Event-related Potentials in Electroencephalography: Characteristics and Single-trial Detection for Rapid Object Search

Yonghong Huang
B.S., South China University of Technology, China, 1990
M.S., University of New Brunswick, Canada, 2005

Presented to the Division of Biomedical Engineering within
The Department of Science & Engineering
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
in
Biomedical Engineering

June 2010

Department of Science & Engineering

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of

Yonghong Huang

has been approved

Dr. Deniz Erdogmus, Thesis Advisor
Adjunct Assistant Professor, Northeast University

Dr. Misha Pavel, Thesis Advisor
Professor

Dr. Todd K. Leen
Professor

Dr. Santosh Mathan
Principal Scientist, Honeywell International

Dr. Barry Oken
Professor

Dedication

To my husband, Zhong, and my sons, Jordan and Matthew

Acknowledgements

First and foremost I would like to express my sincere gratitude to my thesis advisors, Dr. Deniz Erdogmus and Dr. Misha Pavel. Without their guidance and persistent support, this dissertation would not have been possible. Dr. Pavel accepted the daunting task of teaching me the fundamentals of neuroscience. He helped me hone my scientific thinking and communication skills. His advice on my research was invaluable. I am most grateful to my principal advisor, Dr. Erdogmus, for guiding me in the field of signal processing and machine learning in brain-computer interfaces. His passion and constant enthusiasm for research never failed to inspire me. I appreciated his open-door policy that allowed for constant communication flow and his encouragement throughout my research. I have been extremely fortunate to have Dr. Erdogmus and Dr. Pavel as my mentors. I am very grateful to both of them for providing me a smooth transition from the Computer Science & Electrical Engineering (CSEE) division to the Biomedical Engineering (BME) division in the middle of my PhD program. The interdisciplinary nature of my PhD program has profoundly impacted my thesis work.

I would like to thank my committee members for their valuable suggestions and careful assessment of my work. To Dr. Todd K. Leen for teaching me machine learning techniques and technical writing through the Medtracker project and courses. One part of my thesis, the mixed model and the Fisher kernel, was inspired by the work done by Dr. Leen's group. To Dr. Santosh Mathan for advising me throughout the NIA project, which directly led to my thesis. To Dr. Barry Oken for being available at a moment's notice to answer my questions on brain signal characteristics. To Dr. Subo Song for being my advisory committee member.

I would like to especially thank my previous advisors, Dr. Kevin Englehart and Dr. Bernard Hudgins, for helping me build a solid foundation in biomedical engineering. Their encouragement and support was invaluable on my postgraduate scholarship from NSERC in Canada which culminated in this PhD program.

I am thankful to all of my colleagues and friends who made my stay at the university a memorable and valuable experience. I would like to give special thanks to Dr. Kenneth E. Hild for patiently correcting my writing and providing suggestions on my research. Kenny has been a great friend and teammate these past two years. I am very fortunate to have had the chance to work with him. Many thanks to the people in the laboratory of CSEE and BME for being subjects in the experiments, to Zhengdong Lu and Tian Lan for collaborating on the project, discussing the research and helping with data collection.

This work was supported by DARPA under contract HM1582-05-C-0046 and by the NSF under grants ECS-0524835, ECS-0622239, and IIS-0713690. Satellite images included in this paper were produced by DigitalGlobe Inc., Longmont, CO 80501, USA. (c) 2003. The two IA datasets used in the experiments were collected at the Honeywell Human-Centered Systems Laboratory (Minneapolis, MN).

Finally I thank my family for their years of support and encouragement. To my Dad, Zhekung, for teaching me determination and supporting me. To my brother, Ruizhu, for his caring and best wishes. To my best friend and sister, Linda, for her encouragement and being a part of our family. To my lovely two sons, Jordan and Matthew, for being understanding when there wasn't time to visit the zoo or a bedtime story. Most of all to my husband, Zhong, for being beside me on this incredible journey. I am so thankful for your unwavering support, for cheering me up and for standing by me through the good times and bad.

