel, Pz 1-2 Hz, Pz 2-4 Hz, Pz 4-8 Hz, O2 1-2 Hz, Cz 2-4 Hz, T8 1-2 Hz, and target P3 1-2 Hz, P3 2-4 Hz, FP2 1-2 Hz, F8 1-2 Hz, T7 2-4 Hz, T7 4-8 Hz. The most frequently appearing electrode response was Pz, with the remaining electrodes scattered around the head. Already this shows the possibility of complementary information outside of the parietal region. This selection is shown in Figure 4.2 as an example of the selection method depicted in Figure 4.1.

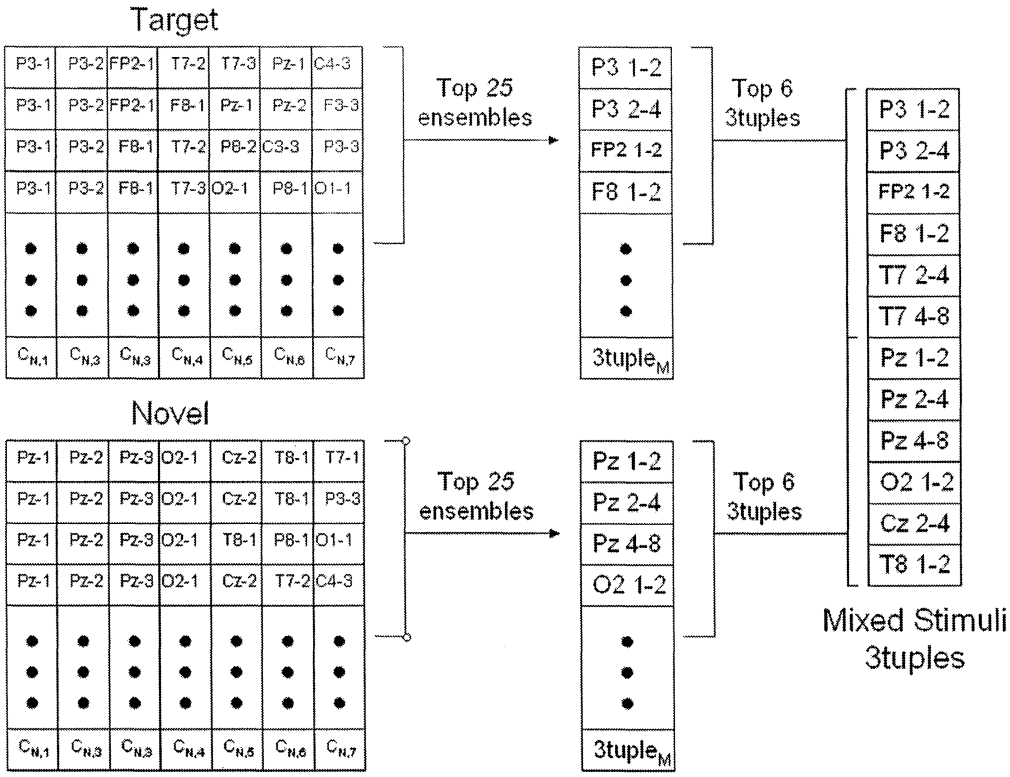

Figure 4.2: Example of how the selection method from Figure 4.1 works.

Before making a final analysis on this set of electrode responses, the necessity of using leave-one-out training was examined. Up to this point all results were averages of 10 leave-one-out trials. Leave-one-out is essentially just a 71 fold cross validation, and ultimately very computationally expensive. The possibility that this extent of validation being unnecessarily expensive was tested by conducting 3-fold, 5-fold, 10-fold, and leave-one-one training/testing for ensembles of 7 classifiers. The number of trials for this experiment was increased from 10 to 100 to shrink the confidence intervals to as small as possible. The resulting ensemble classification performances are shown in the following 4 tables (4.11 through 4.14) with the top five 7-classifier ensembles for each training method along with a 95% confidence interval and combination rule used.

Table 4.11: Top performing ensembles of 7 from mixed stimuli using 3 fold cross validation trained with MLPs for Cohort A.

3-fold cross validation			
Electrodes/Levels	**Mean (%)**	**CI (%)**	**Comb Rule**
Pz_2, Pz_3, $O2_1$, $T8_1$, $P3t_2$, $FP2t_1$, $T7t_3$	75.7	0.95	Sum
Pz_1, Pz_2, Pz_3, $O2_1$, $T8_1$, $P3t_2$, $FP2t_1$	75.6	0.84	Sum
Pz_1, Pz_2, Pz_3, $O2_1$, Cz_2, $P3t_2$, $FP2t_1$	75.5	0.84	Sum
Pz_2, Pz_3, $O2_1$, $P3t_1$, $P3t_2$, $FP2t_1$, $T7t_3$	75.4	0.84	Sum
Pz_1, Pz_2, Pz_3, $O2_1$, $P3t_2$, $FP2t_1$, $T7t_3$	75.4	0.84	Sum

Table 4.12: Top performing ensembles of 7 from mixed stimuli using 5 fold cross validation trained with MLPs for Cohort A.

5-fold cross validation			
Electrodes/Levels	**Mean (%)**	**CI (%)**	**Comb Rule**
Pz_2, Pz_3, $O2_1$, Cz_2, $T8_1$, $P3t_1$, $T7t_3$	78.5	0.69	Sum
Pz_2, Pz_3, $O2_1$, Cz_2, $P3t_1$, $F8t_1$, $T7t_3$	78.4	0.79	Sum
Pz_2, Pz_3, $O2_1$, $T8_1$, $P3t_1$, $FP2t_1$, $T7t_3$	78.3	0.70	Sum
Pz_2, Pz_3, $O2_1$, $P3t_1$, $P3t_2$, $F8t_1$, $FP2t_1$	78.2	0.66	Prod
Pz_2, Pz_3, $O2_1$, $T8_1$, $P3t_1$, $FP2t_1$, $T7t_2$	78.2	0.74	Sum

Table 4.13: Top performing ensembles of 7 from mixed stimuli using 10 fold cross validation trained with MLPs for Cohort A.

10-fold cross validation			
Electrodes/Levels	**Mean (%)**	**CI (%)**	**Comb Rule**
Pz_2, Pz_3, $O2_1$, $T8_1$, $P3t_2$, $FP2t_1$, $T7t_3$	80.8	0.71	Sum
Pz_2, Pz_3, $O2_1$, $P3t_1$, $P3t_2$, $FP2t_1$, $T7t_3$	80.6	0.69	Sum
Pz_2, Pz_3, $O2_1$, Cz_2, $T8_1$, $P3t_2$, $T7t_3$	80.6	0.76	Sum
Pz_2, Pz_3, $O2_1$, Cz_2, $P3t_1$, $F8t_1$, $T7t_3$	80.5	0.67	Sum
Pz_2, Pz_3, $O2_1$, $P3t_1$, P3t2, $FP2t_1$, $T7t_2$	80.5	0.70	Sum

Table 4.14: Top performing ensembles of 7 from mixed stimuli using leave-one-out or 71 fold cross validation trained with MLPs for Cohort A.

Leave one out			
Electrodes/Levels	**Mean (%)**	**CI (%)**	**Comb Rule**
Pz_2, Pz_3, $O2_1$, Cz_2, $T8_1$, $P3t_1$, $T7t_3$	82.7	0.60	Sum
Pz_2, Pz_3, $O2_1$, $P3t_1$, $P3t_2$, $FP2t_1$, $T7t_2$	82.7	0.59	Sum
Pz_1, Pz_2, Pz_3, $O2_1$, Cz_2, $P3t_2$, $FP2t_1$	82.6	0.55	Sum
Pz_2, Pz_3, $O2_1$, $T8_1$, $FP2t_1$, $FP2t_1$, $T7t_2$	82.5	0.54	Sum
Pz_2, Pz_3, $O2_1$, Cz_2, $P3t_1$, $F8t_1$, $T7t_3$	82.3	0.56	Sum

These results show an increasing trend as the number of folds is increased with leave-one-out providing the best classification performance. This trend is not surprising since the leave-one-out method allows for the most possible training data to train the MLPs while only one patient is tested at a time. The performance increase from 10-fold to 71-fold is great enough to support continuing analysis using the leave-one out scheme.

The addition of the remaining electrodes' responses outside the parietal region did provide slightly higher classification performance than the parietal region alone, from 81.7% to 82.7%. Regardless of the size of the increase in performance, this does verify that there are markers in the EEG outside of the top performing parietal region, which can contribute to the diagnosis of AD. An argument can be made, however, that the sample size of 25 ensembles of 3 was not robust enough to select the very best electrode responses.

To investigate this argument, the 29,620 ensembles of 3 classifers for each stimulus were again revisited. This time a threshold was drawn for all ensembles performing better than 75% for the novel combinations and 70% for the target combinations. These thresholds were chosen to select at least 1,000 unique ensembles for each stimulus. This adjustment increased the sample from 25 ensembles for each

stimulus to approximately 1,000 ensembles each for novel combinations and target combinations. The six most frequently appearing electrode responses were once again selected according to Figure 4.1 with n equal to roughly 1,000 and N equal to 29,260. To better visualize how often the different responses appear, a histogram was superimposed over the electrode distribution map. Three bars were placed at each electrode location, one for each frequency band with 1-2 Hz on top and 4-8 Hz on the bottom. The intensity of the color then shows how frequently the electrode/frequency band appeared in the total number of ensembles of three exceeding the given threshold. Figure 4.3 shows the histogram for target responses and Figure 4.4 shows the histogram for novel responses.

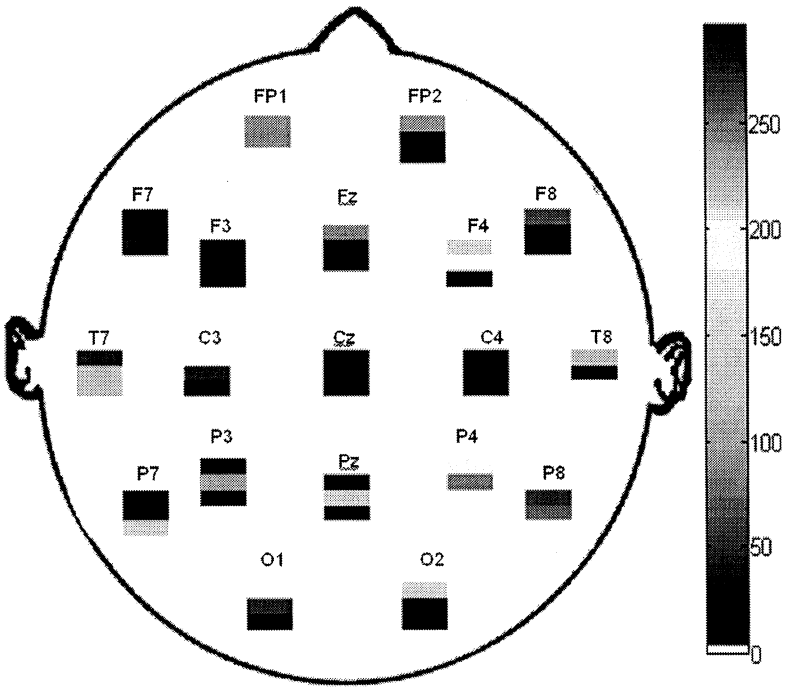

Figure 4.3: Distribution of most frequently occurring target electrode responses in combinations of three performing better than 70% using MLPs for Cohort A.

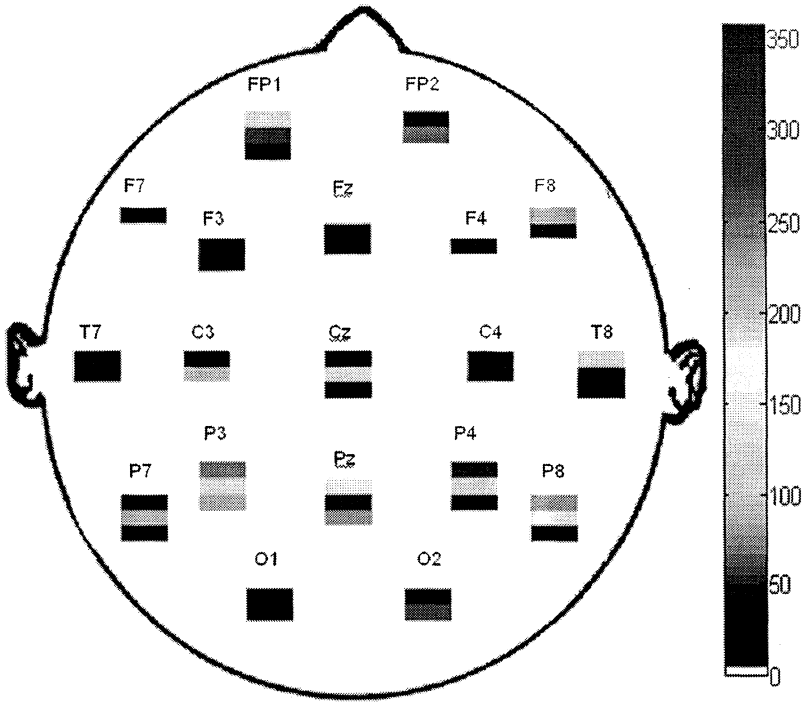

Figure 4.4: Distribution of most frequently occurring novel electrode responses in combinations of three performing better than 75% using MLPs for Cohort A.

The scalp histograms show that the original electrode site selection was not far off from the optimal set of electrode/frequency band combinations. Again the top 6 most frequently appearing three-tuples were selected for each stimulus. The majority of the three-tuples remained the same, although there were some changes as shown in Table 4.15.

Table 4.15: New selection of three-tuples based on larger sample size as compared to original selection of three-tuples using MLPs for Cohort A.

Novel		Target	
Old	**New**	**Old**	**New**
Pz 1-2 Hz	Pz 1-2 Hz	P3 1-2 Hz	P3 1-2 Hz
Pz 2-4 Hz	Pz 2-4 Hz	P3 2-4 Hz	P4 1-2 Hz
Pz, 4-8 Hz	Pz 4-8 Hz	FP2 1-2 Hz	FP2 1-2 Hz
O2 1-2 Hz	O2 1-2 Hz	F8 1-2 Hz	F8 1-2 Hz
Cz 2-4 Hz	P8 2-4 Hz	T7 2-4 Hz	P4 2-4 Hz
T8 1-2 Hz	P3 4-8 Hz	T7 4-8 Hz	T7 4-8 Hz

This slightly modified collection of three-tuples was then used in the same leave-out-out training scheme as before to see if any significant change was made in performance. The top 5 results of 100 trials again are shown in Table 4.16 with confidence intervals, sensitivity, specificity, and positive predicted values.

Table 4.16: Top 5 performing ensembles using the new selection of three-tuples using MLPs for Cohort A.

Electrode/Frequency/Stimulus	Average (%)	CI (%)	Sensitivity (%)	Specificity (%)	PPV (%)	Comb Rule
Pz_2, Pz_3, $O2_1$, $P3_3$, $F8t_1$, $FP2t_1$, $T7t_2$	82.2	0.58	79.2	85.1	83.1	Sum
Pz_2, Pz_3, $O2_1$, $P3_3$, $P3t_1$, $F8t_1$, $T7t_2$	82.0	0.50	79.0	84.8	82.8	Sum
Pz_2, Pz_3, $O2_1$, $P3_3$, $P3t_1$, FP2t1, $T7t_2$	81.3	0.64	79.9	82.6	80.1	Sum
Pz3, $O2_1$, $P3_3$, $P3t_1$, $F8t_1$, $FP2t_1$, $T7t_2$	81.1	0.53	78.9	83.1	81.2	Sum
Pz1, Pz2, Pz_3, $O2_1$, P33, $F8t_1$, $T7t_2$	81.0	0.46	78.2	83.6	81.5	Sum

The new selection of electrodes yielded a best performing ensemble of 82.2%. Overall, the new selection of electrodes performed statistically the same as the old selection. The new set of electrodes also yielded a peak single trial performance of 91.3% classification percentage. This performance shows there is definitely complementary information coming from different parts of the scalp as well as in response to different types of stimuli. This increase also shows there is not necessarily a "best" set of three-tuples to use, but rather particular regions where the most informative signals come from.

4.6 DATA FUSION RESULTS FROM COHORT A USING SVMS

A similar process was conducted using SVMs as the training method. An exhaustive search through all 29,620 ensembles of 3 was performed first. Overall, these performance numbers were slightly lower than their MLP counterparts resulting in a

cutoff threshold of 68% for novel and 64% for target to attain approximately 1,000 ensembles for each stimulus. Again scalp histograms were generated to better visualize the distribution of responses. These histograms can be seen in Figure 4.5 for target and Figure 4.6 for novel stimuli.

While not identical to the histograms from the MLPs, they are indeed very similar with common electrode responses occurring frequently regardless of the neural network used. The top 6 most frequently appearing electrodes were selected as: novel Pz 2-4 Hz, C3 2-4 Hz, P3 4-8 Hz, Cz 2-4 Hz, P8 1-2 Hz, P8 2-4 Hz, and target FP2 1-2 Hz, P4 2-4 Hz, P8 1-2 Hz, P8 2-4 Hz, FP1 1-2 Hz, FP1 2-4 Hz.

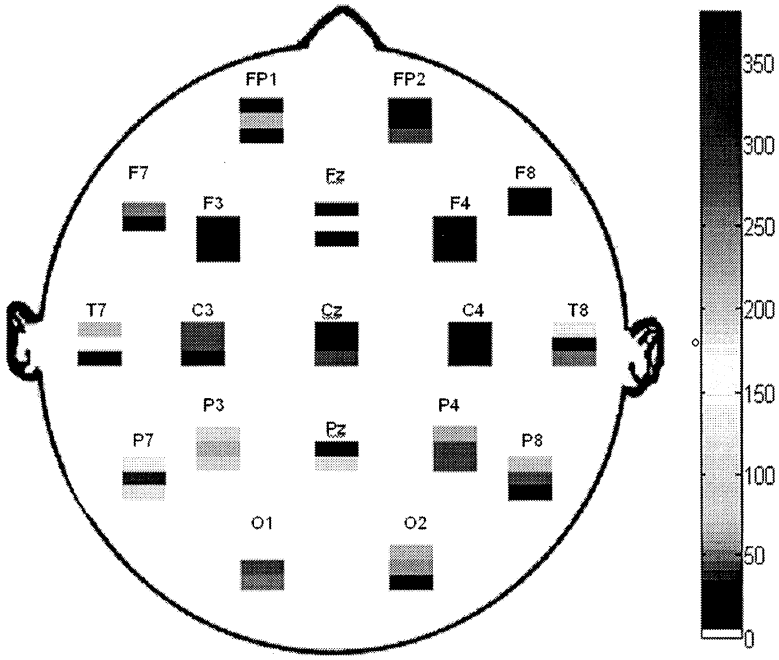

Figure 4.5: Distribution of most frequently occurring target electrode responses in combinations of three performing better than 64% using SVMs for Cohort A.

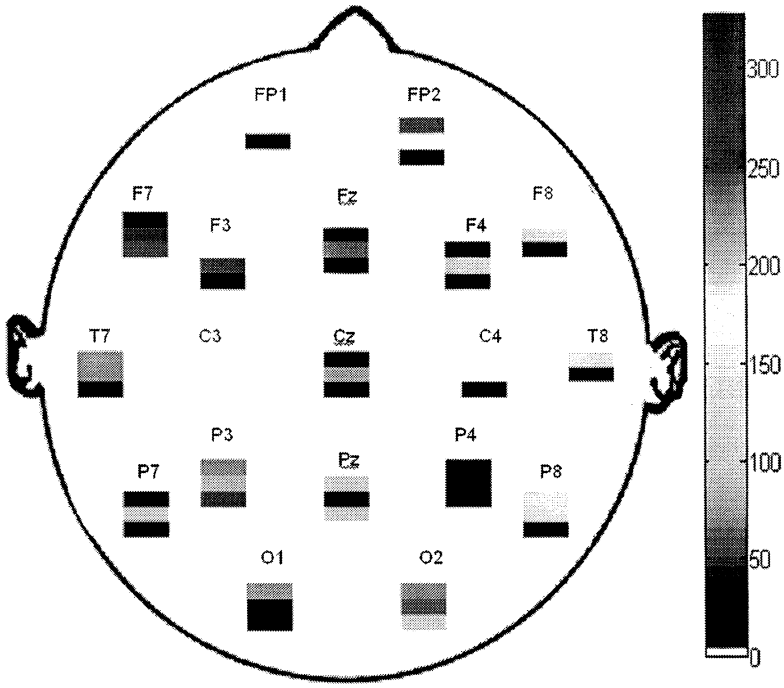

Figure 4.6: Distribution of most frequently occurring novel electrode responses in combinations of three performing better than 68% using SVMs for Cohort A.

These top 6 electrode responses from each stimulus type were then combined together into ensembles of 7 classifiers as previously done with the MLPs. The major difference here is the way the decisions are combined. Since there is no class support given by the SVM like there is for the MLP, only the selected class can be used to make a decision. To combine the decisions, simple majority voting is used as the combination rule for all SVM ensembles. The top 5 performing SVM ensembles are listed in Table 4.17.

Table 4.17: Top 5 performing ensembles using the top 6 three-tuples each from target and novel responses using SVMs for Cohort A.

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Pz_2, $C3_2$, $P3_3$, Cz_2, $P8_1$, $P8_2$, $FP1t_2$	77.46	70.59	83.78	80.00
Pz_2, $C3_2$, $P3_3$, Cz_2, $P8_1$, $P8t_2$, $FP1t_2$	77.46	70.59	83.78	80.00
Pz_2, $C3_2$, $P3_3$, Cz_2, $P8_2$, $P8t_2$, $FP1t_2$	77.46	73.53	81.08	78.13
Pz_2, $C3_2$, $P3_3$, Cz_2, $P8_2$, $FP1t_2$, $P8t_1$	77.46	70.59	83.78	80.00
Pz_2, $C3_2$, $P3_3$, Cz_2, $P8t_2$, $FP1t_2$, $P8t_1$	77.46	70.59	83.78	80.00

As seen earlier, the MLPs slightly outperform the SVMs when using the same process. The top 5 ensembles all performed at 77.46%, a significant reduction from the 82.7% attained by the MLPs.

A slightly more automated technique of electrode selection was also tested using the SVMs. The previous method involved an exhaustive search through all ensembles of 3 classifiers for each stimulus, from which the top 1,000 ensembles were examined to select.6 electrodes for each stimulus. For this new automated method, no ensembles are created before electrode selection. Rather, the top 15 performing individual classifiers were selected, which were then used to create ensembles of 7 classifiers. This process was done for novel and target responses separately first, and then with both stimuli combined and is shown in Figure 4.7. For Cohort A, M is equal to 57 and for Cohort B, M is equal to 48.

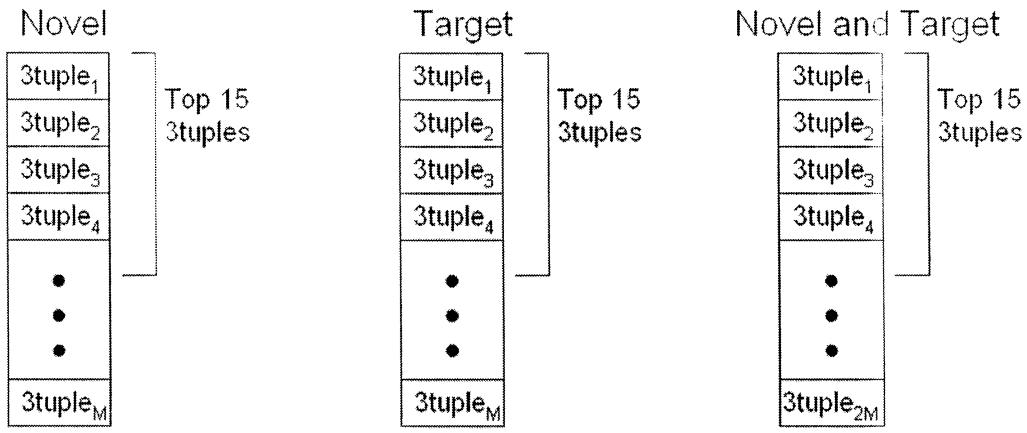

Figure 4.7: Alternate selection method for choosing three-tuples. Each list is ordered by single classifier performance for each three-tuple in descending order.