Contents

List of Tables

List of Figures

Abstract

Event-related Potentials in Electroencephalography: Characteristics and Single-trial Detection for Rapid Object Search

Yonghong Huang, M.S.

Doctor of Philosophy
Division of Biomedical Engineering within
The Department of Science & Engineering
and the Oregon Health & Science University
School of Medicine

June 2010
Thesis Advisors: Dr. Deniz Erdogmus and Dr. Misha Pavel

Brain computer interfaces (BCIs) provide a non-muscular avenue for the user to communicate with others and to control external devices. Over the last two decades BCIs have been developed to assist the severely motor-disabled people, such as traumatic brain injury, stroke, or amyotrophic lateral sclerosis.

Electroencephalography (EEG) is one of the most popular noninvasive BCI approaches. The inputs to EEG-based BCIs are event-related potentials (ERPs), which are neural signatures representing the responses to an external stimulus. Traditional BCI systems,

which have had some success, make inferences based on trial-averaged ERPs, where each trial consists of one stimulus.

In this thesis, (1) we develop a single-trial, EEG-based BCI to increase the throughput of visual image search and (2) we unveil a neural correlate of human visual perception that occurs in rapid visual-recognition tasks.

Our first task is to develop a BCI. Our BCI makes inferences from single-trial ERPs; hence, it is more efficient than traditional methods. It uses cross-session training and a novel, hybrid generative/discriminative classifier (which combines a mixed effect model and a support vector machine via a Fisher kernel) to improve ERP detection performance, and it uses dimension reduction and incremental learning to reduce computational complexity. Based on the analysis of our BCI, we conclude that: single-trial ERP detection is possible; cross-session training outperforms the often-used single-session method; our hybrid classifier has a detection performance that is as good or better than some of the well-known classifiers; and dimension reduction and incremental learning substantially reduces computational complexity and they do so without an associated drop in detection performance.

Our second task is to characterize a neural correlate of human visual perception. Our approach involves measuring physiological signals and behavioral performance as a function of both the difficulty of the task (measured by the length of time images are available for viewing) and the difficulty of the target (estimated by the minimum viewing time required for a fixed detection rate). We find that the neural responses are highly correlated with both target difficulty and task difficulty. Based on these findings we further surmise that, during visual information processing, the brain dynamically allocates additional cognitive resources under increasingly difficult conditions.

Chapter 1

Introduction

Brain computer interfaces (BCIs) provide a non-muscular avenue for the user to communicate with others and to control external devices. Over the last two decades, BCIs have been developed to assist the severely motor-disabled people, such as traumatic brain injury, stroke, or amyotrophic lateral sclerosis. BCIs infer a user's intentions using only recorded neural activity, hence speech ability and motor control are not required. The development of sophisticated BCI is made possible by modern computer technology and our newly-obtained knowledge of brain activity. Most current BCIs focus on assisting physically-disabled subjects. In contrast, we attempt to enhance the abilities of healthy subjects for cognitive tasks. Our particular application is to use a BCI to increase the throughput of visual image search. We examine human visual perception and develop advanced techniques to achieve this goal.

Human visual perception is a complex process engaging sensory, cognitive and motor events. Unveiling the neural correlates of visual detection and its underlying component processes is challenging due to the complexity of the human brain and the massively parallel interactions that exist in the brain. In the recent half century, advances in technology and knowledge of the human brain make it possible to examine visual information processing at a physiological level. Evoked neural responses, elicited by sensory, cognitive and motor events, can be detected by means of noninvasive electrical recordings from the scalp using, e.g., electroencephalography (EEG). These event-related potentials (ERPs), associated with human perceptual judgments, reflect patterned neural activities.

This thesis consists of two main parts: one related to engineering and one related to science. In this thesis, (1) we develop a single-trial, EEG-based BCI to increase the throughput of visual image search and (2) we unveil a neural correlate of human visual perception that occurs in rapid visual-recognition tasks.

First, we develop an ERP-based BCI that increases the synergy between the image search platform and the human operator by tapping into the split-second perceptual judgments of humans. The goal is to assist intelligence analysts to improve their efficiency in scanning images. ERPs that are elicited by rare target images and non-target images (distractors) have different patterns. The BCI is able to identify the presence of a target image seen briefly by an expert simply by detecting the ERP patterns associated with the presentation of target stimuli on a single-trial basis. A successful ERP-based BCI depends on robust ERP detection. We develop advanced machine learning methods for single-trial ERP detection. ERP detection is hampered by many factors, such as noise, outliers, high dimensionality, non-stationarity, and small training sets. For this reason, our BCI uses cross-session training and incremental learning. To further improve the single-trial detection performance, we apply a hierarchical Baysian model to analyze the ERP patterns on both the population and the individual levels and we develop an advanced discriminant classifier based on modern statistical learning theory and on population-informed models of individual ERP trials.

Second, we use the ERP as a physiological marker to examine the human visual perception and cognition in a recognition task, in contrast to the previous studies. The main difference is that they control only the target difficulty, whereas we control both the target difficulty and the task difficulty. We simulate a natural (visual) environment, which often involves rapidly changing stimuli, using the dynamic paradigm known as rapid serial visual presentation (RSVP). The goal is to assess the dynamics of resource management, such as attention allocation, using brain signatures (ERPs) involved in human visual information system. We examine the dynamics of attention allocation as a function of the task requirement and the complexity of visual stimuli.

1.1 Objectives

The primary objectives of this work are to unveil the neural signatures of human visual information processing and improve single-trial ERP detection performance. The direct motivation of this research is to speed up image analysis by enhancing human information processing capacity based on ERP pattern classification. Central to this work are two assertions: (1) allocation of attention and cognitive processing resources is a fundamental determinant of performance in perceptual tasks and (2) the process of classification plays a crucial role in the detection performance of BCIs.

In an effort to improve performance, this work seeks to generalize the single-trial ERP detection problem. Correspondingly, our secondary objectives are:

- Design a framework for rapid visual image search based on single-trial ERP detection. Specifically, demonstrate the relative efficacy of different statistical pattern recognition techniques, explain the factors that affect their performance, and specify their optimum configuration for single-trial ERP detection.
- Demonstrate the feasibility of adaptive cross-session training and the efficiency of incremental learning for single-trial ERP detection.
- Develop a hybrid generative/discriminative method combining the complementary strengths from the generative model and the discriminative method to improve single-trial ERP detection performance.
- Demonstrate that the channel projection method successfully extracts representative features, reduces dimensions and improves the ERP detection performance.
- Establish a better understanding of the neural correlates of human visual cognitive processes. Specifically, assess the dynamics of attention allocation using a rapidly changing environment to identify neural activity directly responsible for human visual perceptual judgements.

This work accomplishes these objects using a complement of empirical and theoretical investigation. The major contributions include the discovery of neural signatures of

human visual information processing in a dynamic task, the development of a framework of ERP-based BCI using cross-session incremental training and the development of an improved hybrid classifier combining a generative model and a discriminative model for single-trial ERP detection.

1.2 Background

1.2.1 Event-related Potentials

The primary focus of this thesis is ERP, a time-locked EEG brain response associated with perceptual and cognitive events. ERPs consist of a series of positive and negative components. In general, the earlier components are most commonly associated with sensory events and the later components with more cognitive events [119]. A specified ERP has a characteristic amplitude, latency and scalp distribution.