The results are shown in Table 4.18 for target, Table 4.19 for novel, and Table 4.20 for the mixed stimuli.

Table 4.18: Top 5 performing ensembles using the top 15 three-tuples from target responses using SVMs for Cohort A.

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$FP2_1$, $P4_2$, $P8_3$, $P8_1$, $FP1_1$, $P7_3$, $F8_1$	84.51	79.41	89.19	87.10
$FP2_1$, $P4_2$, $P8_3$, $P8_1$, $FP1_1$, $P7_1$, $F8_1$	83.10	79.41	86.49	84.38
$FP2_1$, $P4_2$, $P8_3$, $P8_1$, $P7_1$, $P7_3$, $F8_1$	83.10	79.41	86.49	84.38
$FP2_1$, $P4_2$, $P8_1$, $P7_3$, $F8_1$, $P3_1$, $P3_3$	83.10	79.41	86.49	84.38
$FP2_1$, $P4_2$, $P8_3$, $P8_1$, $FP1_1$, $FP1_2$, $F8_1$	81.69	79.41	83.78	81.82

Table 4.19: Top 5 performing ensembles using the top 15 three-tuples from novel responses using SVMs for Cohort A.

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Pz_2, $C3_2$, Cz_2, Pz_3, $P8_1$, Pz_1, $FP1_1$	84.51	82.35	86.49	84.85
Pz_2, $C3_2$, $P3_3$, Cz_2, Pz_3, Pz_1, $FP1_1$	83.10	79.41	86.49	84.38
Pz_2, $C3_2$, $P3_3$, Cz_2, $P8_1$, Pz_1, $FP1_1$	83.10	82.35	83.78	82.35
Pz_2, $C3_2$, $P3_3$, Pz_3, $P8_1$, Pz_1, $FP1_1$	83.10	85.29	81.08	80.56
Pz_2, $C3_2$, $P3_3$, $P7_2$, Pz_1, $FP1_1$, Fz_2	83.10	79.41	86.49	84.38

Table 4.20: Top 5 performing ensembles using the top 15 three-tuples from target and novel responses combined using SVMs for Cohort A.

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Pz_2, $C3_2$, $P3_3$, $FP2t_1$, $P8t_2$, $FP1t_2$, $P8_1$	88.73	88.24	89.19	88.24
$C3_2$, $P3_3$, $FP2t_1$, $C4_2$, $P8t_2$, $P8_1$, $P4t_2$	88.73	88.24	89.19	88.24
Pz_2, $C3_2$, $P3_3$, $P8t_2$, $P8_2$, $FP1t_2$, $P8_1$	87.32	88.24	86.49	85.71
Pz_2, $C3_2$, $FP2t_1$, $P4t_2$, $P8t_2$, $P8_1$, $FP1t_1$	87.32	88.24	86.49	85.71
Pz_2, $P3_3$, $FP2t_1$, $P8t_2$, $FP1t_2$, $P8_1$, $P8t_1$	87.32	82.35	91.89	90.32

As seen from these three tables, there is a significant increase over both the previous method of selecting responses and the best performing MLP ensembles. Both the target and novel ensembles individually performed better than the MLP mixed ensembles at 84.51%. When adding the two stimuli responses together, an additional increase in performance is realized up to 88.73%. In continuation with the MLP ensembles, all mixed ensembles contain responses from both target and novel stimuli, which is what attributes to the performance increase. Inclusion of responses to both stimuli is further evidence that there is complementary information shared between the responses of the two different stimuli. These numbers also show that the SVMs can outperform the MLPs in only a fraction of the time. The performances shown here also exceed the 75% accuracy of community clinics and are not much lower than the 90% accuracy of an expert neurologist.

4.7 DATA FUSION RESULTS FROM COHORT B USING MLPS

The ensembles for Cohort B using MLPs were started with the exhaustive search through all ensembles of 3 using the selection method from Figure 4.1 with N=17,296, M=48, n=1,000. Due to the reduction from 19 to 16 electrode locations, the total number of

ensembles was reduced to 17,296. Thresholds were set to 70% and 67% to obtain approximately 1,000 samples from novel and target responses, respectively. The scalp histograms are shown in Figure 4.8 for target responses and Figure 4.9 for novel responses. The top 6 electrodes for each stimulus type were: novel Oz 2-4 Hz, P7 1-2 Hz, C3 4-8 Hz, T8 2-4 Hz, F4 2-4 Hz, Oz 1-2 Hz and target Pz 1-2 Hz, Fz 4-8 Hz, F4 4-8 Hz, C3 1-2 Hz, FP2 2-4 Hz, P8 1-2 Hz. This list is different when compared with both lists from Cohort A. Oz is the predominant electrode in response to novel tones as opposed to Pz, although Pz was still the most frequently appearing electrode in the target list.

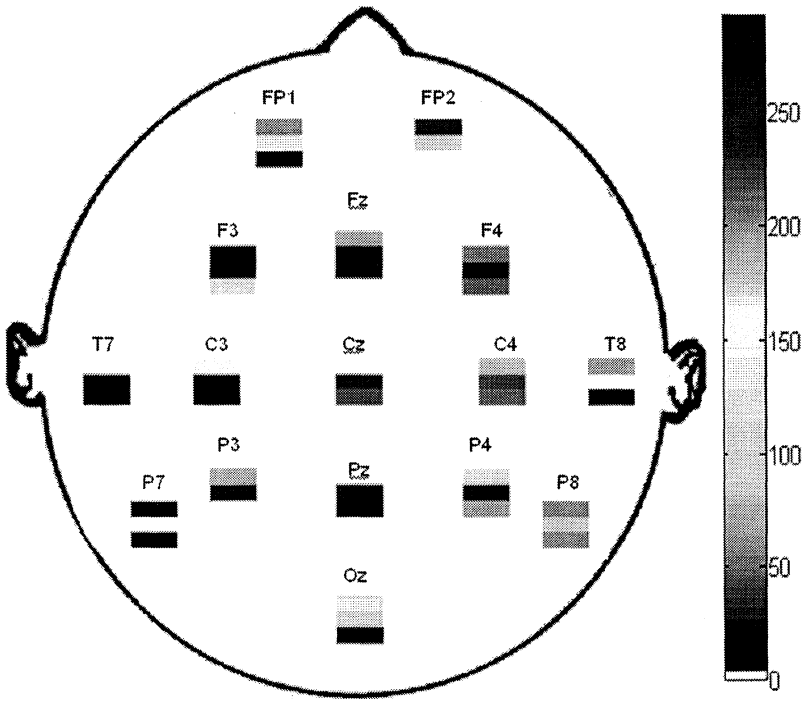

Figure 4.8: Distribution of most frequently occurring target electrode responses in combinations of three performing better than 67% using MLPs for Cohort B.

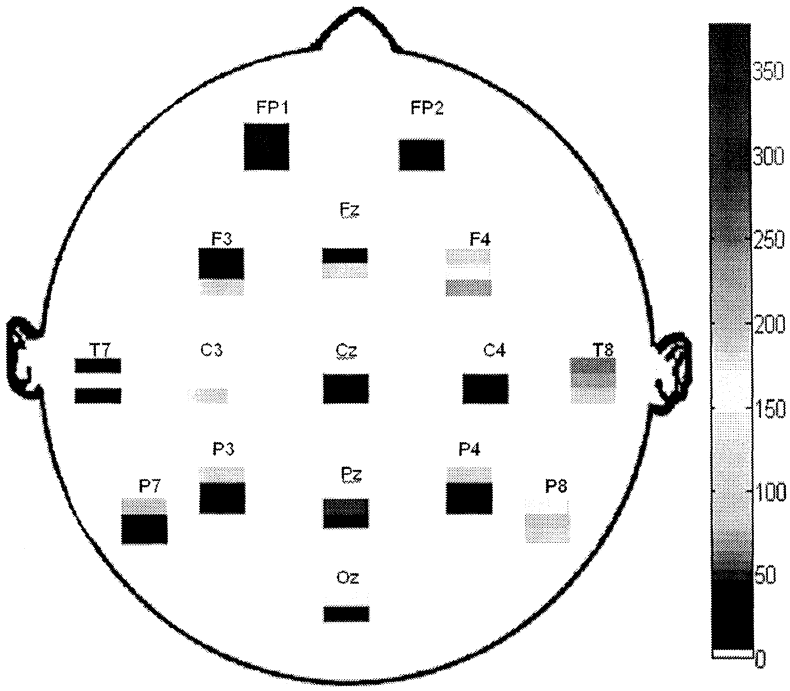

Figure 4.9: Distribution of most frequently occurring novel electrode responses in combinations of three performing better than 70% using MLPs for Cohort B.

The results from the top 5 ensembles are shown in Table 4.21. Despite using slightly different electrode locations from Cohort A, performances were still high. Performance figures are again shown as averages of 10 trials since MLPs were used with the top average being 84.68% from the sum rule. The change of top electrode responses is most likely due to some combination of different equipment, different artifact rejection, and a slightly different electrode montage used. Overall the performance is still in the same neighborhood as the performances achieved from Cohort A.

Table 4.21: Top 5 performing ensembles for Cohort B using the top 6 three-tuples each from target and novel responses using MLPs for Cohort B.

Electrode/Frequency/Stimulus	Average (%)	CI (%)	Sensitivity (%)	Specificity (%)	PPV (%)	Comb Rule
Oz_2, $P7_1$, $C3_3$, $F4_2$, $C3t_1$, $FP2t_2$, $P8t_1$	84.68	2.17	91.94	77.42	80.32	Sum
Oz_2, $C3_3$, $F4_2$, Pzt_1, Fzt_3, $FP2t_2$, $P8t_1$	84.19	1.85	92.26	76.13	79.44	Sum
Oz_2, $C3_3$, $T8_4$, Pzt_1, Fzt_3, $FP2t_2$, $P8t_1$	84.03	3.11	91.61	76.4	79.59	Sum
Oz_2, $P7_1$, $F4_2$, Fzt_3, $C3t_1$, $FP2t_2$, $P8t_1$	83.87	2.52	92.26	75.48	79.21	Sum
Oz_2, $P7_1$, $T8_4$, $F4_2$, Fzt_3, $FP2t_2$, $P8t_1$	83.71	2.34	90.97	76.45	79.57	Sum

4.8 DATA FUSION RESULTS FROM COHORT B USING SVMS

As in Cohort A, Cohort B was next tested using SVMs to see if performance numbers followed the same trend when comparing MLPs with SVMs. The top 6 most frequently appearing electrodes were selected from each stimulus by means of an exhaustive search through ensembles of 3 according to Figure 4.1. The scalp histograms are shown in Figures 4.10 and 4.11 for target and novel stimuli. To achieve approximately 1,000 results, thresholds used were 69% for target responses and 72% for novel responses. The selected electrodes were: novel F4 2-4 Hz, Oz 1-2 Hz, T8 2-4 Hz, F3 4-8 Hz, Oz 2-4 Hz, P8 2-4 Hz and target C3 1-2 Hz, P8 2-4 Hz, FP2 2-4 Hz, C4, 1-2 Hz, P4 4-8 Hz, Pz 1-2 Hz. These locations are similar, but again not identical, to the locations selected for MLPs.

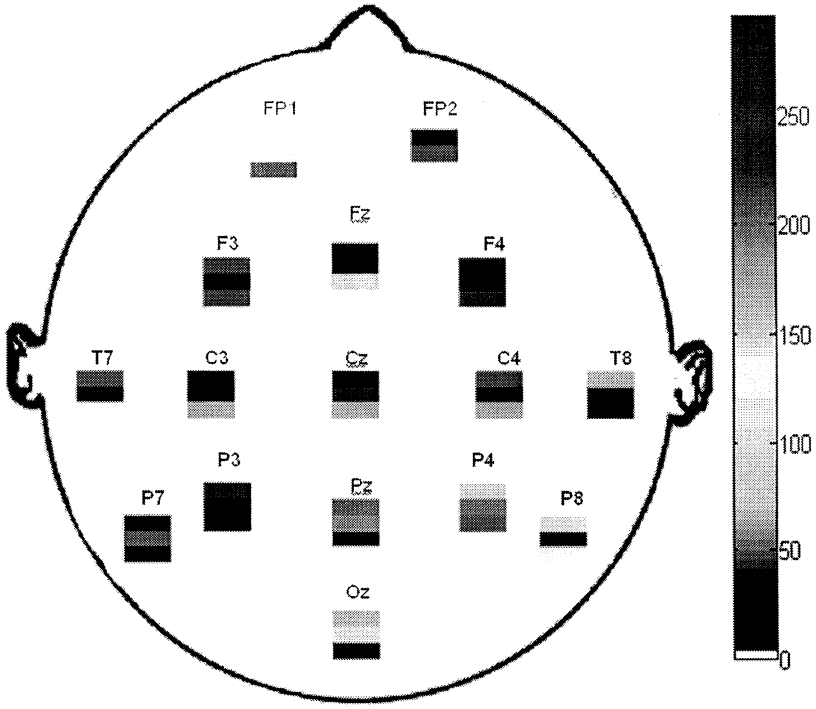

Figure 4.10: Distribution of most frequently occurring target electrode responses in combinations of three performing better than 69% using SVMs for Cohort B.

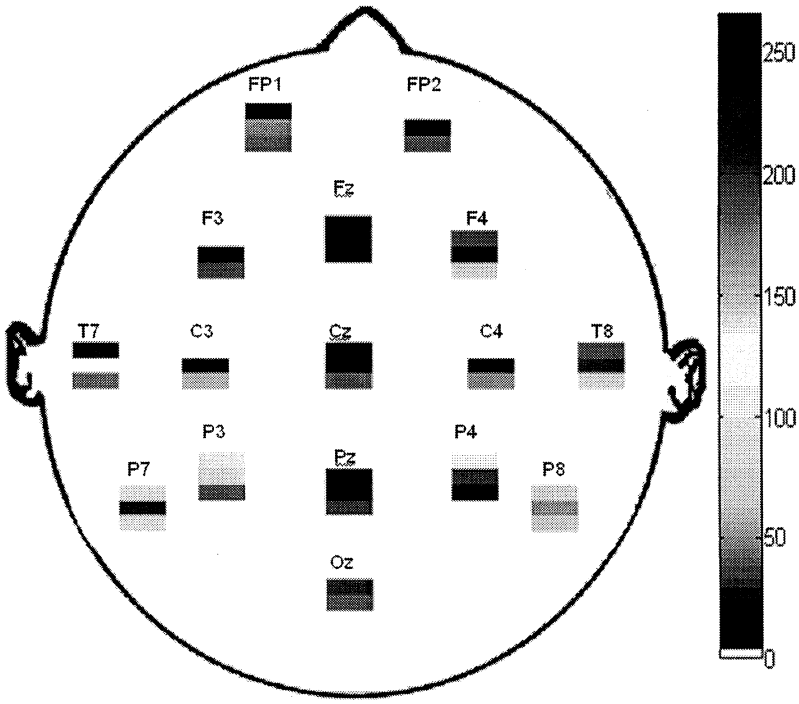

Figure 4.11: Distribution of most frequently occurring novel electrode responses in combinations of three performing better than 72% using SVMs for Cohort B.

The top 5 ensembles of 7 are shown in Table 4.22. Since SVMs were used, only a single performance value is shown. The top performing ensemble reached 93.55%, by far the best performance reached by any ensemble tested thusfar. These numbers also agree with the previous numbers showing that the SVMs can outperform the MLPs in a fraction of the time.

Table 4.22: Top 5 performing ensembles for Cohort B using the top 6 three-tuples each from target and novel responses using SVMs for Cohort B.

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$F4_2$, $T8_2$, $F3_3$, Oz_2, $P8_2$, $C3t_1$, $FP2t_2$	93.55	93.55	93.55	93.55
$F4_2$, Oz_1, $T8_2$, $F3_3$, Oz_2, $P8_2$, $P8t_2$	91.94	87.10	96.77	96.43
$F4_2$, Oz_1, $T8_2$, $F3_3$, Oz_2, $C3t_1$, $P8t_2$	91.94	90.32	93.55	93.33
$F4_2$, Oz_1, $T8_2$, $F3_3$, Oz_2, $P8t_2$, Pzt_1	91.94	87.10	96.77	96.43
$F4_2$, Oz_1, $T8_2$, $F3_3$, $P8_2$, $C3t_1$, $P8t_2$	91.94	87.10	96.77	96.43

Since for Cohort A, performance was higher when just using an ensemble of the best performing single SVM classifiers, the same procedure according to Figure 4.6 was tested with Cohort B. Novel and target responses were first tested individually and then combined together. Table 4.23 shows the results from the target responses, Table 4.24 shows the results from the novel responses, and Table 4.25 shows the mixed stimulus results.

Table 4.23: Top 5 performing ensembles using the top 15 three-tuples from target responses using SVMs for Cohort B.

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$P8_2$, $C3_1$, Pz_1, $C4_1$, $P4_3$, Oz_1, $FP1_1$	85.48	77.42	93.55	92.31
$P8_2$, $C3_1$, Pz_1, $P4_3$, Oz_2, Oz_1, $FP1_1$	85.48	80.65	90.32	89.29
$P8_2$, $C3_1$, $FP2_2$, Pz_1, $P8_3$, Oz_2, $FP1_1$	83.87	77.42	90.32	88.89
$P8_2$, $C3_1$, Pz_1, $C4_1$, $P8_3$, $FP1_1$, Fz_3	83.87	77.42	90.32	88.89
$P8_2$, $C3_1$, Pz_1, $C4_1$, Oz_2, Oz_1, $FP1_1$	83.87	77.42	90.32	88.89

Table 4.24: Top 5 performing ensembles using the top 15 three-tuples from novel responses using SVMs for Cohort B.

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Oz_1, $F4_2$, $T8_1$, $P8_2$, $F3_3$, $P4_1$, $P7_3$	91.94	93.55	90.32	90.63
Oz_1, $F4_2$, $T8_1$, $P8_2$, $F3_3$, $P3_1$, $P7_3$	91.94	90.32	93.55	93.33
Oz_1, $F4_2$, Oz_2, $T8_1$, $P8_2$, $F3_3$, $P3_1$	90.32	87.10	93.55	93.10
Oz_1, $F4_2$, Oz_2, $T8_1$, $P8_2$, $F3_3$, $P7_3$	90.32	83.87	96.77	96.30
Oz_1, $F4_2$, Oz_2, $T8_1$, $F3_3$, $P7_3$, $P3_2$	90.32	87.10	93.55	93.10

Table 4.25: Top 5 performing ensembles using the top 15 three-tuples from target and novel responses combined using SVMs for Cohort B.

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Oz_1, $F4_2$, $T8_1$, $P8_2$, $F3_3$, $FP2t_2$, $P3_1$	93.55	85.29	91.18	90.63
Oz_1, $F4_2$, $T8_1$, $P8_2$, $FP2t_2$, Pzt_1, $P3_1$	93.55	87.88	83.33	82.86
Oz_1, $F4_2$, $T8_1$, $F3_3$, $P8t_2$, $C3t_1$, $P7_1$	93.55	84.85	85.71	84.85
Oz_1, $F4_2$, $T8_1$, $P8t_2$, $C3t_1$, $FP2t_2$, $P3_1$	93.55	87.50	73.81	71.79
Oz_1, $F4_2$, Oz_2, $T8_1$, $P8_2$, $F3_3$, $P8t_2$	91.94	85.29	88.24	87.88

For Cohort B, the method of simply selecting the highest performing single SVM classifiers worked well, but not to the same degree as the exhaustive ensemble of 3 method. The top performance came from the mixed stimulus group at 93.55%. All ensembles still performed significantly better than the single classifiers and overall performance figures were higher for Cohort B than for Cohort A.

4.9 20% TEST RESULTS FROM COHORT A

A question was raised as to the validity of the leave-one-out training and test method used thus far. Since the selection of classifiers to be used in the ensemble is based on the full leave-one-out training and testing with the full set of data, the final performance is

not based on *previously unseen* data. To validate the process, an alternate training and testing method was tested to see if the previous performance results could be verified.

This new method is basically the same leave-one-out procedure nested inside of a 5-fold cross validation. The original set of data is split into 5 folds of approximately 20% of the total data set each. Each block of 20% is removed one at a time while the remaining 80% of the data is used in a leave-one-out fashion as previously done. The ensembles are generated based on the performance of the 80% of training data. Each classifier is then retrained with the full 80% of training data and those same ensembles are then used to test on the 20% of data that were removed and never seen by any classifier. The difference between this method and the 5-fold method used earlier is the inclusion of the nested leave-one-out training within the 5-fold cross validation. This process is repeated five times such that each fold serves as the test data once, with unique ensembles creates during each fold based solely on the training data of that fold. For increased validation, the entire process was also repeated 5 times with a completely random selection of patients in each fold every time.

This new training and testing method was first tried using MLPs on the data from Cohort A. Based on the superior performance in all other tests, only the sum rule was used for combinational purposes. The resulting performances for 5 trials, each containing 5 folds is shown in Table 4.26. The numbers are the top performing ensembles from each trial.

Table 4.26: Classification performance from 20% test method using MLPs for Cohort A.

	Classification Performance (%)				
	Trial 1	**Trial 2**	**Trial 3**	**Trial 4**	**Trial 5**
Fold 1	78.57	78.57	78.57	71.43	85.71
Fold 2	85.71	92.86	92.86	85.71	85.71
Fold 3	92.86	85.71	78.57	92.86	71.43
Fold 4	92.86	92.86	64.29	85.71	78.87
Fold 5	86.67	73.33	86.67	80	93.33
Average	87.33	84.66	80.19	83.14	83.01

Ensembles were created in a similar fashion to the alternate method used in the SVMs. The 12 three-tuples with the best single classifier performance were selected to create ensembles of 7 classifiers according to Figure 4.6. The best average for a single trial was 87.33% with the average across all folds from all 5 trials at 84.24%, which is right in the range of the leave-one-out method including the entire set of data. The electrodes used to create the top performing ensemble for each fold are shown in Table 4.27.

Table 4.27: Three-tuples used in the best ensemble for each trial and fold of the 20% method using MLPs for Cohort A.

Trial 1

Fold 1	$T8_1$	PZ_2	$P8_2$	PZ_3	$C3t_2$	$FP1_1$	$P3_3$
Fold 2	$O2_1$	$P8_2$	PZ_3	PZ_1	$FP1_1$	$T8_1$	$FP1_2$
Fold 3	$P7_2$	$T8_1$	FZ_2	PZ_2	PZ_3	PZt_2	$O2_1$
Fold 4	PZ_3	$FP2_2$	$P3_2$	$P4_2$	PZ_2	$P3t_1$	$FP1_1$
Fold 5	$P3_1$	$FP1_2$	$FP2t_1$	FZ_2	$FP1_1$	$P3_2$	PZ_2

Trial 2

Fold 1	$P3_1$	$FP1t_1$	$O2_1$	$FP1_1$	$P3_3$	PZ_1	PZ_2
Fold 2	PZ_2	CZ_2	$P4_2$	$P8_1$	$P8_2$	FZt_1	$P3_3$
Fold 3	FZ_2	$O2t_1$	$P4_2$	$P3_1$	PZ_2	$O2_1$	PZ_1
Fold 4	$FP2_2$	$P3_2$	$P3_3$	FZ_1	PZt_2	$T7t_3$	$O2_1$
Fold 5	$F8_1$	$P3_2$	$T8_1$	$FP2t_1$	$P4t_2$	$F7t_3$	$C4_3$

Trial 3

Fold 1	$P8_2$	$P4_1$	PZ_2	$F3_3$	$P4t_2$	$P8t_1$	$T8_2$
Fold 2	$P4t_2$	$T8_1$	FZt_1	$O2_1$	$FP1t_1$	$P3t_1$	$P7t_3$
Fold 3	$P3_3$	$P3_2$	$P4_1$	$F8t_1$	$FP2t_1$	FZ_2	$T7t_3$
Fold 4	PZ_2	$P8_2$	$P3_2$	$FP1_3$	$P4_2$	$F4t_1$	$P3t_1$
Fold 5	$P4t_2$	PZ_3	PZ_2	$FP1_1$	$O1_1$	$P4_2$	$O2_1$

Trial 4

Fold 1	$P3_1$	$P3_3$	PZ_3	$P3_2$	PZ_1	$P4_2$	$P7_2$
Fold 2	$T8_1$	$F8_1$	$FP1_1$	$P4_2$	PZ_2	CZt_2	PZ_3
Fold 3	$O2_1$	$P3_3$	PZ_2	$C4_2$	$FP1_1$	$FP1_2$	$P8_2$
Fold 4	$P7_2$	$P4t_2$	PZ_2	$P3_2$	$P3t_1$	$P8_2$	$FP1t_1$
Fold 5	$FP1_2$	CZ_2	PZ_1	$C3t_3$	$F8_1$	$O2_2$	$P4_2$

Trial 5

Fold 1	$P7_2$	$P7t_3$	PZ_2	$T8t_2$	$FP2t_1$	$P3_2$	$P4_2$
Fold 2	$P3_3$	PZ_2	CZ_2	PZ_3	$O2_1$	$FP1_1$	$P4_2$
Fold 3	$F3_3$	$FP2_2$	$FP1_2$	$P8_2$	$P8_3$	$F8t_1$	$O1_2$
Fold 4	PZ_3	PZ_2	$P3t_1$	$P8_2$	$FP1_2$	$FP2t_1$	$F3t_1$
Fold 5	PZ_3	$P8_2$	$F3_3$	$C3_2$	$F8_2$	$P3_1$	$P3_3$

To gain a better visual of which three-tuples were selected most often, the scalp histograms were again used. The scales are significantly smaller since there are only 175 possible selections from five 5-fold trials. Figure 4.12 shows the scalp histogram from target responses and 4.13 shows the histogram from novel responses. Despite using significantly fewer instances, the histograms are not very much different than the leave-

one-out histograms generated earlier. The combination similar classification performance and similar scalp histograms from the 20% test and the leave-one-out methods show that the leave-one-out method was valid in its testing.

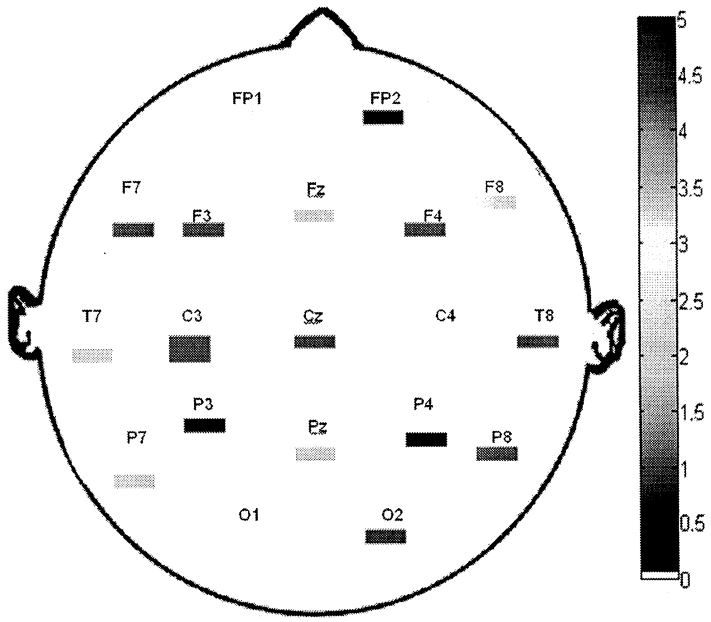

Figure 4.12: Scalp histogram from 20% test data method from target responses using MLPs for Cohort A.

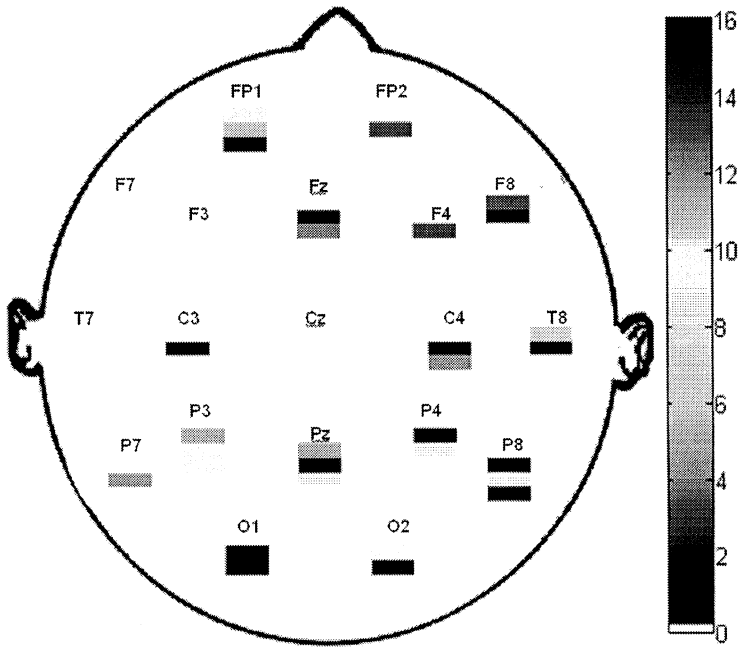

Figure 4.13: Scalp histogram from 20% test data method from novel responses using MLPs for Cohort A.

The same process was completed using SVMs instead of MLPs for a comparison. The results from that test are shown in Table 4.28 with the three-tuples used in each ensemble listed in Table 4.29.

Table 4.28: Classification performance from 20% test method using SVMs for Cohort A.

	Classification Performance (%)				
	Trial 1	**Trial 2**	**Trial 3**	**Trial 4**	**Trial 5**
Fold 1	100	78.57	92.86	100	92.86
Fold 2	78.57	100	100	78.57	92.86
Fold 3	100	92.86	92.86	100	78.57
Fold 4	92.86	92.86	92.86	9286	78.57
Fold 5	80	80	80	80	93.33
Average	90.28	88.86	91.71	90.28	87.23

Table 4.29: Three-tuples used in the best ensemble for each trial and fold of the 20% method using SVMs for Cohort A.

Trial 1

Fold 1	$P3_3$	CZ_2	$P3_2$	$P8t_1$	$P4t_2$	$FP1t_1$	PZt_2
Fold 2	PZ_2	$P4_1$	$F8_1$	$P8_2$	FZ_2	CZ_1	$O1t_2$
Fold 3	PZ_1	$P3_2$	$P7_2$	$P7_3$	$F8t_3$	$P7t_3$	$FP1t_2$
Fold 4	PZ_2	FZ_2	$P3_3$	$P4_2$	$P8_2$	$P7t_3$	FZt_1
Fold 5	$O2_1$	CZ_2	PZ_1	PZ_2	$F4_1$	$P4t_2$	$FP2t_2$

Trial 2

Fold 1	FZ_2	$F4_2$	PZ_2	$P3_2$	$FP2t_2$	PZt_2	$P4t_2$
Fold 2	PZ_2	$P3_3$	$P8_2$	$C4_2$	PZ_1	$C3t_2$	$FP1t_2$
Fold 3	CZ_2	PZ_1	$P3_3$	$T8_3$	$FP2t_2$	$FP1t_2$	$F8t_1$
Fold 4	$P3_3$	$P8_2$	CZ_2	$P3_1$	PZ_2	$FP1t_2$	$O2t_1$
Fold 5	PZ_2	$P3_2$	PZ_3	$P3_3$	FZ_2	$P3t_2$	$P3t_3$

Trial 3

Fold 1	$P3_1$	$FP1_2$	$T7_2$	$FP1t_2$	$FP2t_2$	$FP1t_1$	$P4t_2$
Fold 2	PZ_2	$C4_2$	CZ_2	PZ_3	$P3_1$	$P4t_2$	$FP2t_2$
Fold 3	$P3_2$	P33	FZ_2	$C4_2$	$F8t_2$	$F7t_3$	$F8t_1$
Fold 4	PZ_2	$P3_1$	$FP1_2$	$P3_3$	CZ_2	$P4t_2$	$FP2t_3$
Fold 5	PZ_2	PZ_1	CZ_2	$T8_3$	$O2_1$	$FP2t_1$	PZt_1

Trial 4

Fold 1	$O1_1$	PZ_1	$P3_3$	$P4_3$	$P7_2$	$P8_1$	$FP1t_1$
Fold 2	$P3_3$	CZ_2	$P3_2$	$P8t_1$	$P4t_3$	$FP1t_1$	PZt_2
Fold 3	PZ_2	$P4_1$	$F8_1$	$P8_2$	FZ_2	CZ_1	$O1t_2$
Fold 4	PZ_2	FZ_2	$P3_3$	$P4_2$	$P8_2$	$P7t_3$	FZt_1
Fold 5	$O2_1$	CZ_2	PZ_1	PZ_2	$F4_1$	$P4t_2$	$FP2t_2$

Trial 5

Fold 1	CZ_2	PZ_2	$P8_1$	$F7_2$	PZt_2	$FP2t_2$	$P7t_3$
Fold 2	CZ_2	$FP1_3$	PZ_1	$P3_2$	PZ_2	$P8t_2$	$FP1t_1$
Fold 3	$C3_2$	$P3_1$	$P3_3$	PZ_2	$P3_2$	$FP2t_2$	$P3t_3$
Fold 4	PZ_2	$FP1_1$	$C4_2$	$P4_1$	$O2_1$	PZ_1	$P8_1$
Fold 5	$FP1_1$	$C3_2$	$P8_2$	PZ_3	$P3_3$	$P4_1$	$P8t_1$

As in the leave-one-out training and testing procedure, the SVM ensembles performed significantly better with a top average trial of 91.71% and an average across all trials of 89.67%. The difference in computational expense was extremely apparent during the course of this test. The entire process was completed in roughly 3 hours using

SVMs on a single computer, while the MLPs took approximately 45 hours utilizing 10 different computers for a total computational time of 450 hours.

The histograms were again generated and shown in Figure 4.14 for target responses and 4.15 for novel responses. Again, the histograms are similar to the leave-one-out histograms.

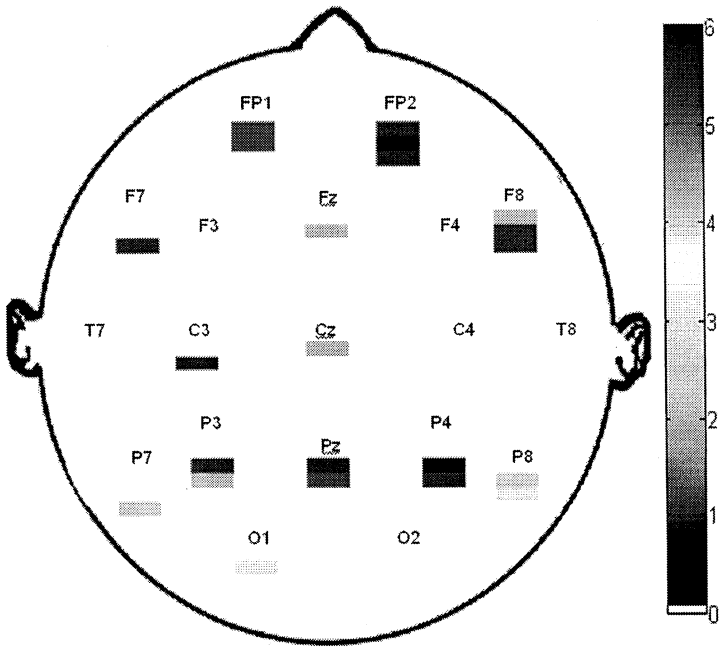

Figure 4.14: Scalp histogram from 20% test data method from target responses using SVMs for Cohort A.

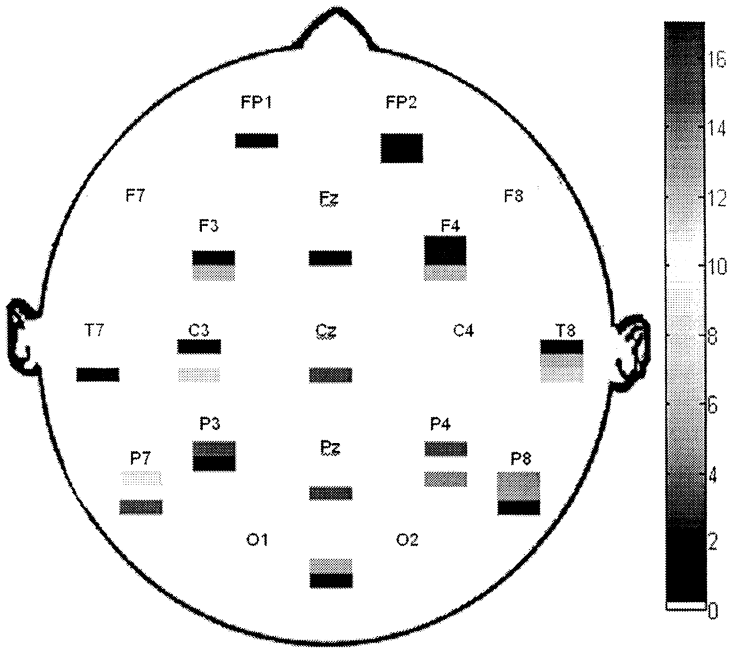

Figure 4.15: Scalp histogram from 20% test data method from novel responses using SVMs for Cohort A.

4.10 20% TEST RESULTS FROM COHORT B

The 20% test method was performed using SVMs on Cohort B. The experimental setup was identical except the number of subjects in the 20% dropped from 14 or 15 to 12 or 13. The resulting fold performances are shown in Table 4.30.

Table 4.30: Classification performance from 20% test method using SVMs for Cohort B.

	Classification Performance (%)				
	Trial 1	**Trial 2**	**Trial 3**	**Trial 4**	**Trial 5**
Fold 1	100	91.67	100	100	100
Fold 2	91.67	91.67	91.67	100	83.33
Fold 3	91.67	83.33	91.67	91.67	100
Fold 4	100	92.31	92.31	100	92.31
Fold 5	91.67	91.67	100	91.67	91.67
Average	95.002	90.13	95.13	96.668	93.462

For Cohort B, again there is slightly higher performance numbers than from Cohort A. The top trial yielded an average of 96.66% correct classification of the test data. The overall average was 94.08%, which is also right in the range of the leave-one-

out training method. The three-tuples used in the best ensembles are shown in Table 4.31.

Table 4.31: Three-tuples used in the best ensemble for each trial and fold of the 20% method using SVMs for Cohort B.

Trial 1

Fold 1	$F4_2$	$T8_3$	$P7_1$	$C3_3$	$FP2t_2$	$C4t_1$	PZt_1
Fold 2	PZ_3	$F4_3$	$P7_1$	$P8_2$	$F3_3$	$C3t_1$	OZt_2
Fold 3	$T8_2$	OZ_1	$F4_2$	$C3_3$	$FP1t_1$	$FP2t_2$	$C4t_3$
Fold 4	$F4_2$	$F3_3$	OZ_2	$F4_1$	$P8t_1$	$FP2t_2$	$FP2t_1$
Fold 5	$P7_1$	$F3_3$	$FP1_2$	$P8_2$	$P8_1$	OZ_2	$P8t_2$

Trial 2

Fold 1	$F4_2$	$F4_3$	$P7_3$	$P4t_2$	CZt_3	$C3t_1$	$P8t_1$
Fold 2	$F4_2$	$T8_3$	$F4_3$	$P8_1$	OZ_2	$C4t_1$	FZt_2
Fold 3	PZ_3	$P4_3$	$C3_3$	OZ_2	$F4_3$	$P8_2$	$P7_1$
Fold 4	$P7_1$	OZ_2	OZ_1	$P8_2$	$F3_3$	$P4t_1$	$P8t_2$
Fold 5	$T8_3$	$F4_2$	$P8_2$	$T8_2$	$P4_1$	$C3_1$	$FP2_2$

Trial 3

Fold 1	$P8_1$	$P8_3$	CZ_3	$P4_3$	$P4_1$	$P4t_3$	$P8t_2$
Fold 2	$P7_1$	$P7_3$	$F4_2$	$P8_2$	$P3_1$	$P3t_2$	$FP2t_2$
Fold 3	$P7_1$	$C3_3$	$P8_2$	$F3_3$	$F4_2$	$P4t_1$	$FP2t_2$
Fold 4	$T8_3$	OZ_2	$P8_2$	$C4t_3$	$P8t_2$	$FP2t_2$	$P8t_1$
Fold 5	$F4_2$	$T7_3$	$C3_3$	$F3_3$	$F3_2$	OZ_2	$FP1t_1$

Trial 4

Fold 1	$P4_1$	$T8_1$	OZ_2	$F4_2$	OZ_1	$F3_3$	$C3t_1$
Fold 2	$P8_2$	$P3_1$	$F4_2$	$T8_2$	OZ_2	$F3_2$	$P4t_3$
Fold 3	$P8_2$	FZ_3	$P7_1$	$F3_3$	PZt_1	$P8t_2$	$C3t_1$
Fold 4	$F4_2$	OZ_1	$C3_3$	$F4_3$	$T8_3$	$P8t_1$	FZt_2
Fold 5	$F3_3$	$C3_3$	OZ_2	$P4_3$	$P7_3$	PZt_1	$P4t_3$

Trial 5

Fold 1	$F4_2$	$T8_3$	$F3_3$	OZ_2	$F4_1$	$P3_1$	PZt_1
Fold 2	$P8_2$	$F4_2$	OZ_2	CZ_3	PZt_1	$C3t_1$	$FP2t_2$
Fold 3	$F4_2$	$T8_2$	OZ_1	OZ_2	$P8_3$	$F3_3$	OZt_2
Fold 4	$P8_2$	CZ_3	$T8_2$	$P4_3$	OZ_2	$FP1t_2$	$FP2t_1$
Fold 5	$FP2_3$	$P3_2$	$F4_2$	$P8_1$	PZ_3	OZ_2	OZt_2

The scalp histograms were generated and shown in Figures 4.16 and 4.17 for target and novel. Similarities are again seen between these and the previous histograms.

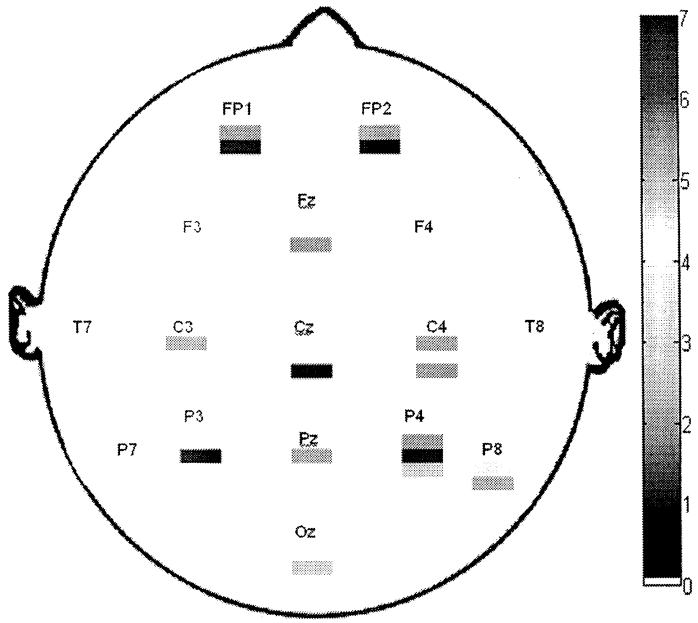

Figure 4.16: Scalp histogram from 20% test data method from target responses using SVMs for Cohort B.

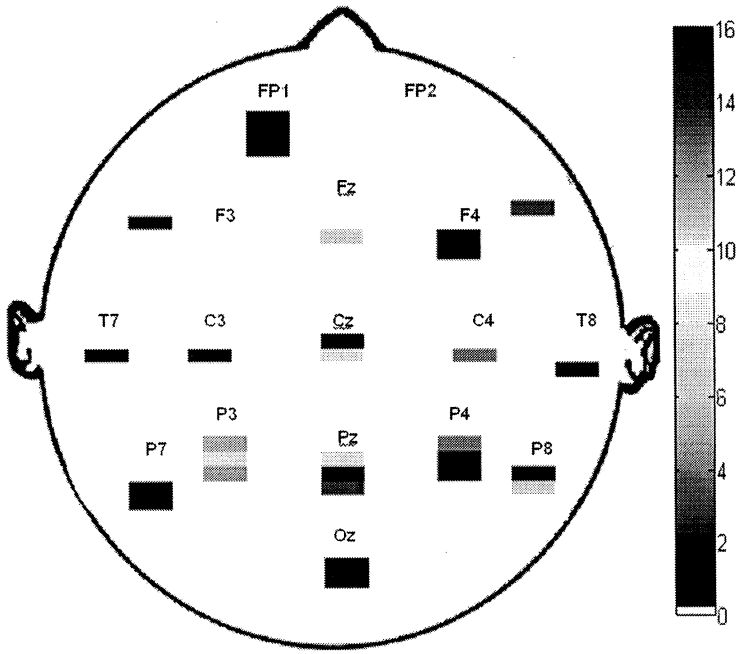

Figure 4.17: Scalp histogram from 20% test data method from novel responses using SVMs for Cohort B.

4.11 MIXED COHORT RESULTS

As a final test, the two cohorts were combined together to see how the entire cohort would perform in terms of classification. This combination brought the total cohort size to 133 subjects. SVMs were used due to their higher performance on the cohorts separately as well as the speed of training and consistency of results. Classifiers were trained individually first on the entire set of three-tuples. Since Cohort B only used 16 electrodes for EEG recording, only those electrodes were selected from Cohort A. In the case of Oz, the O2 electrode was used from Cohort A. The individual electrode performances are shown in Tables 4.32 and 4.33 for target and novel responses, respectively.

Table 4.32: Results for classifiers trained on target stimuli responses from all 16 electrodes at all 3 feature levels for the subjects of both cohorts using SVMs.

Electrode Response	1-2 Hz	2-4 Hz	4-8 Hz
C3	55.64%	47.37%	57.89%
C4	57.89%	54.89%	48.87%
CZ	53.38%	47.37%	45.86%
F3	48.87%	54.89%	**63.16%**
F4	53.38%	52.63%	51.13%
FZ	57.89%	**60.90%**	47.37%
FP1	57.14%	54.14%	41.35%
FP2	58.65%	58.65%	51.88%
OZ/O2	58.65%	57.14%	51.13%
PZ	**63.91%**	**60.90%**	42.11%
P3	**63.91%**	**63.16%**	53.38%
P4	55.64%	**65.41%**	50.38%
P7	55.64%	55.64%	**60.90%**
P8	**69.17%**	48.87%	46.62%
TZ	53.38%	47.37%	50.38%
T8	54.14%	51.88%	**61.65%**

Table 4.33: Results for classifiers trained on novel stimuli responses from all 16 electrodes at all 3 feature levels for the subjects of both cohorts using SVMs.