ERPs have drawn a lot of attention in the field of cognitive neuroscience. Long before computers were available for the recording EEGs, the first ERP recording from awake humans was performed by Pauline and Davis in 1935. The first published computer-averaged ERP waveform was from Galambos et al. in 1962 [33]. The modern era of the ERP study began in 1964 with Walter et al.'s finding of the first cognitive ERP component, contingent negative variation, which reflected the subject's preparation of the upcoming target [130]. The next major advance was the discovery of the P300 component by Sutton et al. in 1965 [120]. They found that the unpredictable stimuli elicited large, positive P300s and predictable stimuli elicited the P300s with much smaller amplitudes. The P300 can be affected by attention level, task difficulty, stimulus occurrence probability, stimulus randomization, and subject age. Over the last few decades, many researchers have employed the ERPs, due to its high temporal resolution, to assess highly specific neural processes [84, 2, 96, 70, 27, 119, 54, 73]. Picton [96] and Oken [84] provide overviews of the $P300$, which include technical aspects of $P300$ recording , psychological aspects of $P300$, $P300$ characteristics, subject parameters, and clinical studies of $P300$.

1.2.2 Brain Computer Interfaces

The primary source of information for BCI is EEG, a signal with a number of frequency components. The first rhythmic electrical activity was recorded from the cortical surface in animals by Richard Caton in 1875 [13]. Hans Berger first reported the measurements of electrical activity from the human scalp in 1929 [4], which later became known as EEG. EEG signals are generated inside the brain and are recorded by a set of electrodes using standard locations on the scalp. EEG analysis is very challenging because EEG data represents a conglomeration of numerous neural sources of activity, making it difficult to isolate individual neural cognitive processes. In the more than half a century, since Hans Berger's finding, EEG and other brain electrophysiological measures have been applied mainly to diagnose neurological disorders in the clinic and assess brain function in the laboratory.

Researchers have long speculated that the measurements of brain activity might convey messages or commands to the external world so that it may be possible to replace or augment the brain's normal neuromuscular interaction with the environment. This new non-motor channel for communicating a user's intent to control an external computer-based device can be achieved through a BCI. The goal of BCI is to control an external device by decoding a user's intents from brain signals, instead of using motor controls. The term BCI was first used by Jacques Vidal in the 1970s [128, 129]. Vidal's system used visual evoked scalp-potentials recorded from the visual cortex to determine the gaze direction, which was used to move a cursor based on the user's intent. The first convincing demonstration of a direct functional interface between a brain and a robotic arm was documented in 1999 by Chapin [15].

In the past decade, there has been a large amount of research performed in the highly multidisciplinary field of BCI. A wide range of practical applications of BCI have already been demonstrated in the laboratory and in limited clinical use [132, 134, 55, 6, 67, 38]. BCI is primarily designed for providing a means of communication and control for those

with severe motor disorders, particularly those suffering from spinal cord injury, amyotrophic lateral sclerosis, brain stem stroke, cerebral palsy and muscular dystrophy. Simple BCI applications include basic binary control systems (lights, temperature, TV, and hand orthosis), basic word-processing and Internet access. More complex BCI applications can be used to control devices such as motorized wheelchairs, robotic arms, and limb prostheses. BCI technology makes a huge difference in the productivity and quality of life of severely disabled people.

There are a variety of other methods for recording brain signals that might be used in a BCI. For example, we can record electrical or magnetic fields, or use functional magnetic resonance imaging, positron emission tomography and infrared imaging. However, most of these methods are technically intricate, prohibitively expensive, have slow temporal response or limited real-time capabilities, or are at an early stage of development. In reality, therefore, recording electrical potentials is more practical for clinical applications at the present and for the near future [131]. There are alternative electrophysiological recording methods. The electrical fields produced by brain activity can be recorded from the scalp (EEG), from the cortical surface (electrocorticogram (ECoG)), or from areas inside the cortex (local field potentials (LFP) or action potentials from a single neuron). Presently, intracortical BCI data (LFP and action potentials) come mainly from animals, primarily monkeys. Human BCI data are confined almost entirely to EEG and short-term ECoG. Notice that EEG, unlike the invasive options, avoids the risks of brain surgery (tissue damage and scarring), and is one of the most popular BCI approaches.