Electrode Response	1-2 Hz	2-4 Hz	4-8 Hz
C3	59.40%	**60.15%**	57.89%
C4	50.38%	52.63%	56.39%
CZ	**61.65%**	49.62%	54.89%
F3	49.62%	37.59%	59.40%
F4	45.11%	57.14%	50.38%
FZ	42.11%	53.38%	57.14%
FP1	48.87%	53.38%	43.61%
FP2	48.87%	52.63%	45.86%
OZ/O2	**66.17%**	56.39%	51.13%
PZ	**64.66%**	**60.90%**	54.89%
P3	**69.92%**	**61.65%**	**61.65%**
P4	**73.68%**	**65.41%**	59.40%
P7	**64.66%**	**60.90%**	54.14%
P8	**63.91%**	**66.92%**	57.89%
TZ	53.38%	50.38%	45.11%
T8	**60.90%**	54.89%	56.39%

Overall, the single classifier performance numbers are close to the performance numbers from the two cohorts individually. The highest performing target three-tuple was P8 1-2 Hz at 69.17% and highest performing novel three-tuple was P4 1-2 Hz at 73.68%. The parietal electrodes again have the most performances over 60% as shown by bold and highlight in the table.

Ensembles were then created using the top 15 three-tuples selected automatically by the algorithm according to Figure 4.6 and combined exhaustively in ensembles of 7 classifiers. Ensembles were first created independently for each stimulus, and then for both stimuli together. Table 4.34 shows the results from target responses, Table 4.35 for novel responses, and 4.36 for combined responses.

Table 4.34: Top 5 performing ensembles using the top 15 three-tuples from target responses using SVMs for both cohorts.

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$P8_1$, $P4_2$, $P3_1$, $F3_3$, $T8_3$, Fz_2, $FP2_1$	78.95	72.31	85.29	82.46
$P8_1$, $P4_2$, $P3_2$, $F3_3$, $P7_3$, Fz_2, $FP2_1$	78.95	73.85	83.82	81.36
$P8_1$, $P3_1$, $P3_2$, $F3_3$, $T8_3$, $P7_3$, $FP2_2$	78.95	70.77	86.76	83.64
$P8_1$, $P3_2$, $F3_3$, $T8_3$, $P7_3$, $FP2_1$, $FP2_2$	78.95	75.38	82.35	80.33
$P8_1$, $P4_2$, Pz_1, $P3_1$, $F3_3$, $T8_3$, $FP2_2$	78.20	69.23	86.76	83.33

Table 4.35: Top 5 performing ensembles using the top 15 three-tuples from novel responses using SVMs for both cohorts.

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$P4_1$, $P3_1$, $P8_2$, Oz_1, $P4_2$, $P7_1$, $P8_1$	80.45	75.38	85.29	83.05
$P4_1$, $P3_1$, $P8_2$, Oz_1, $P7_1$, $P8_1$, Pz_2	79.70	75.38	83.82	81.67
$P4_1$, $P3_1$, Pz_1, $P3_3$, Cz_1, $P7_2$, $T8_1$	79.70	70.77	88.24	85.19
$P4_1$, $P3_1$, $P8_2$, Pz_1, Pz_2, $T8_1$, $C3_2$	78.95	72.31	85.29	82.46
$P4_1$, $P3_1$, $P8_2$, Oz_1, $P4_2$, $P8_1$, $P7_2$	78.20	73.85	82.35	80.00

Table 4.36: Top 5 performing ensembles using the top 15 three-tuples from target and novel responses combined using SVMs for both cohorts.

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$P4_1$, $P8t_1$, $P8_2$, Oz_1, $P4_2$, $P4t_2$, Pz_1	82.71	74.65	81.69	80.30
$P4_1$, $P8t_1$, $P8_2$, Oz_1, $P4_2$, $P4t_2$, Pzt_1	81.95	70.67	79.22	76.81
$P4_1$, $P8t_1$, $P8_2$, Oz_1, $P4_2$, $P4t_2$, $P3t_3$	81.95	70.67	78.48	75.36
$P4_1$, $P8t_1$, $P8_2$, Oz_1, $P4_2$, $P4t_2$, $P3t_2$	81.20	70.27	78.67	76.47
$P4_1$, $P3_1$, $P8t_2$, Oz_1, $P4_2$, $P4t_2$, Pz_1	81.20	69.23	78.67	77.14

The classification performances are noticeably lower than in either cohort independently, only 82.71% versus the 88% and 93% seen from Cohorts A and B, respectively. This decrease in performance could be due to a number of factors including unique testing locations, a larger cohort with more subjects exhibiting features closer to

the boundaries drawn by the SVM, less intricate boundaries created by the SVM resulting in more misclassifications, and other problems like noisy patient data, which could have been brought on by the different artifact rejection schemes. On the positive side, however, the best performance of 82.71% is still significantly higher than the local physician average of 75%.

4.11.1 CROSS COHORT FEATURE SELECTION

The next test was to see how each cohort would perform during testing when the other cohort was used for training (ie. train on Cohort A and test on Cohort B). This analysis was completed using the best performing ensembles from each of the respective cohorts. First, classifiers were trained using data from Cohort B. The best ensembles were selected according to table 4.25. These ensembles were then used to test on the data from Cohort A. The results are shown in Table 4.37.

Table 4.37: SVM classifiers trained on Cohort B and tested on Cohort A using the best ensembles from Cohort B.

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Oz_1, $F4_2$, $T8_1$, $P8_2$, $F3_3$, $FP2t_2$, $P3_1$	46.48	41.03	45.95	44.44
Oz_1, $F4_2$, $T8_1$, $P8_2$, $FP2t_2$, Pzt_1, $P3_1$	60.56	48.89	56.76	57.89
Oz_1, $F4_2$, $T8_1$, $F3_3$, $P8t_2$, $C3t_1$, $P7_1$	42.24	43.90	32.43	41.86
Oz_1, $F4_2$, $T8_1$, $P8t_2$, $C3t_1$, $FP2t_2$, $P3_1$	56.34	53.06	45.10	48.15
Oz_1, $F4_2$, Oz_2, $T8_1$, $P8_2$, $F3_3$, $P8t_2$	59.15	47.73	56.76	56.76

The resulting performances from the ensembles are significantly lower than the low 90% range exhibited in Table 4.25. Only one ensemble performed better than 60%, with two of the five performing only in the 40% range. These low performances could be

due to either differences in the data based on the collection and artifact removal or physiologic differences in the data itself from the two different cohorts.

The same process was used to train classifiers using Cohort A and test Cohort B using the best performing ensembles from Table 4.20. The results are shown in Table 4.38.

Table 4.38: SVM classifiers trained on Cohort A and tested on Cohort B using the best ensembles from Cohort A.

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Pz_2, $C3_2$, $P3_3$, $FP2t_1$, $P8t_2$, $FP1t_2$, $P8_1$	59.68	70.97	48.39	57.89
$C3_2$, $P3_3$, $FP2t_1$, $C4_2$, $P8t_2$, $P8_1$, $P4t_2$	59.68	71.43	38.71	56.82
Pz_2, $C3_2$, $P3_3$, $P8t_2$, $P8_2$, $FP1t_2$, $P8_1$	66.13	63.64	64.52	65.63
Pz_2, $C3_2$, $FP2t_1$, $P4t_2$, $P8t_2$, $P8_1$, $FP1t_1$	58.06	55.88	54.84	57.58
Pz_2, $P3_3$, $FP2t_1$, $P8t_2$, $FP1t_2$, $P8_1$, $P8t_1$	70.97	61.54	64.52	68.57

The performance numbers for this test are significantly higher with the lowest performance being 58% and the highest performance at almost 71%. These performance figures are still considerably lower than the high 80% range shown in Table 4.20, due to the same reasons.

4.11.2 CROSS COHORT REANALYSIS

Previously, only the top five ensembles were examined from each of the alternate cohorts. To validate any differences, a complete reanalysis was performed for each cohort. For this experiment, one cohort was used to fully train classifiers on all electrode locations from both target and novel. Those classifiers were then tested on the alternate cohort. The top 15 performing classifiers based on this training were then combined into ensembles of seven classifiers. This process mimics the process used for each cohort

previously, with the only change coming from the test data. SVMs were used as the classifier and ensembles were created for each stimuli separately first, and then both stimuli combined. The results for training on Cohort B and testing on Cohort A are shown in tables 4.39-41.

Table 4.39: Top 5 performing ensembles using the top 15 three-tuples from target responses using SVMs trained on Cohort B, tested on Cohort A

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$P3_1$, $F3_2$, Fz_2, $P8_1$, Oz_1, $P3_3$, $F3_3$	77.46	79.41	75.68	75.00
$P4_1$, $F3_2$, $P8_1$, Pz_1, $T8_3$, $P3_3$, $F4_1$	77.46	79.41	75.68	75.00
$P4_1$, $F3_2$, $P8_1$, $P7_1$, $T8_3$, $P3_3$, $F3_3$	77.46	88.24	67.57	71.43
$P3_1$, $P4_1$, $F3_2$, $P8_1$, $T8_3$, $P3_3$, $F3_3$	76.06	79.41	72.97	72.97
$P3_1$, $P4_1$, $F3_2$, $P8_1$, $P3_3$, $F3_3$, $F4_1$	76.06	82.35	70.27	71.79

Table 4.40: Top 5 performing ensembles using the top 15 three-tuples from novel responses using SVMs trained on Cohort B, tested on Cohort A.

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Pz_1, $P4_1$, $P8_2$, $P3_1$, Oz_1, Cz_1, $P8_1$	85.92	91.18	81.08	81.58
Pz_1, $P4_1$, $P8_2$, $P3_1$, Oz_1, Cz_1, Oz_2	84.51	91.18	78.38	79.49
Pz_1, $P4_1$, $P8_2$, $P3_1$, Oz_1, $P8_1$, Oz_2	84.51	91.18	78.38	79.49
Pz_1, $P4_1$, $P3_1$, $P4_2$, Cz_1, $P8_1$, Oz_2	84.51	88.24	81.08	81.08
Pz_1, $P4_1$, $P4_2$, $C3_1$, $P7_3$, $P4_3$, Oz_2	84.51	91.18	78.38	79.49

Table 4.41: Top 5 performing ensembles using the top 15 three-tuples from target and novel responses combined using SVMs trained on Cohort B, tested on Cohort A.

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Pz_1, $P4_1$, $P3t_1$, $P8_2$, $P4t_1$, Fzt_2, Pz_2	83.10	73.81	71.43	72.09
Pz_1, $P8_1$, $P3t_1$, $P8_2$, $P3_1$, $F3t_2$, $P8t_1$	83.10	74.42	70.21	69.57
Pz_1, $P3t_1$, $P8_2$, $P3_1$, $P4t_1$, Fzt_2, Pz_2	83.10	70.45	71.43	72.09
Pz_1, $P3t_1$, $P8_2$, $P3_1$, $P4t_1$, $P8t_1$, Oz_1	83.10	70.45	71.74	70.45
$P4_1$, $P8_1$, $P3t_1$, $P8_2$, $P4t_1$, $P8t_1$, Oz_1	83.10	78.57	64.58	66.00

The five best performing ensembles are listed in each table. Performances reached into the mid 80% range for the responses to novel stimuli and dropped slightly when responses to novel and target were combined. The top performance of almost 86% is close to the best performances from experiment using Cohort A for training and testing, which shows that the possibility of combining the two cohorts is definitely feasible since the performance decrease was not too significant.

The same process was used for training on Cohort A and testing on Cohort B. Results are in Tables 4.42-44 for novel, target, and combined stimuli.

Table 4.42: Top 5 performing ensembles using the top 15 three-tuples from target responses using SVMs trained on Cohort A tested on Cohort B

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$P8_1$, $F3_3$, Fz_1, $P3_1$, Pz_1, $P4_3$, Cz_3	83.87	96.77	70.97	76.92
$P8_1$, $F3_3$, Fz_1, Pz_1, Oz_1, $P4_3$, $P8_2$	82.26	90.32	74.19	77.78
$P8_1$, $F3_3$, Fz_1, $P7_1$, Pz_1, Cz_1, $P4_3$	80.65	90.32	70.97	75.68
$P8_1$, $F3_3$, Fz_1, $P7_1$, $P8_3$, Cz_1, Cz_3	80.65	90.32	70.97	75.68
$P8_1$, $F3_3$, Fz_1, Pz_1, $P3_3$, Cz_1, Cz_3	80.65	83.87	77.42	78.79

Table 4.43: Top 5 performing ensembles using the top 15 three-tuples from novel responses using SVMs trained on Cohort A tested on Cohort B

Electrode/Frequency	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$P8_1$, Oz_1, Pz_1, Cz_1, $P4_1$, $P4_2$, $P3_3$	83.87	83.87	83.87	83.87
$P8_1$, Oz_1, Cz_1, $P8_2$, $P4_1$, $P3_3$, $T8_1$	83.87	80.65	87.10	86.21
$P8_1$, Oz_1, $P8_2$, $P4_1$, $C4_3$, $P3_3$, Pz_2	83.87	83.87	83.87	83.87
$P8_1$, Oz_1, $P4_1$, $P4_2$, $C4_3$, $P3_3$, Pz_2	83.87	80.65	87.10	86.21
$P8_1$, Oz_1, $P8_2$, $P4_1$, $P7_1$, $P3_3$, Pz_2	83.87	83.87	83.87	83.87

Table 4.44: Top 5 performing ensembles using the top 15 three-tuples from target and novel responses combined using SVMs trained on Cohort A tested on Cohort B

Electrode/Frequency/Stimulus	Performance (%)	Sensitivity (%)	Specificity (%)	PPV (%)
$P8t_1$, $P8_1$, Oz_1, Cz_1, $P8_2$, $F3t_3$, Fzt_1	91.94	81.58	72.50	73.81
$P8t_1$, $P8_1$, Oz_1, Cz_1, $F3t_3$, Fzt_1, $P4_1$	91.94	85.71	71.43	71.43
$P8t_1$, $P8_1$, Oz_1, Cz_1, $F3t_3$, Fzt_1, $P4_2$	90.32	81.58	68.18	68.89
$P8t_1$, $P8_1$, Oz_1, Cz_1, $F3t_3$, Fzt_1, $P3t_1$	90.32	86.11	62.22	64.58
$P8t_1$, $P8_1$, Oz_1, $P8_2$, $F3t_3$, Fzt_1, $P4_1$	90.32	83.78	71.79	73.81

The performances again increased from the selection of 5 best ensembles previously. The top performance of almost 92% is also close to the performance of Cohort B by itself. This high performance further enforces the ability to combine the two datasets.

CHAPTER 5

CONCLUSIONS

The primary application in this work was the creation of an automated means of Alzheimer's disease diagnosis from a cohort of AD patients and cognitively normal subjects. Two separate cohorts were used, each tested at different facilities by different technicians, with unique means of artifact rejection. From there, ERPs were averaged together and features were extracted from responses to target and novel stimuli using wavelet analysis to provide coefficients from three frequency bands, 1-2 Hz, 2-4 Hz, and 4-8 Hz. Automated classification was performed using MLP and SVM neural networks to create classifiers, which were then combined using decision level data fusion to provide final classifications.

5.1 SUMMARY OF ACCOMPLISHMENTS

Previous work focused on only specific electrodes rather than a full exhaustive analysis of all data available. The main focus was the parietal region, although the combination scheme was still not completely exhaustive. In this portion of the study, all available data was examined from the previous cohort as well as an additional cohort. Ensembles, therefore, contained data from 19 electrodes for Cohort A and 16 electrodes from Cohort B.

Trends from previous studies were still examined including novel responses performing better overall than their target response counterparts. The parietal region was

still the most informative region of the head when compared with the other regions in both cohorts. There is, however, significant contribution from other regions in the head that were not able to be seen in previous work.

The metric used for comparison was the diagnostic accuracy of local physicians. This figure of 75% diagnostic accuracy would serve as the very minimum performance the algorithm must achieve to accept its validity. The algorithm presented here yielded performances of 93.6% for Cohort B and 88.7% for Cohort A using SVMs and 84.7% for Cohort B and 82.2% for Cohort A using MLPs, all of which significantly exceed the goal of 75%.

In almost every case the SVMs performed better than the MLPs using the same algorithm with the neural network being the only variable. In addition, the nature of the SVM allows for identical performance every time it is run, unlike the MLP, which uses random initialization every time and provides slightly different performances. The SVM also completes its training in a fraction of the time it takes to complete with the MLP, which allows for faster, more accurate diagnoses. Based on this information, future work on this project should focus primarily on the SVM as the classifier of choice.

All previous work in this study was done using only a leave-one-out training method. In this method, each subject used during the testing phase was tested by a classifier that had not previously seen that patient. This method allows for the most possible training data to be used to train the classifier before testing. This approach is completely valid when looking at a single classifier. A question arose when this method was still used for ensemble selection. Since the electrodes used in the ensemble were selected based on their single classifier performances, the test performance of the

ensemble is not entirely based on previously unseen data. This dilemma is despite the fact that the ensemble testing was still done in a leave-one-out fashion where the particular test patient was still unseen by the all the classifiers in the ensemble.

A second evaluation structure was created to test the hypothesis that the ensemble selection was being affected by including all patients in the single classifier performance numbers. A 5-fold process was used where by 20% of the total cohort was excluded from initial training and electrode selection and only used for testing by the ensemble. When comparing performance numbers of each algorithm, there is very little difference. The MLP version using Cohort A provided 82.2% using leave-one-out and 83.7% using the 20% test data. The SVM version using Cohort A provide 88.73% using leave-one-out and 89.7% using 20% test data. The SVM version using Cohort B provided 93.6% using leave-one-out and 94.1% using 20% test data. Even more important are the scalp histograms. When comparing the electrode selections from leave-one-out versus 20% test using the same neural network, cohort, and stimuli, the histograms are very similar. The only differences between the two are the number of trials used, although the concentrations are all in the same electrode areas.

This experiment shows that the 20% test method selected almost identical electrode locations to the leave-one-out method, which means that the leave-one-out selection is completely valid and the 20% method adds additional computation, which is unneeded. Secondly, and more importantly, this shows that there are specific electrodes, which constantly provide better performance than others regardless of the training method used for selection.

This second observation is important in selecting the "best" ensemble to use for the algorithm. While, the performance numbers are important in addressing how affective the algorithm is in diagnosing between AD and normal, it's the regions of the head where the most informative signals are being generated that is of the most value. This analysis gives a better representation of where the discerning information is being generated from within the brain, which could lead to a better selection of recording sites for a future implementation and an idea of specifically which electrodes to choose for ensemble creation.

The following two Figures show all the scalp histograms side by side from each stimulus. Figure 5.1 shows the target histograms and Figure 5.2 shows the novel histograms. In each figure the histograms are arranged as follows: top left is Cohort A MLPs, bottom left is Cohort A SVMs, top right is Cohort B MLPs, and bottom right is Cohort B SVMs.

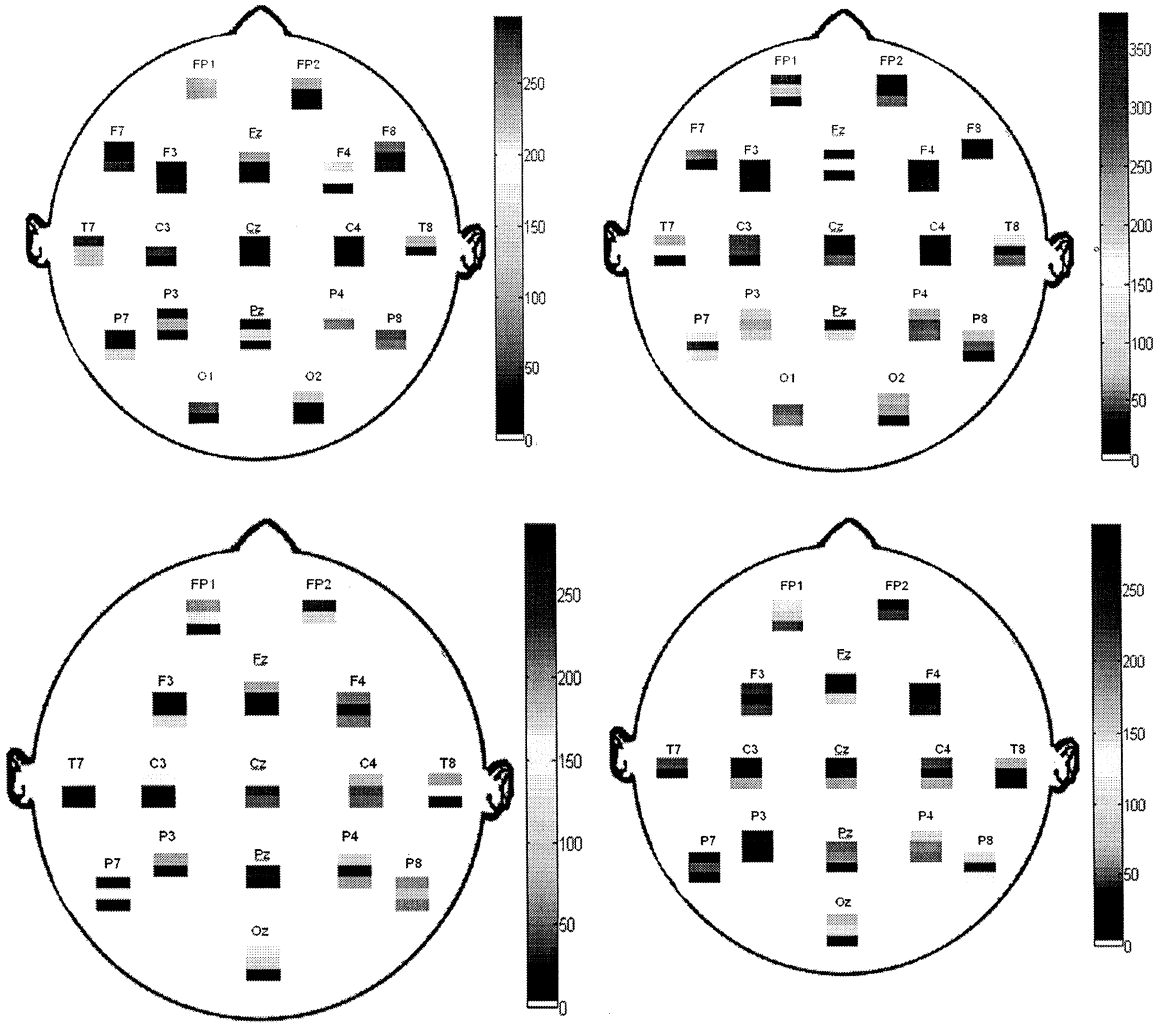

Figure 5.1: Target scalp histograms from both cohorts.

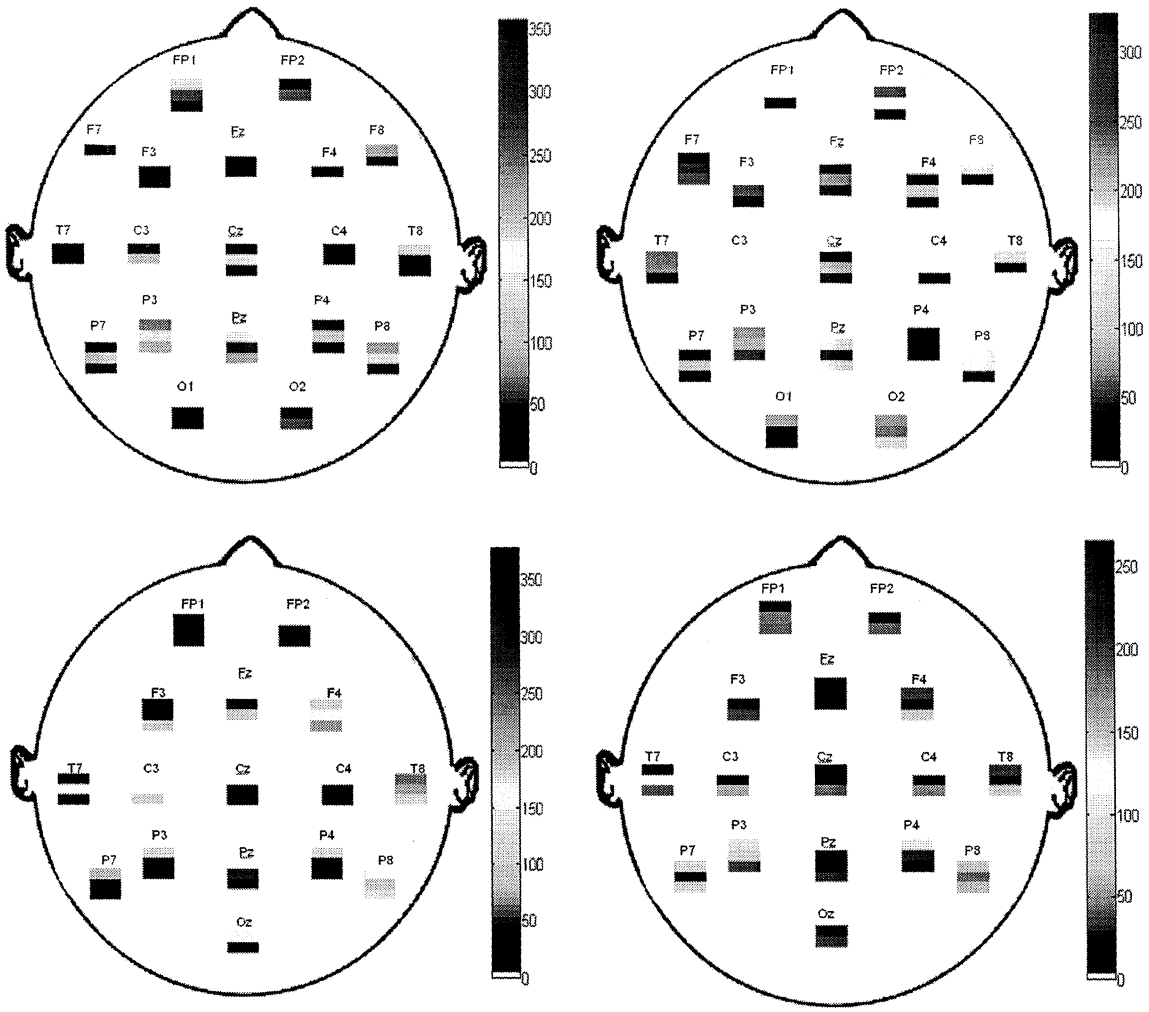

Figure 5.2: Novel scalp histograms from both cohorts.

There are definitely differences in the plots, specifically when comparing cohorts. It's clear that the electrodes in the parietal and occipital regions are consistently appearing along with the frontal most electrodes, FP1 and FP2. The rest of the electrodes seem to be more cohort dependant. This trend was also very evident when the two cohorts were combined. Again, the parietal electrodes were the best performing and occurred most frequently in the ensembles. In section 4.11.2, performances as high as 91.9% were achieved. The most frequently appearing electrodes were again drawn from the parietal and occipital regions. Based on all this information, when selecting

electrodes to combine together to diagnose AD, the parietal and occipital regions should be the focus for diagnosis with the possible inclusion of the FP electrodes. The other electrodes are too sporadic to be considered reliable, despite showing signs of providing some complementary information. This information comes as expected since the P300 and the majority of memory activity originate in the parietal region. This trend is further emphasized in Figures 5.3 and 5.4, which show the most frequently appearing electrodes for target and novel stimuli, respectively, with no relation to frequency band.

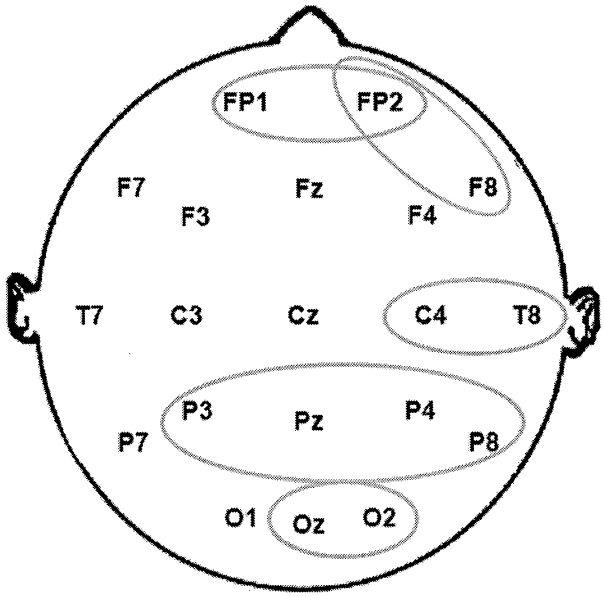

Figure 5.3: Most frequently appearing electrodes in response to target stimuli with no relation to frequency bands.

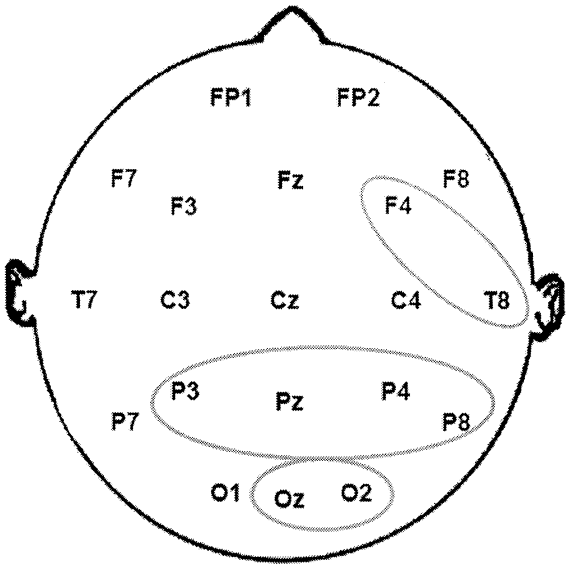

Figure 5.4: Most frequently appearing electrodes in response to novel stimuli with no relation to frequency bands.

This portion of the study has shown significant improvement in classification accuracy over previous versions of the algorithm. Stepenosky, in 2006, showed a maximum average performance for the entire 71 patient Cohort A of 83.1% using MLPs. The version of the algorithm presented in this work shows an increase of over 5% to 88.7% using SVMs. This increase is both statistically and clinically significant. The algorithm was also able to achieve a classification accuracy of 93.6% for Cohort B. Since this is the first application of the algorithm to this second cohort, that number will serve as the benchmark for future versions of the algorithm.

5.2 SOURCES OF ERROR

The primary source of error in this study comes from the diagnostic process of the patients included in the study. The classes used for each subject were derived from clinical evaluation, which as stated earlier is at best 90% accurate when performed by an expert neurologist. Since there is no way to improve upon this accuracy without

postmortem analysis, that 90% accuracy is the 'gold standard' for diagnosis of the subjects. As a result, the diagnoses made for all subjects are assumed to be 100% accurate for the testing within this study. This assumption could lead to inherent error if the original diagnosis was in fact incorrect. Inclusion of autopsy findings at a later date could potentially increase the performance and validity of all results.

Other errors could lie within the classification algorithms as well as the testing procedures. The classifiers themselves could be a source of error if the most ideal parameters were not selected. In addition, the MLP allows for more error since the initialization is random each time it is trained. The error goal was also set to 0.01 for all trials, however, that error goal was seldom reached.

As far as the testing procedures, the two different cohorts were tested at different facilities using different equipment. These differences could lead to variations in the data between the two cohorts. The data from Cohort B was also cleaned of artifacts by an automated procedure rather than by an expert electroencephalographer, which could lead to data instances either being removed that shouldn't be removed or instances remaining in the data that should have been removed. An analysis of the automated removal records by an expert could help to alleviate some this error.

5.3 RECOMMENDATIONS FOR FUTURE WORK

At this point, the ERP analysis between AD and normal subjects has been exhausted for the most part. While more subjects are still currently being recruited for Cohort B, analysis should be shifted to other areas of the study. In addition to the AD and normal subjects, Cohort B also includes Parkinson's disease patients and MCI patients. As the

size of these cohorts is increased, focus should be directed towards 3 and 4-class classification algorithms based on the results of the 2-class algorithm. Cohort B data acquisition protocols also include other biomarkers besides the EEG. Analysis of these other biomarkers including MRI, CSF, and PET scan data is critical to developing the most accurate and complete algorithm possible.

REFERENCES

[1] National Center for Health Statistics, "Health, United States, 2006 with chartbook on trends in the health of Americans," Hyattsville, MD: 2006

[2] American Health Assistance Foundation, "The facts on Alzheimer's disease," [online], 2007: <http://www.ahaf.org/SubIndex/AD_PDF_FactSheets/AD_stats.pdf>, Last accessed: November, 2007.

[3] Parkinson's Disease Society, "How many people have Parkinson's?," [online], 2007: <http://www.parkinsons.org.uk/about-parkinsons/what-is-parkinsons/how-many-people-have-parkinson.aspx>, Last accessed: November, 2007.

[4] Alzheimer's Association, "Alzheimer's disease facts and figures, 2007," [online], 2007: <http://www.alz.org/national/documents/Report_2007FactsAndFigures.pdf>, Last accessed: November, 2007.

[5] Wimo, A., *et al.*, "An estimate of the total worldwide societal costs of dementia in 2005," *Alzheimer's and Dementia*, 2007, Vol. 3, Issue 2, pp. 81-91.

[6] GPO Access, "Department of Education," [online], 2006: <http://www.gpoaccess.gov/usbudget/fy06/pdf/budget/education.pdf>, Last accessed: November, 2007.

[7] Alzheimer's Association, "Basics of Alzheimer's disease," [online], 2007: <http://www.alz.org/documents/national/basics_of_alz_low.pdf>, Last accessed: November, 2007.

[8] Help-Guide, "Alzheimer's diagnosis and risk factors," [online], 2007: <http://www.helpguide.org/elder/alzheimers_diagnosis_risk_factors.htm>, Last accessed: November, 2007.

[9] Lim, A., *et al.*, "Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series," *Journal of the American Geriatrics Society*, 1999, Vol. 47. Issue 5, pp. 564-569.

[10] Shenk, D., *The Forgetting, Alzheimer's: Portrait of an Epidemic.* Doubleday. New York: 2001.

[11] Small, D. H., "Recent findings on the biology of Alzheimer's disease," *Research and Practice in Alzheimer's disease*, Vol 3., Springer Publishing Company. New York: 2000. pp.27-32.

[12] National Institute on Aging, "Alzheimer's disease: Unraveling the mystery," [online], 2006: <http://www.nia.nih.gov/NR/rdonlyres/A294D332-71A2-4866-BDD7-A0DF216DAAA4/0/Alzheimers_Disease_Unraveling_the_Mystery.pdf >, Last Accessed: November, 2007.

[13] Soukup, J. E., *Alzheimer's disease: A Guide to Diagnosis, Treatment, and Management*. Praeger. Westport, Connecticut: 1996.

[14] Hansson, O., *et al.*, "Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow up study," *The Lancet Neurology*, 2006, Vol. 5, Issue 3, pp. 228-234.

[15] Bouwman, F. H., *et al.*, "CSF biomarkers and medial temporal lobe atrophy predict dementia in mild cognitive impairment," *Neurobiology of Aging*, Vol. 28, Issue 7, 2007, pp. 1070-1074.

[16] Schoonenboom, N.S.M., "CSF and MRI markers independently contribute to the diagnosis of Alzheimer's disease," *Neurobiology of Aging*, In Press, 2007.

[17] De Leon, M.J., Klunk, W., "Biomarkers for the early diagnosis of Alzheimer's disease," *The Lancet Neurology*, Vol. 5, Issue 3, 2006, pp. 198-199.

[18] Mosconi, L., *et al,*. "Early detection of Alzheimer's disease using neuroimaging," *Experimental Gerontology*, 2007, Vol. 42, Issue 1-2, pp. 129-138.

[19] Bucher, L.A., "Potential of functional MRI for evaluation of memory loss: successes and pitfalls," *Research and Practice in Alzheimer's disease*, Vol 3., Springer Publishing Company. New York: 2000. pp. 111-118.

[20] Ito, K., "PET/SPECT for dementia—early diagnosis of Alzheimer's disease," *International Congress Series*, 2006, Vol. 1290, pp. 123-127.

[21] Zimny, A., "Does perfusion CT enable differentiating Alzheimer's disease from vascular dementia and mixed dementia? A preliminary report," *Journal of the Neurological Sciences*, 2007, Vol. 257, Issues 1-2, pp. 114-120.

[22] Fritzsche, K., *et al.*, "Automated MRI-based quantification of the cerebral atrophy providing diagnostic information on mild cognitive impairment and Alzheimer's disease," *Proceedings of IEEE Symposium on Computer-Based Medical Systems*, 2006.

[23] Gianotti, L.R., *et al.,* "Correlation between disease severity and brain electric LORETA tomography in Alzheimer's disease," *Clinical Neurophysiology*, 2007, Vol. 118, Issue 1, pp. 186-196.

[24] Griebe, M., *et al.,* "Infrared spectroscopy: A new diagnostic tool in Alzheimer's disease," *Neuroscience Letters*, 2007, Vol. 420, Issue 1, pp. 29-33.

[25] Arai, H., *et al.*, "A quantitative near-infrared spectroscopy study: A decrease in cerebra hemoglobin oxygenation in Alzheimer's disease and mild cognitive impairment," *Brain and Cognition*, 2006, Vol. 61, Issue 2, pp 189-194.

[26] Allsop, D., *et al.*, "Fluorescence anisotropy: A method for early detection of Alzheimer β-peptide (Aβ) aggregation", *Biochemical and Biophysical Research Communications*, 2001, Vol. 285, Issue 1, pp. 58-63.

[27] Marcotte, E., *et al.*, "cDNA microarray and proteomic approaches in the study of brain diseases: focus on schizophrenia and Alzheimer's disease,*'Pharmacology & Therapeutics*, 2003, Vol. 100, Issue 1, pp. 63-74.

[28] Sornmo, L., Laguna, P., *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Elsevier Academic Press, Boston: 2005.

[29] Berger, H., "Über das Elektroenkephalogramm des Menschen,*" Archiv für Psychiatrie und Nervenkrankheiten*, 1931; 94:16-60.

[30] Lopes de SilvaF., *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 4th Ed., Lippincott, Williams and Wilkins, Philadelphia, 1999.

[31] Malmivue, J. and Plonsey, R., "Bioelectromagnetism – principles and applications of bioelectric and biomagnetic fields," *Oxford University Press, New York,* 1995; 13. Available online at <http://butler.cc.tut.fi/~malmivue/bem/bembook/13/13.htm>, Last accessed: November, 2007.

[32] Epilepsy Professionals, "Implanted EEG electrodes," [online], 2007: < http://professionals.epilepsy.com/page/surgery_electrodes.html>, Last Accessed: November, 2007.

[33] Crossroads Institute, "Brainwaves and EEG," [online], 2007, <http://www.crossroadsinstitute.org/eeg.html>, Last Accessed: November, 2007.

[34] Touge, T., *et al.*, "The interaction of somatosensory and auditory stimuli on event-related potentials," *International Congress Series*, 2005, Vol. 1278, pp. 149-152.

[35] Murali, S., Kulish, V., "Modeling of evoked potentials of electroencephalograms: An overview," 2007, *Digital Signal Processing*, Vol 17, Issue 3, pp. 665-674.

[36] Polich, J., Margala, C., "P300 and probability: comparison of oddball and single-stimulus paradigms," *International Journal of Psychophysiology*, 1997, Vol. 25, Issue 2, pp. 169-176.

[37] Yamaguchi, S., *et al.*, "Event-related brain potentials in response to novel sounds in dementia," *Clinical Neurophysiology*, 2000, Vol. 111, Issue 2, pp. 195-203.

[38] Cohen, J., Polich, J., "On the number of trials needed for P300," *International Journal of Psychophysiology*, 1997, Vol. 25, Issue 3, pp. 249-255.

[39] Polich, J., Kok, A., "Cognitive and biological determinants of P300: an integrative review," *Biological Psychology*, 1995, Vol 41, Issue 2, pp. 103-146.

[40] Bennington, J. Y., Polich, J., "Comparison of P300 from passive and active tasks for auditory and visual stimuli," *International Journal of Psychophysiology*, 1999, Vol. 34, Issue 2, pp. 171-177.

[41] Strüber, D., Polich, J., "P300 and slow wave from oddball and single-stimulus visual tasks: inter-stimulus interval effects," *International Journal of Psychophysiology,* 2002, Vol. 45, Issue 3, pp. 187-196.

[42] Polich, J., 'Updating P300: An integrative theory of P3a and P3b," *Clinical Neurophysiology*, 2007, doi:10.1016/j.clinph.2007.06.019.

[43] Van Vugt, M., *et al.*, "Comparison of spectral analysis methods for characterizing brain oscillations," *Journal of Neuroscience Methods*, 2007, Vol. 162, Issues 1-2, pp. 49-63.

[44] Başar-Eroglu, C., *et al.*, "Topological distribution of oddball 'P300' responses," *International Journal of Psychophysiology*, 2001, Vol. 29, Issues 2-3, pp. 213-220.

[45] Demiralp, T., *et al.*, "Time-frequency analysis of single-sweep event-related potentials by means of fast wavelet transform," 1999, Vol. 66, Issue 1, pp. 129-145.

[46] Demiralp, T., *et al.*, "Detection of P300 waves in single trials by the wavelet transform (WT)," 1999, Vol 66, Issue 1, pp. 108-128.

[47] Başar, E., *et al.*, "Event-related oscillations are 'real brain responses' — wavelet analysis and new strategies," *International Journal of Psychophysiology*, 2001, Vol. 39, Issues 2-3, pp. 91-127.

[48] Demiralp, T., *et al.*, "Wavelet analysis of oddball P300," *International Journal of Psychophysiology*, 2001, Vol. 39, Issues 2-3, pp. 221-227.

[49] Demiralp, T., *et al.*, "Wavelet analysis of P3a and P3b," *Brain Topography,* 2001, Vol. 13, Issue 4, pp. 251-267.

[50] Jeong, J., "EEG dynamics in patients with Alzheimer's disease," *Clinical Neurophysiology*, 2004, Vol. 155, Issue 7, pp. 1490-1505.

[51] Olichney, J., Hillert, D., "Clinical applications of cognitive event-related potentials in Alzheimer's disease," *Physical Medicine and Rehabilitation Clinics of North America*, 2004, Vol. 15, Issue 1, pp. 205-233.

[52] Goodin, D. S., Aminoff, M.J., "Electrophysiological differences between subtypes of dementia," *Brain*, 1986, Vol. 109, pp. 1103-1113.

[53] Ford, J.M., Pfefferbaum, A., "Age-related changes in event-related potentials," *Advanced Psychophysiology*, 1985, Issue 1, pp. 301-339.

[54] Squires, K.C., *et al,* "Electrophyiological assessment of mental function in aging and dementia," *Aging in the 1980s*, Washington, DC: American Psychology Association, 1980, pp. 125-134.

[55] Gordon, E., *et al.*, "The differential diagnosis of dementia using P300 latency," *Biological Psychiatry*, 1986, Vol. 21, pp. 1123-1132.

[56] Ball, S.S., *et al.* "Longitudinal P300 latency changes in Alzheimer's disease," *Journal of Gerontology and Medical Sciences*, 1989, Vol. 44, pp. 195-200.

[57] Pfefferbaum, A., *et al.,* "Clinical application of the P3 component of even-related potentials: II. Dementia, depression and schizophrenia," *Electroencephalography and Clinical Neurophysiology*, 1986, Vol. 59, pp. 104-124.

[58] Polich, J., *et al.*, "P300 assessment of early Alzheimer's disease," *Electroencephalography and Clinical Neurophysiology,* 1990, Vol. 77, pp.179-189.

[59] Golob, E.J., Starr, A., "Effects of stimulus sequence on event-related potentials and reaction time during target detection in Alzheimer's disease," *Clinical Neurophysiology*, 2000, Vol. 111, pp. 1438-1439.

[60] Hidasi, Z., *et al.*, "Changes of EEG spectra and coherence following performance in a cognitive task in Alzheimer's disease," *International Journal of Psychophysiology*, Vol. 65, pp. 252-260.

[61] Saitoh, E., *et al.*, "Comparison of visual evoked potentials in patients with psychogenic visual disturbance and malingering," *Journal for Pediatrics, Opthalmology, and Strabismus*, 2001, Vol. 38, pp. 21-26.

[62] Tanaka, F., *et al.,* "Auditory and visual event-related potentials and flash visual evoked potentials in Alzheimer's disease: correlations with Mini-Mental State Examination and Raven's Coloured Progressive Matrices." *Journal of Neurological Sciences,* 1998, Vol. 156, pp. 83-88.

[63] Van der Hiele, K., *et al.,* "EEG and MRI correlates of mild cognitive impairment and Alzheimer's disease," *Neurobiology of Agining*, 2007, Vol. 28, pp. 1322-1329.

[64] Szelies, B., *et al.*, "P300 in Alzheimer's disease: relationships to dementia severity and glucose metabolism," *Journal of the Neurological Sciences*, 1995, Vol. 130, pp. 77-81.

[65] Boutros, N., *et al.,* "Evoked potentials in subjects at risk for Alzheimer's disease," *Psychiatry Research,* 1995, Vol. 57, pp. 57-63.

[66] Ally, B., *et al.*, "The P300 component in patients with Alzheimer's disease and their biological children," *Biological Psychology*, 2006, Vol. 72, pp. 180-187.

[67] Polikar, R., *et al.*, "Multiresolution wavelet analysis of ERPs for the detection of Alzheimer's disease," *Proceedings IEEE EMBS, 1997.*

[68] Petrosian, A., *et al.*, "Recurrent neural network and wavelet transform based distinction between Alzheimer's and control EEG," *Proceedings of the First Joint BMES/EMBS Conference Serving Humanity, Advancing Technology*," Atlanta, GA, 1999.

[69] Petrosian, A., *et al.*, "Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG," *Clinical Neurophysiology,* 2001, Vol. 112, pp. 378-387.

[70] Park, E.H., *et al.*, "Alzheimer's disease detection and analysis using P3 component of ERP in Alzheimer type dementia," *Proceedings of the 23rd Annual EMBS International Conference,* Istanbul, Turkey, 2001.

[71] Yagneswaran, S., *et al.*, "Power frequency and wavelet characteristics in differentiating between normal and Alzheimer EEG,' *Proceedings of the Second Joint EMBS/BMES Conference*, 2003.

[72] De Thad,C. H., "Signal processing techniques predict characteristic frequencies of the Alzheimer's disease beta-amyloid protein and its precursor," *Proceedings of the Second Joint EMBS/BMES Conference,* 2003.

[73] Cho, S.Y., *et al.*, "Automatic recognition of Alzheimer's disease with single channel EEG recording," *Proceedings of the 25th Annual International Conference of the IEEE EMBS*, Cancun, Mexico, 2003.

[74] Abasolo, D., *et al.*, "Electroencephalogram analysis with approximate entropy to help in the diagnosis of Alzheimer's disease," *Proceedings of the 4th Annual IEEE Conference on Information Technology Applications in Biomedicine*, 2003.

[75] Abasolo, D., *et al.*, "Approximate entropy and mutual information analysis of the electroencephalogram in Alzheimer's disease patients," *Advances in Medical, Signal and Information Processing*, 2006.

[76] Tao, H. and Tian, X., "Coherence characteristics of gamma-band EEG during rest and cognitive task in MCI and AD," *Proceedings of the 2005 IEEE EMBC,* 2005.

[77] Chapman, R., *et al.*, "Brain event-related potentials: Diagnosing early-stage Alzheimer's disease," *Neurobiology of Aging*, 2007, Vol. 28, pp. 194-201.

[78]Gomez, C., *et al.*, "Magnetoencephalogram background activity analysis in Alzheimer's disease patients using auto mutual information,'*Proceedings of the 28th IEEE EMBS*, New York City, 2006.

[79] Henderson, G, *et al.*, "Development and assessment of methods for detecting dementia using the human electroencephalogram," *IEEE Transactions on Biomedical Engineering*, 2006, Vol. 53, No. 8, pp. 1557-1568.

[80] Jacques, G., 'Daubechies and quadratic B-spline wavelets for automated early diagnosis of Alzheimer's disease,' *Thesis*: Department of Electrical and Computer Engineering – Rowan University, 2004.

[81] Topalis, A., 'Multiresolution wavelet analysis of event-related EEG potentials using ensemble of classifier data fusion techniques for the early detection of Alzheimer's disease,' *Thesis*: Department of Electrical and Computer Engineering – Rowan University, 2006.

[82] Stepenosky, N., 'Data fusion of complementary information from parietal and occipital event related potentials for early diagnosis of Alzheimer's disease,' *Thesis*: Department of Electrical and Computer Engineering – Rowan University, 2006.

[83] Gandhi, H., 'Stacked generalization for early diagnosis of Alzheimer's disease,' *Thesis*: Department of Electrical and Computer Engineering – Rowan University, 2007.

[84] McKhann, G., *et al.*, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease," *Neurology*, 1984, 34:939-944.

[85] Polikar, R., "The wavelet tutorial – The engineer's ultimate guide to wavelet analysis," [online], 2001:

<http://users.rowan.edu/polikar/WAVELETS/WTtutorial.html>, Last accessed: November, 2007.

[86] Tang, Y.Y., *et al*, *Wavelet Theory and Its Application to Pattern Recognition*. World Scientific. River Edge, NJ: 2000.

[87] Addison, P.S., *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine, and Finance*, Institute of Physics Publishing, Philadelphia, 2002.

[88] Croiser, A., Esteban, D., Galand, C., "Perfect channel splitting by use of interpolation/decimation/ tree decomposition techniques," *Int. Symp. Info, Circuits, Systems*, Patras Greece, 1976.

[89] Crochiere, R., Webber, S., Flanagan, J., "Digital coding of speech in sub-bands," *Proceedings of ICASSP,* 1976, Vol. 1, pp. 233-236.

[90] Burt, P., Adelson, E., "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, 1983, Vol. COM-31, No. 4.

[91] Vetterli, M., le Gall, D., "Perfect reconstruction filter banks: Some properties and factorizations," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 1989, Vol. 37, pp. 1057-1071.

[92] Goswami, J.C., Chan, A.K., *Fundamentals of Wavelets: Theory, Algorithms, and Applications*, John Wiley and Sons, Inc., New York: 1999.

[93] Walnut, David, F., *An Introduction to Wavelet Analysis*, Birkhauser, Boston: 2004.

[94] Haykin, S., *Neural Networks: A comprehensive Foundation*, Prentice Hall, Upper Saddle River, NJ: 1999.

[95] Burges, Christopher, J.C., "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, 1998, 2, 121-167.

[96] Polikar, Robi, *Advanced Topics in Pattern Recognition* class website: <http://users.rowan.edu/~polikar/CLASSES/ECE555>, Last accessed: November, 2007.

[97] Bishop, C., *Pattern Recognition and Machine Learning*, Spring, New York, 2006.

[98] Kuncheva, L., *Combining Pattern Classifiers: Methods and Algorithms*, John Wiley & Sons, Inc., Hoboken, NJ: 2004.

APPENDIX A

ERP GRAND AVERAGES FROM COHORT A

The graphs included here are the grand averages from each electrode for Cohort A. In each figure the top graph is the response to target stimuli and the bottom graph is the response to novel stimuli. Normal subjects and AD patients are shown on the same graph.

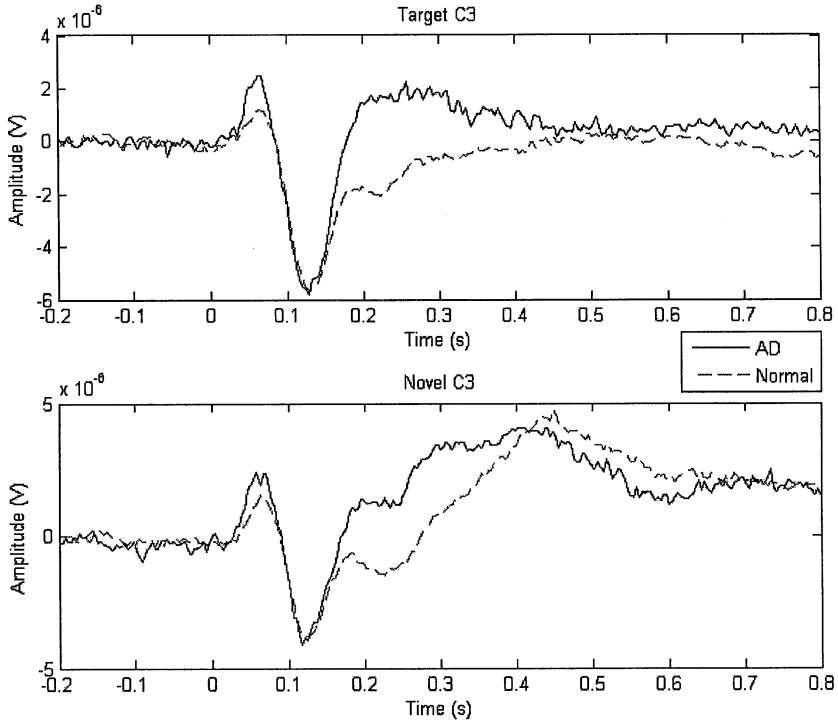

Figure A.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from Cohort A.

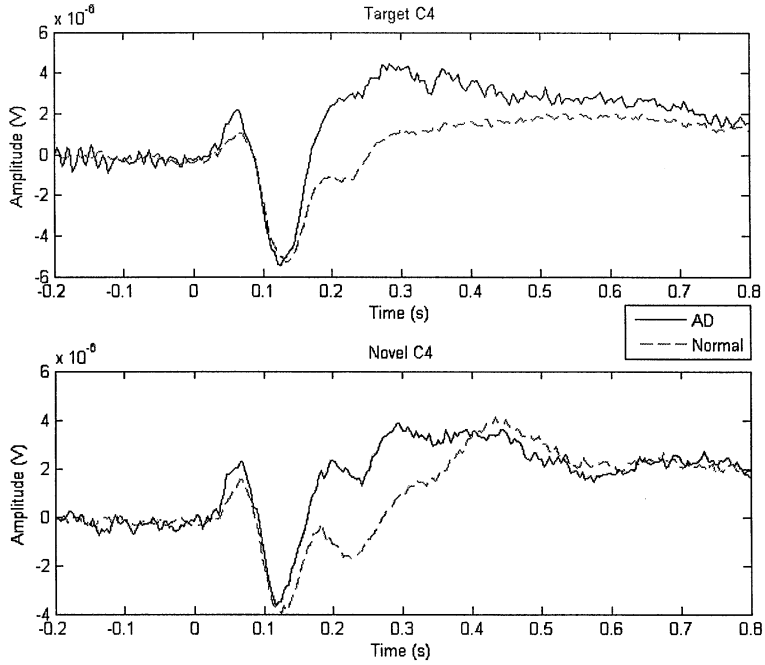

Figure A.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from Cohort A.

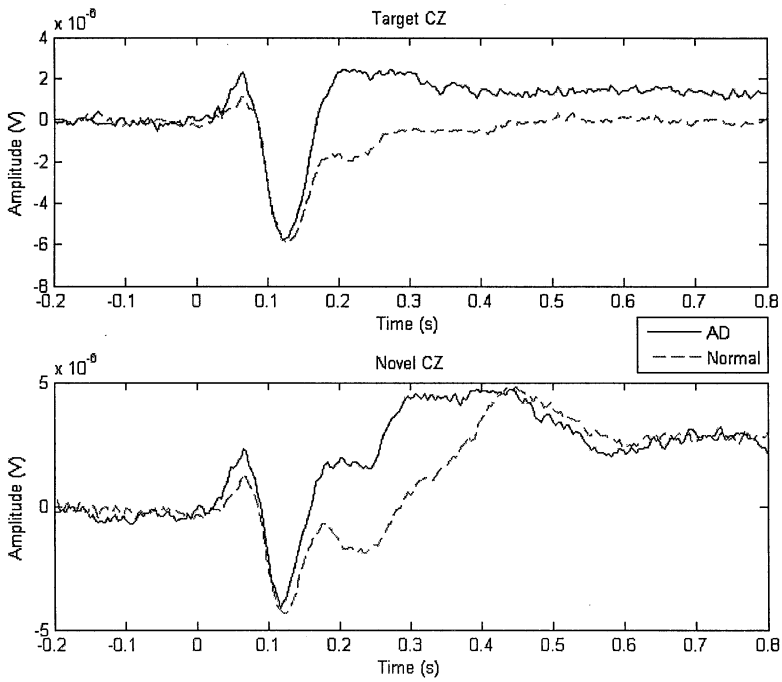

Figure A.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from Cohort A.

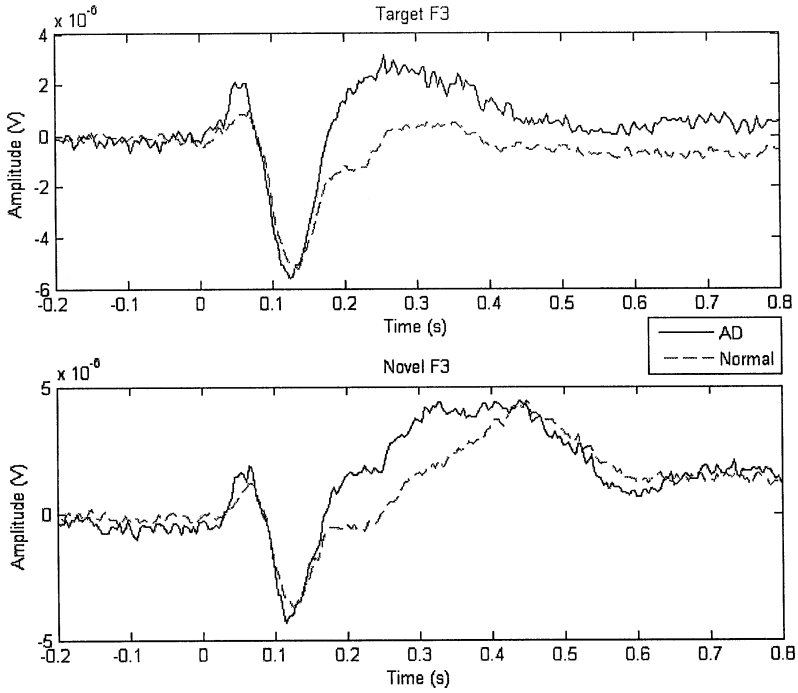

Figure A.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from Cohort A.

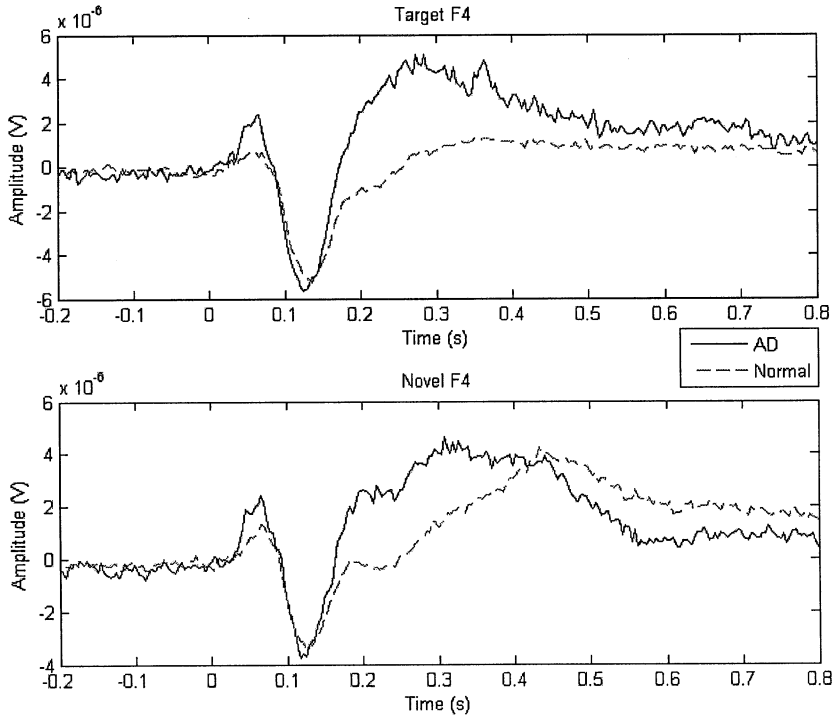

Figure A.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from Cohort A.

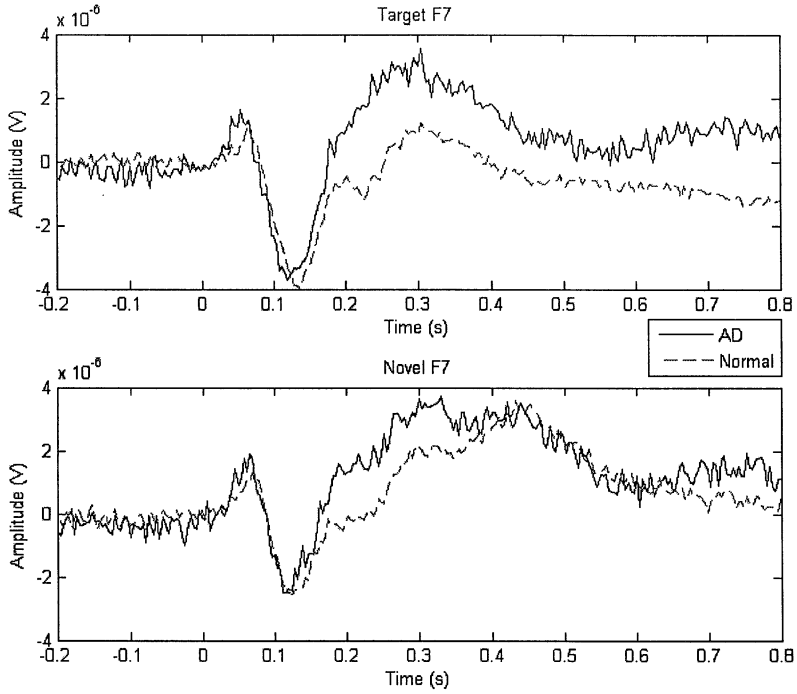

Figure A.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F7 electrode from Cohort A.

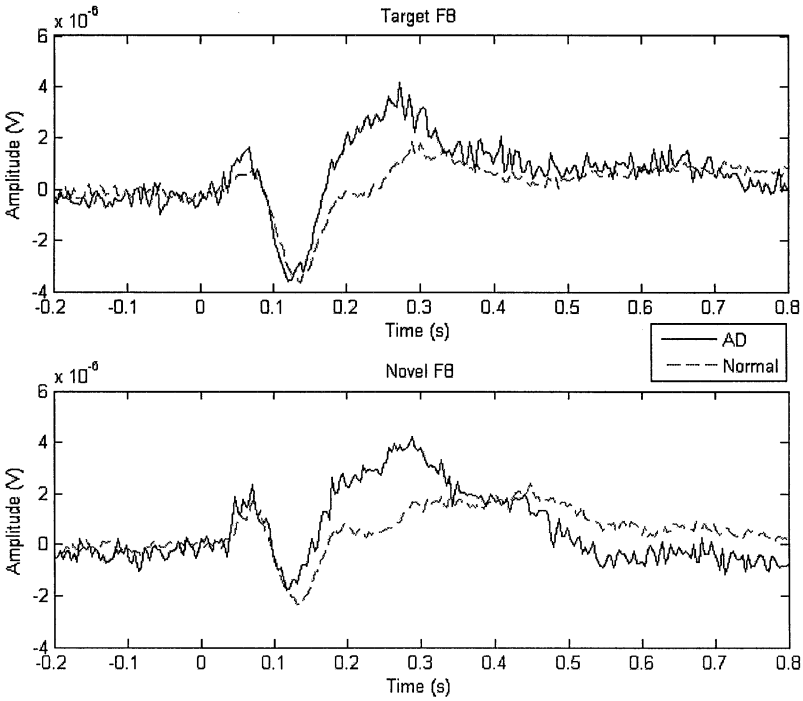

Figure A.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F8 electrode from Cohort A.

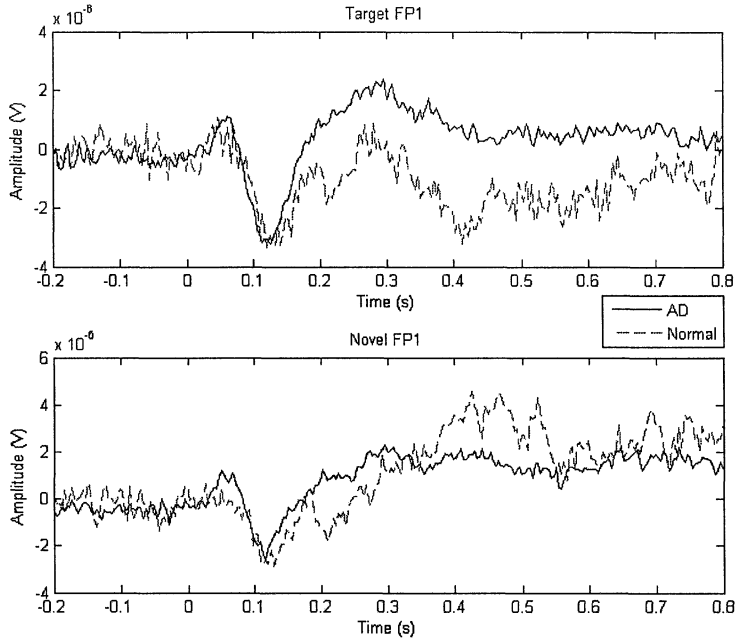

Figure A.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from Cohort A.

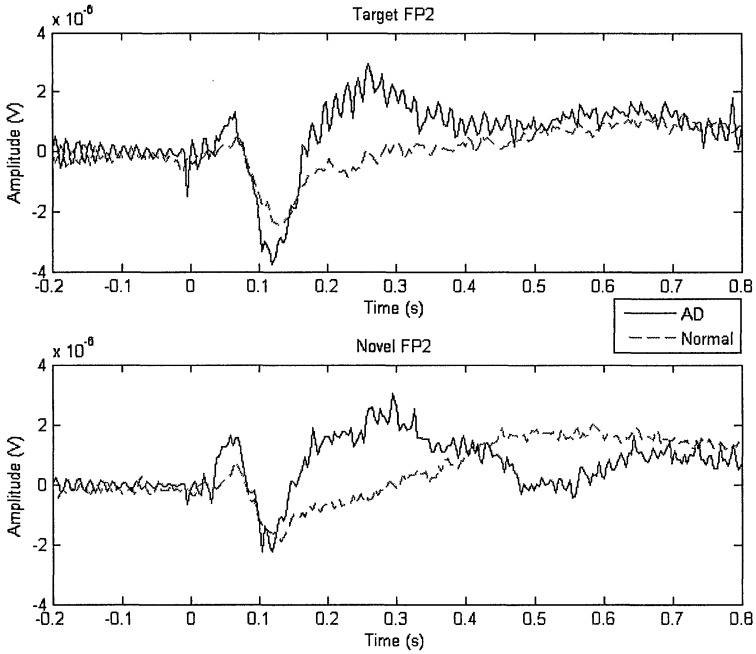

Figure A.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from Cohort A.

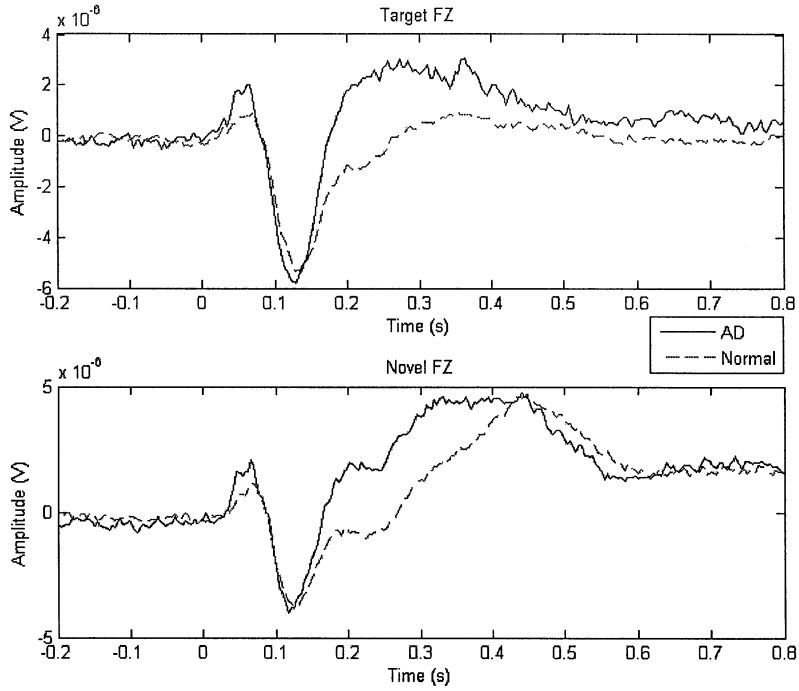

Figure A.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from Cohort A.

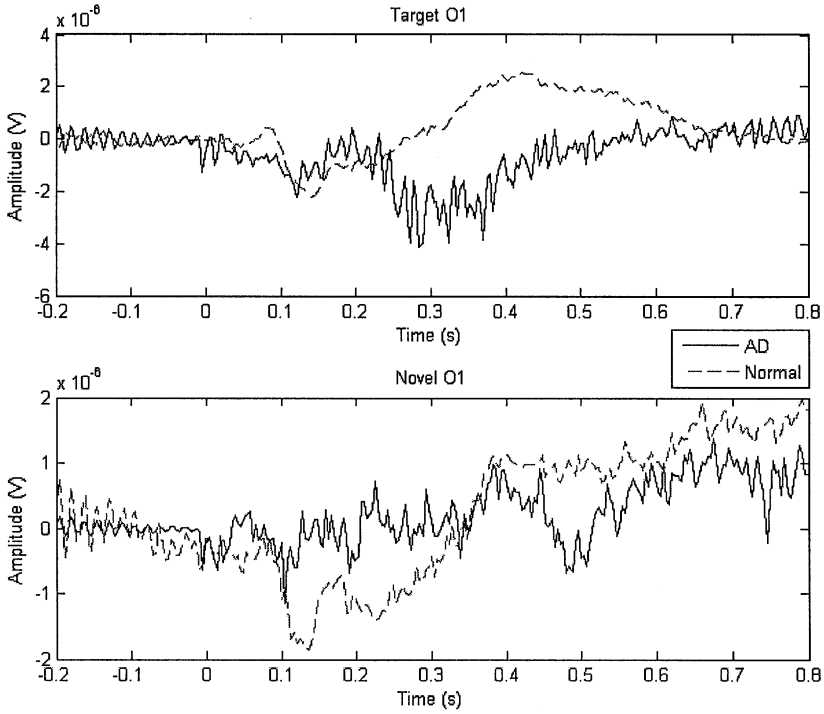

Figure A.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the O1 electrode from Cohort A.

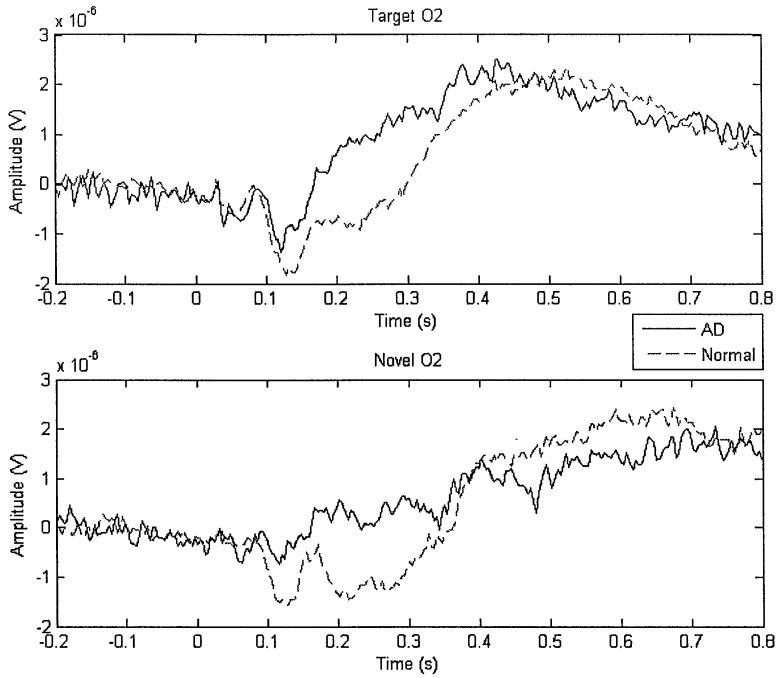

Figure A.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the O2 electrode from Cohort A.

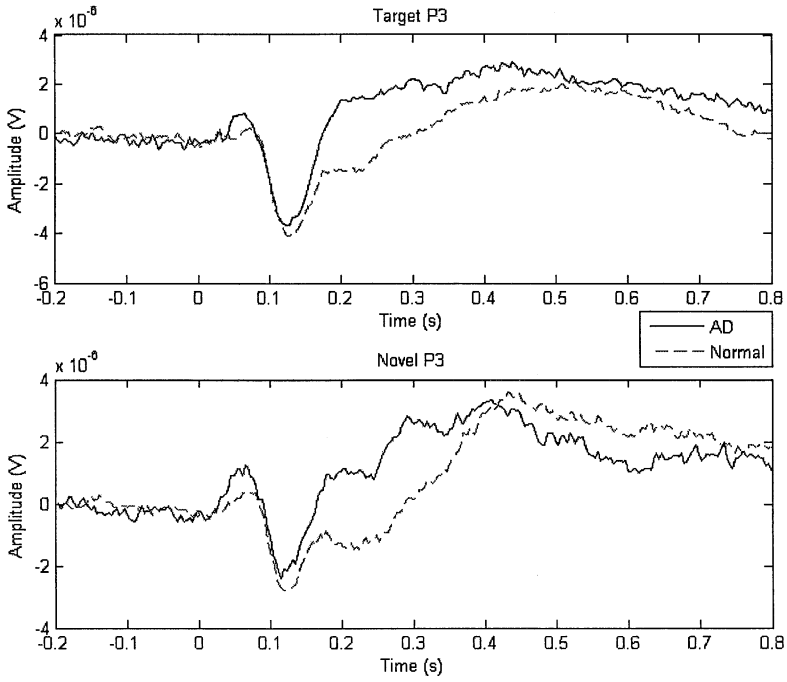

Figure A.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from Cohort A.

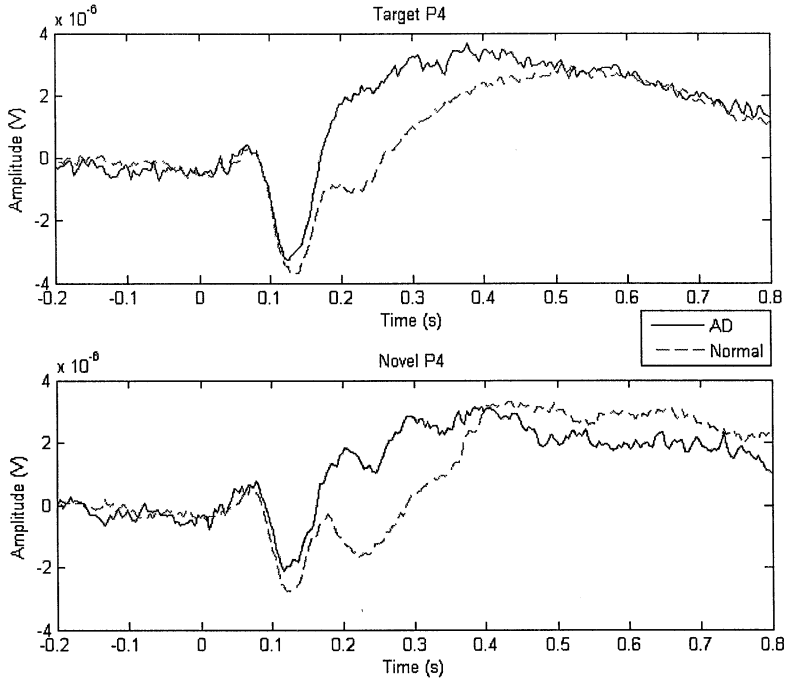

Figure A.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from Cohort A.

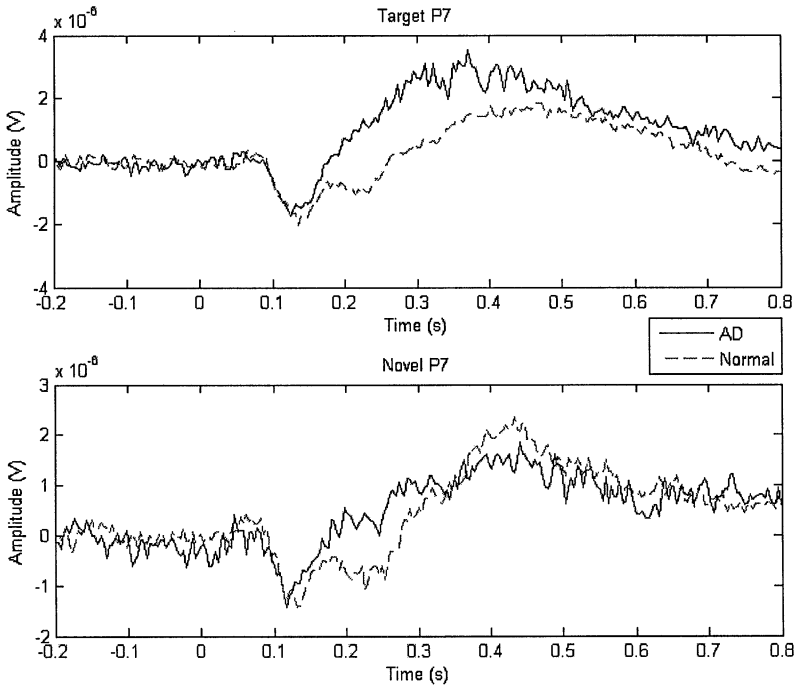

Figure A.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from Cohort A.

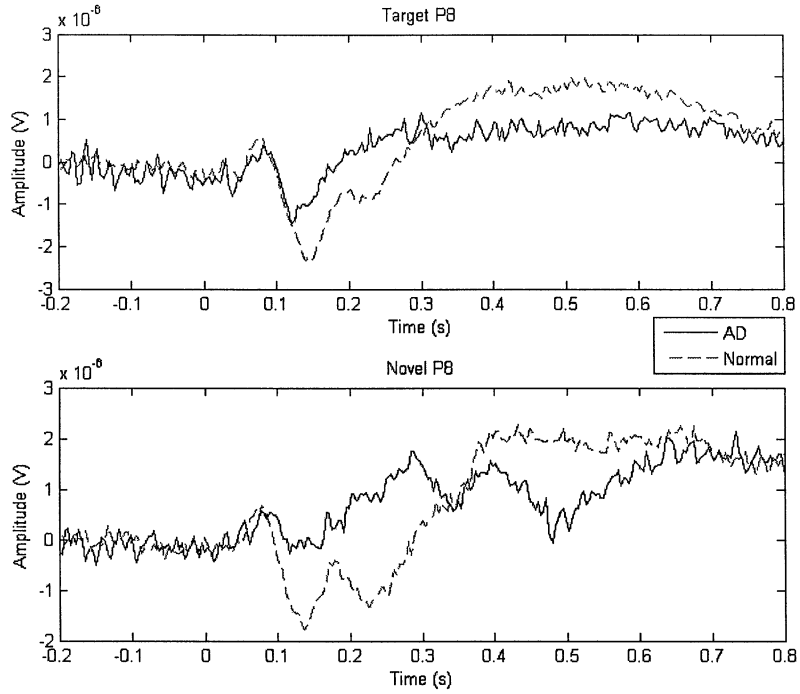

Figure A.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from Cohort A.

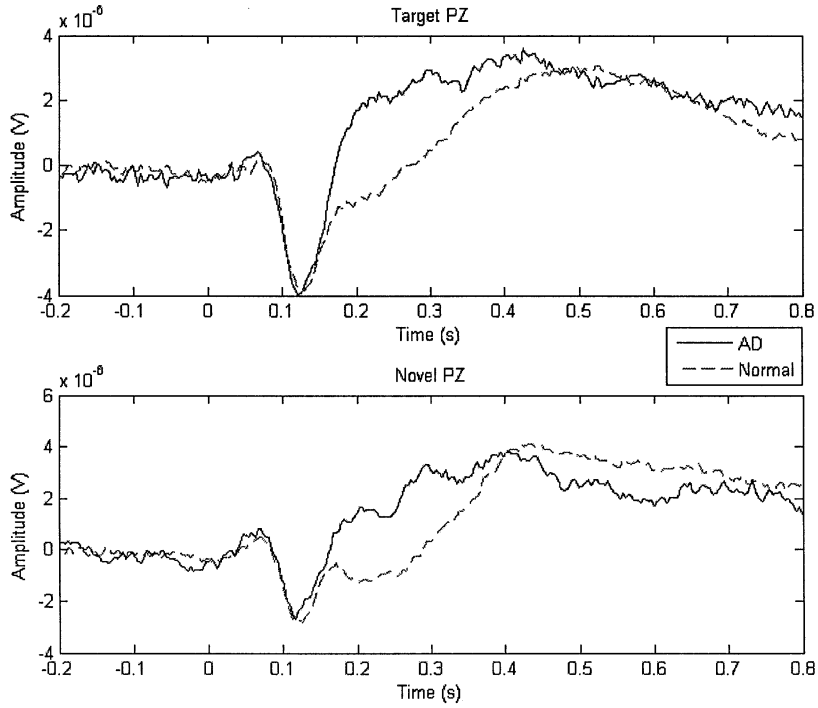

Figure A.17: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from Cohort A.

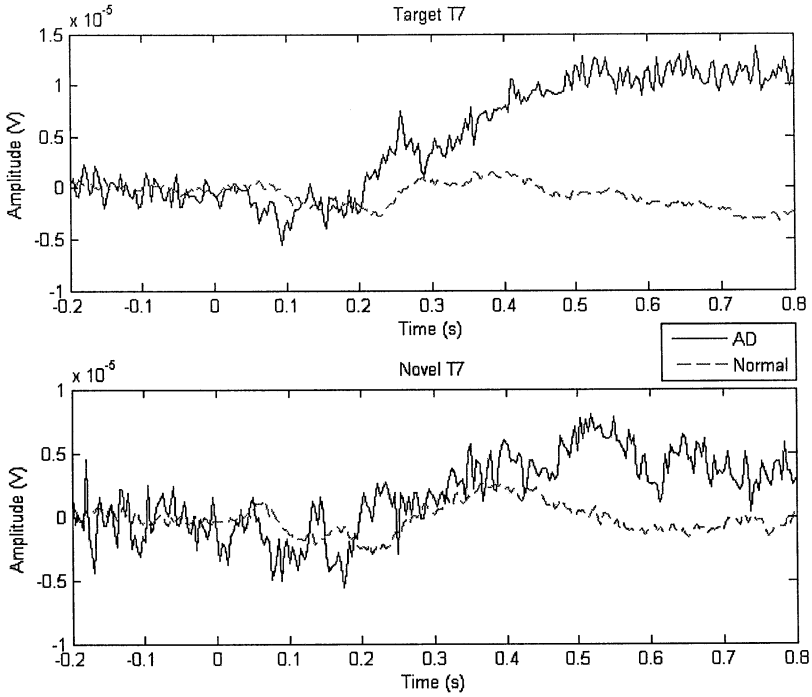

Figure A.18: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from Cohort A.

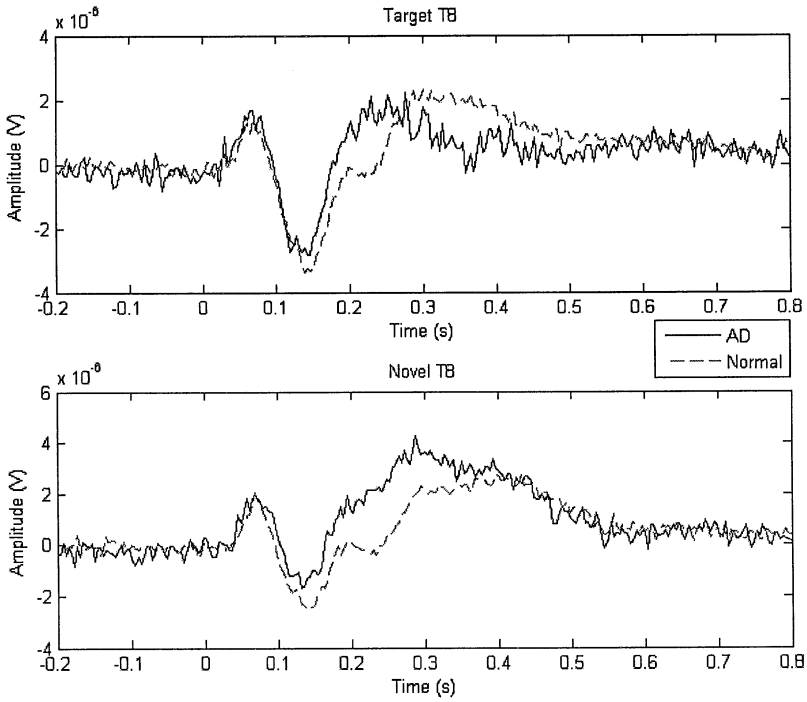

Figure A.19: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from Cohort A.

APPENDIX B

ERP GRAND AVERAGES FROM COHORT B

The graphs included here are the grand averages from each electrode for Cohort B. In each figure the top graph is the response to target stimuli and the bottom graph is the response to novel stimuli. Normal subjects and AD patients are shown on the same graph.

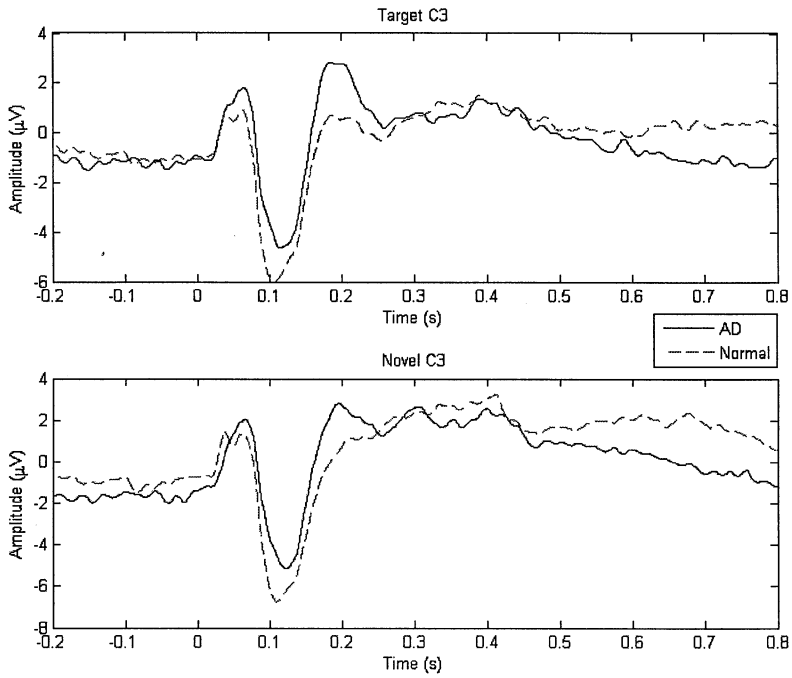

Figure B.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from Cohort B.

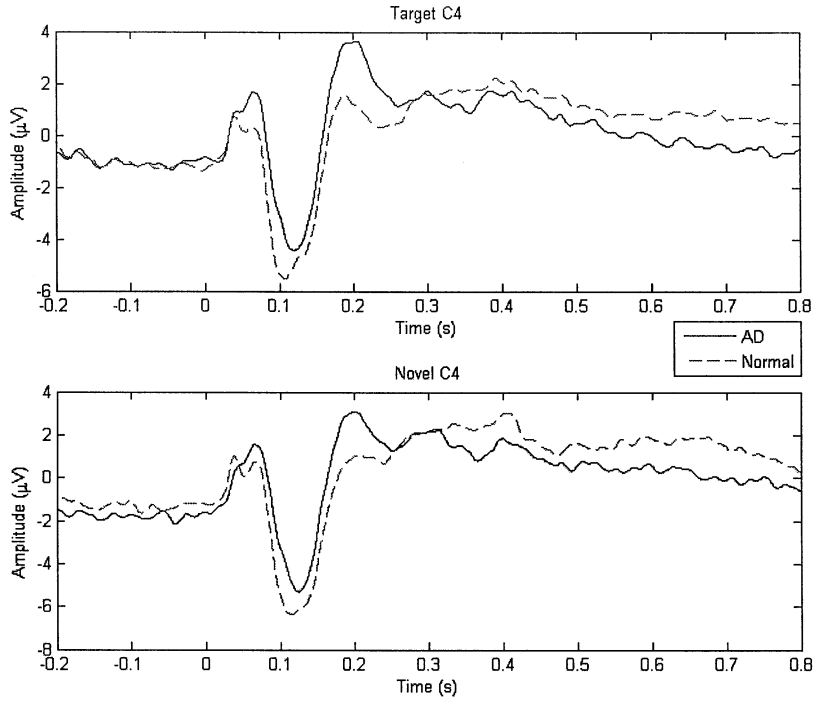

Figure B.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from Cohort B.

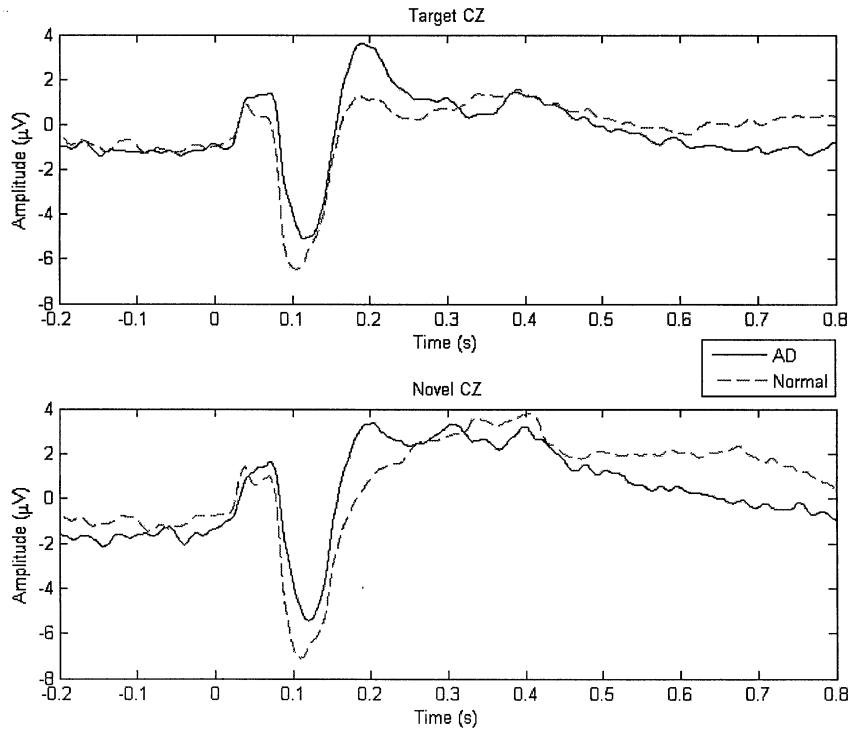

Figure B.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from Cohort B.

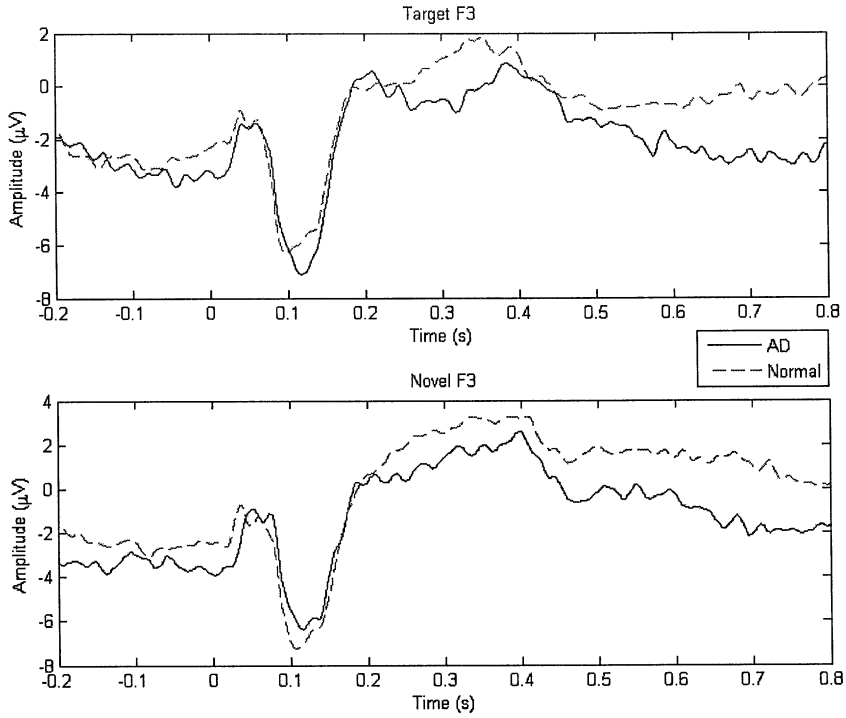

Figure B.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from Cohort B.

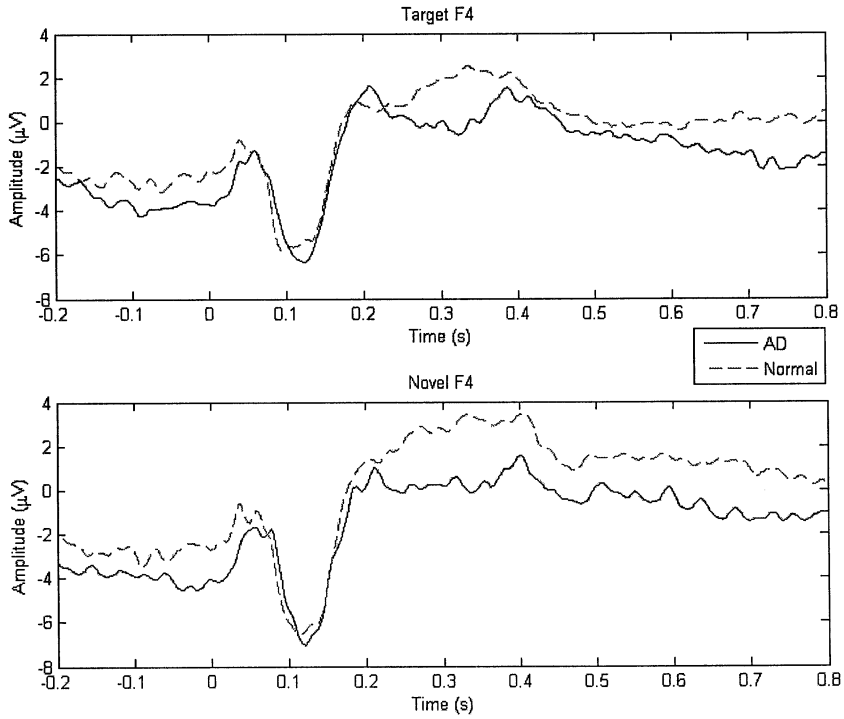

Figure B.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from Cohort B.

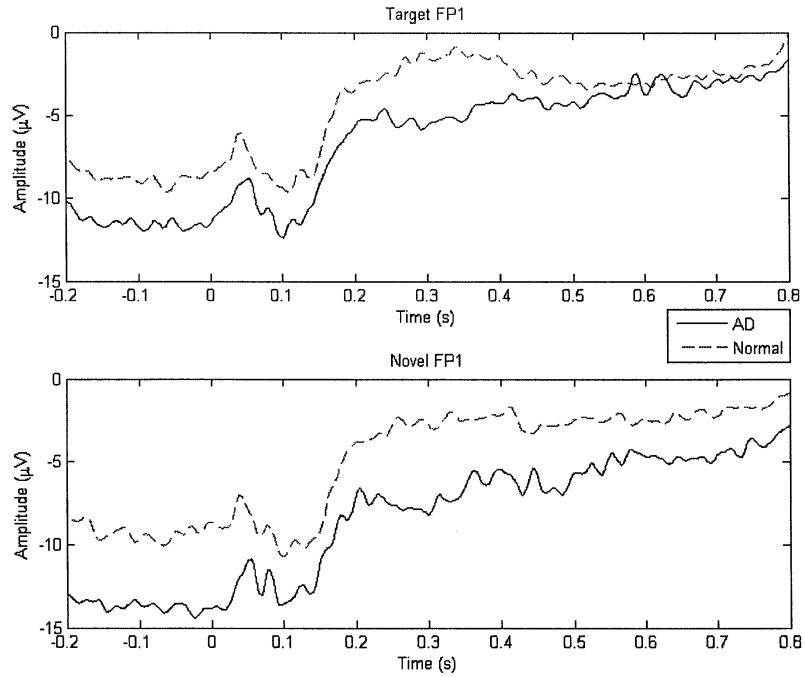

Figure B.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from Cohort B.

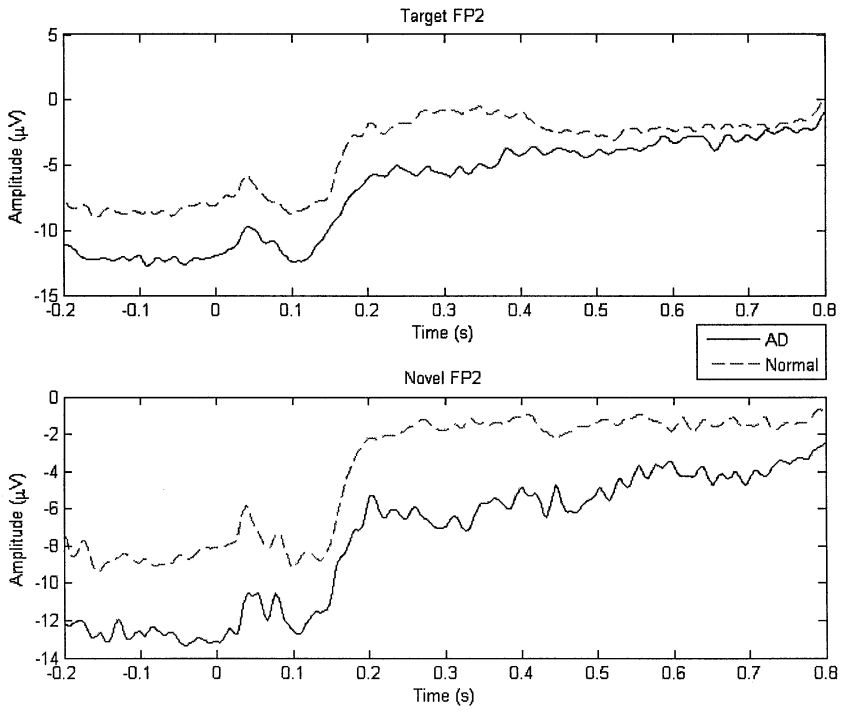

Figure B.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from Cohort B.

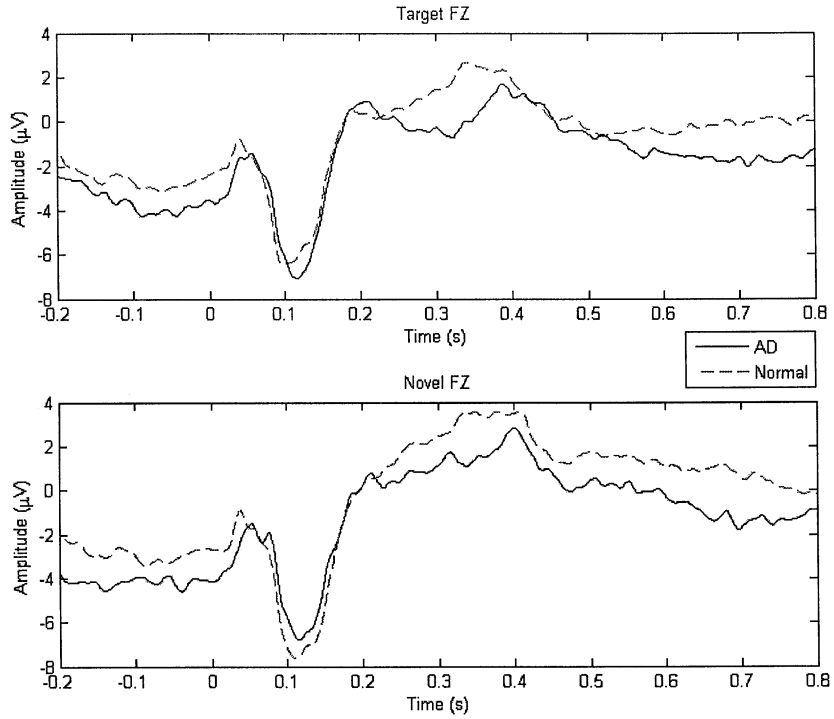

Figure B.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from Cohort B.

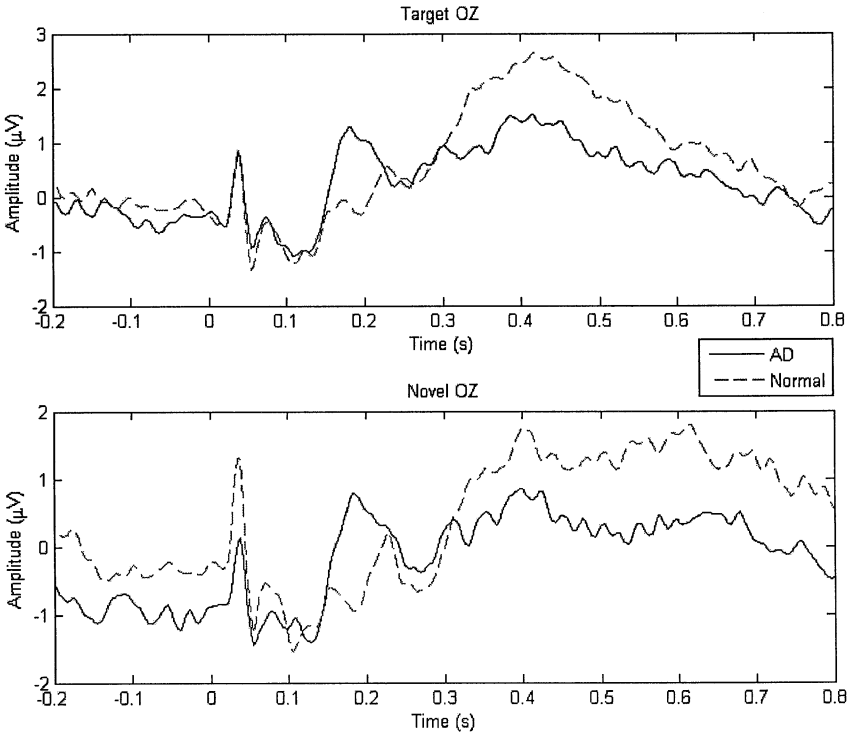

Figure B.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from Cohort B.

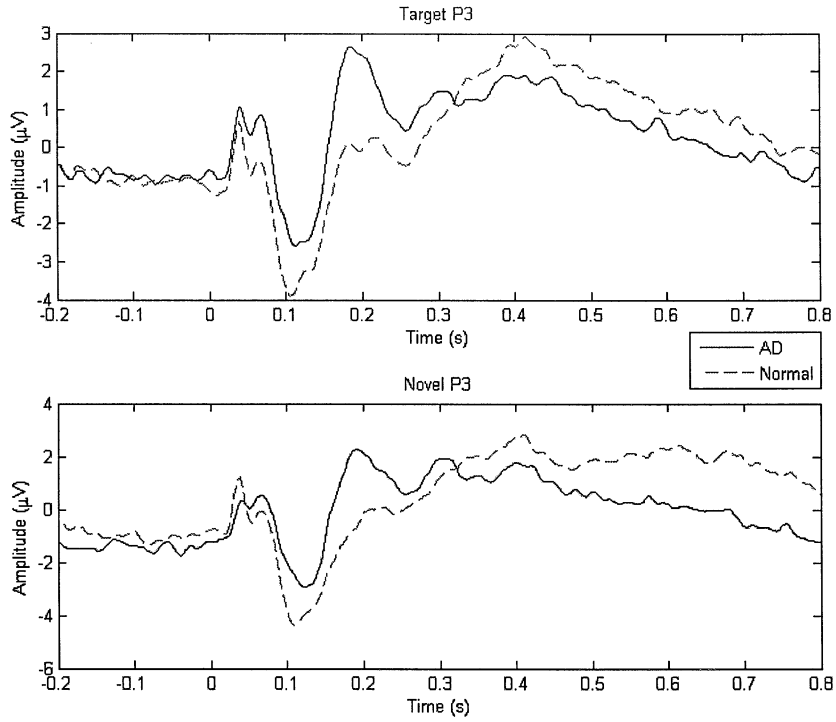

Figure B.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from Cohort B.

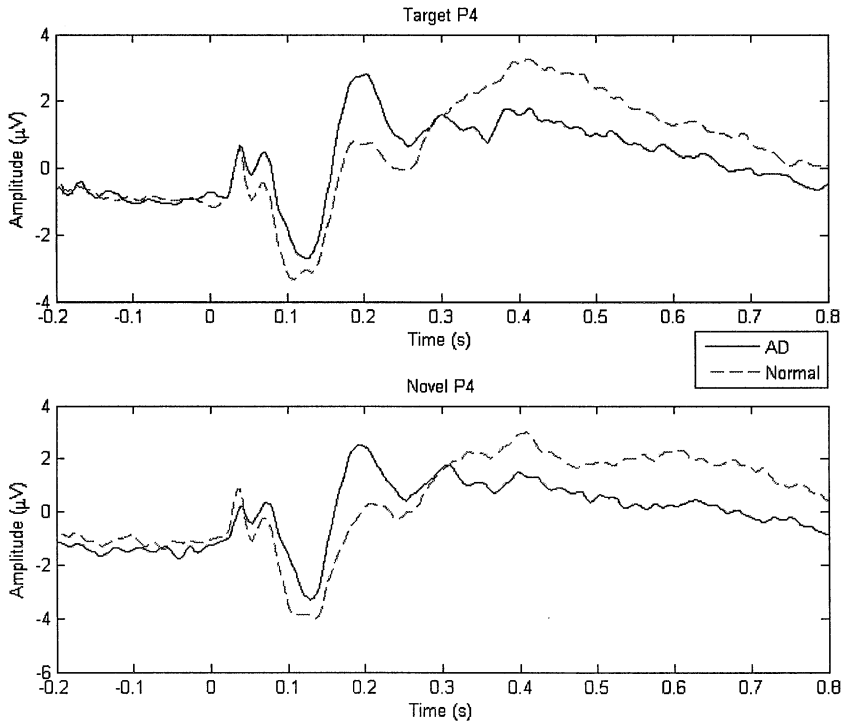

Figure B.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from Cohort B.

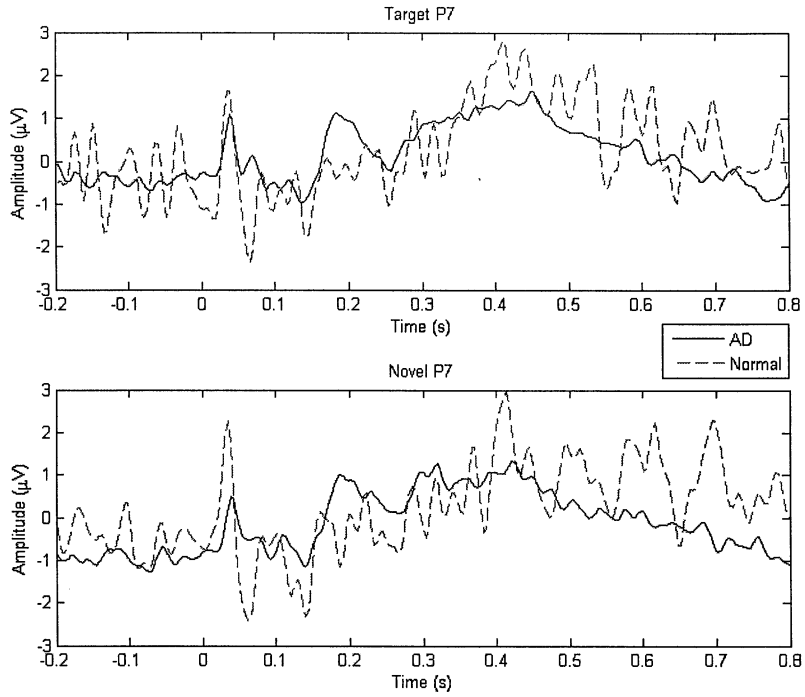

Figure B.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from Cohort B.

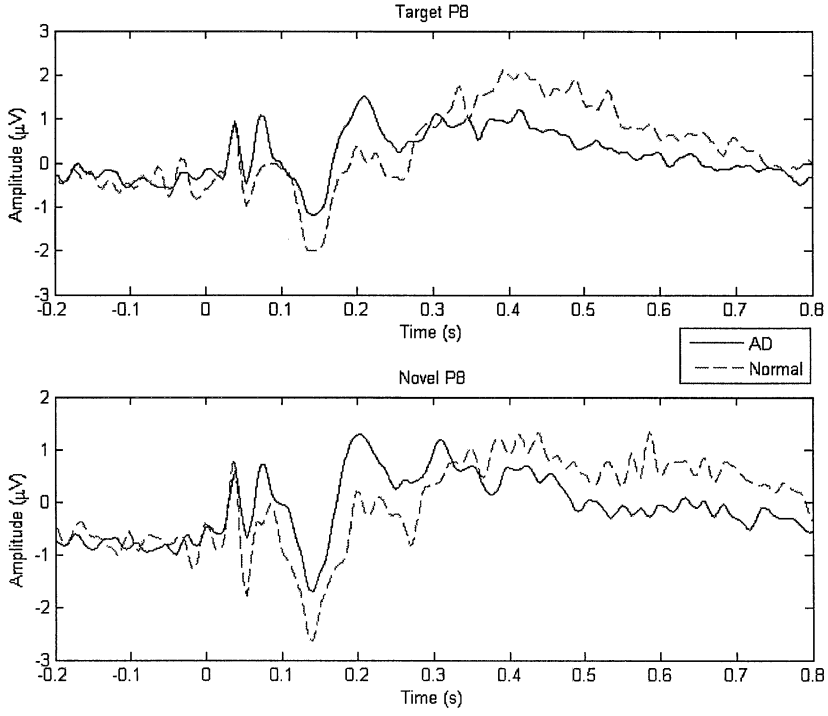

Figure B.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from Cohort B.

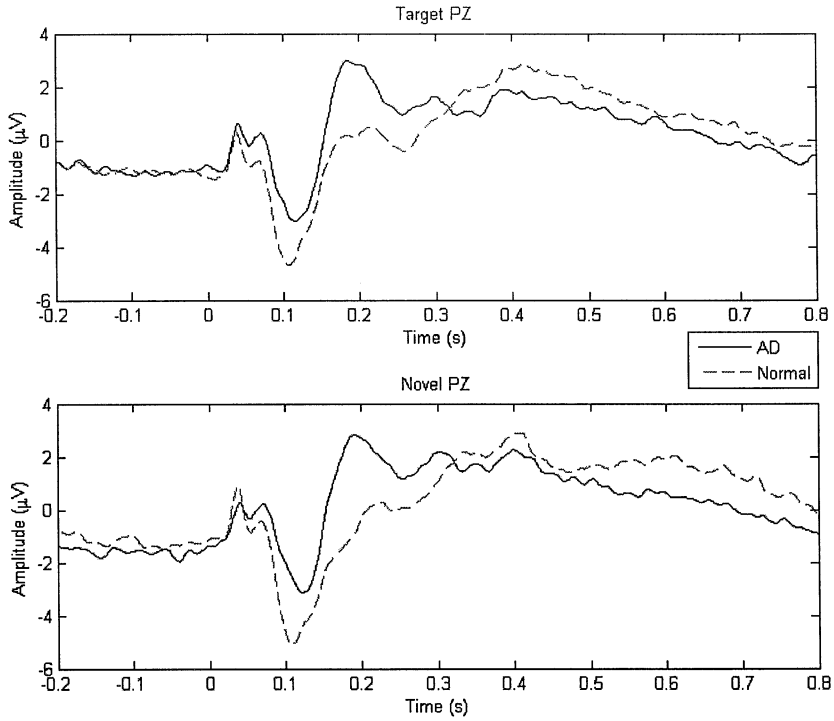

Figure B.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from Cohort B.

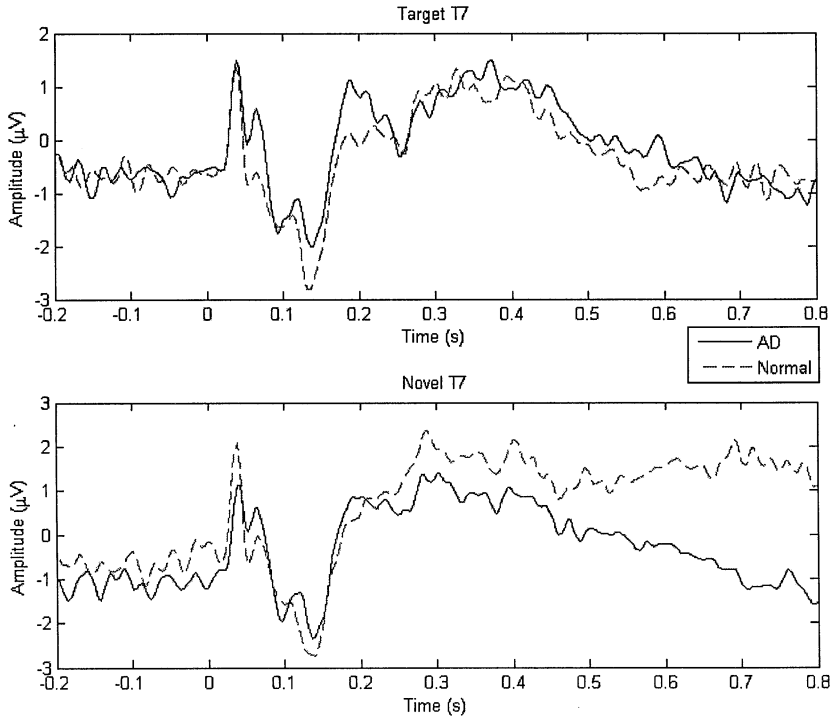

Figure B.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from Cohort B.

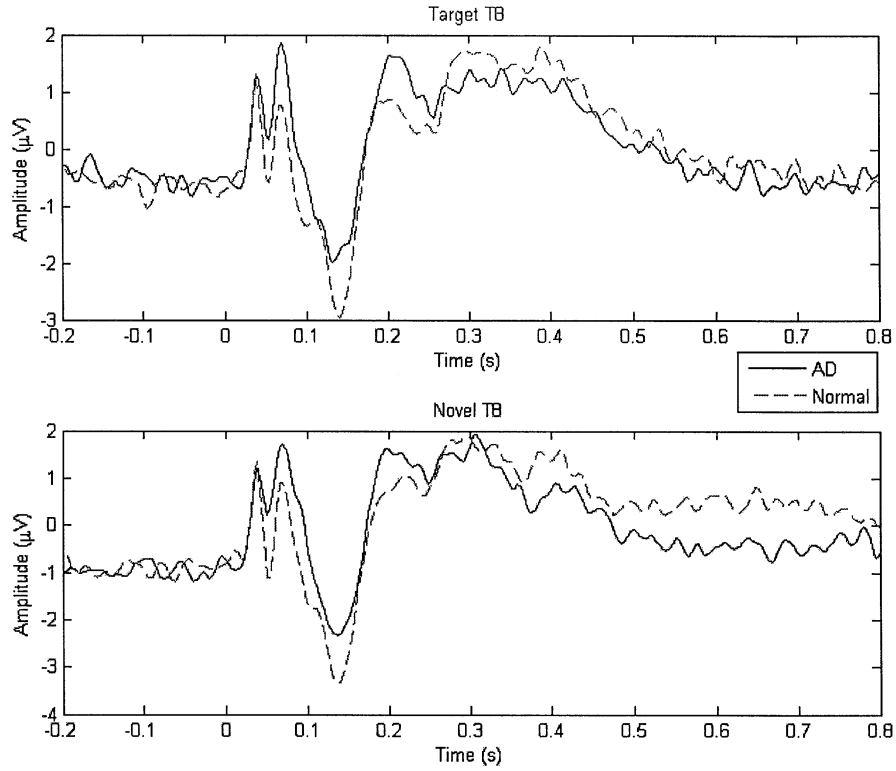

Figure B.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from Cohort B.