Different EEG features can be used to convey user's intent. Based on the categories of EEG components, there are three major types of EEG-based BCIs: $P300$ BCI, slow cortical potential (SCP) BCI, and sensorimotor rhythm BCI. The $P300$, which is one of the most important ERP components, represents brain activity that is elicited in response to a salient or attended event (stimulus). It appears in the centroparietal EEG about 300ms after presentation of the stimulus. Donchin's group [30, 25] demonstrated a $P300$-based BCI that uses the oddball paradigm [104]. The BCI presented the user with

a matrix of 6 by 6 cells, each containing one letter or symbol. EEG was recorded while the letters/symbols were flashed in succession. Only the letters/symbols desired by the user evoked a large $P300$. By detecting the $P300$s associated with the presentation of the desired letters/symbols, the $P300$-based BCI system achieved a communication rate of 7.8 characters per minute for healthy subjects and with an accuracy of 80% accuracy. Their report confirms that a $P300$-based BCI enables a virtual keyboard operation without requiring any activation of skeletal muscles. Another completely different type of EEG component is the SCP, which has a latency ranging from 500ms up to 10s after the stimulus onset. The SCPs indicate the overall preparatory/excitation level of a cortical network. Behavioral and cognitive performance improves after subjects learn to increase the negativity of the SCP, while cognitive and behavioral performance reduces during positive cortical potentials. Birbaumer and his coworker demonstrated a SCP-based BCI on locked-in patients. These users were able to move a cursor to a targets located at the bottom (more positive SCP) or top (more negative SCP) of a computer screen, which allowed them to select letters, words, or pictograms [7, 61]. The sensorimotor rhythm approach, which differs from the $P300$ and the SCP approaches, uses $8-12$ Hz (mu) and $18-26$ Hz (beta) oscillations in the EEG recorded over sensorimotor cortices. Several laboratories have demonstrated BCIs based on sensorimotor rhythms [135, 93, 105, 60], which are used for typing on a virtual keyboard. All three types of EEG-based BCI can be used to support basic word-processing or other simple control tasks.

There are two major advantages of $P300$-based BCI over other BCI approaches. Present-day BCIs mostly depend on visual stimuli. In real applications, people who are locked-in or otherwise severely disabled, may not be able to follow such stimuli, especially if they change rapidly. In this case, $P300$-based BCIs that use auditory cues may prove effective. Another important property of the $P300$-based BCI is that it does not require subject training. Both trained and naive subjects generate a $P300$ response to rare events in an oddball sequence, as shown in Sutton's original report [120]. In addition, the $P300$ has low inter-subject and intra-subject variation and is reliable [29].

Modern computer technology and our newly-obtained knowledge of brain activity make the development of effective BCI possible. The BCI system consists of several components: brain signal acquisition and processing, signal feature extraction, algorithm translation and device operation. Electrophysiological signals reflecting brain activity can be acquired from the scalp, from the cortical surface or from within the brain, and are processed to measure specific signal features, such as amplitudes of ERP. The signal features reflect the user's intent and are translated into commands that operate a device, such as a wheelchair. One of the fundamental aspects of BCI operation is that the system adapts to the user at the same time the user adapts to the system. The user must generate brain signals that encode intent and the BCI system must translate these signals into commands that accomplish the user's intent. The translation algorithm, also known as classification algorithm, is crucial for an effective BCI. Recently, a lot of effort has been spent on the development of classification algorithms for BCIs [68, 78, 9, 133, 127]. When designing classification algorithms for an ERP-based BCI, several aspects should be carefully considered, such as: noise, outliers, high dimensionality, non-stationarity, and insufficient data. A successful ERP-based BCI system depends on robust ERP detection.

1.2.3 ERP Detection: Single Trial vs. Trial Averaging

Extracting ERPs from EEG signals is very challenging. The amplitude of the ERP signals is quite small (1-30μV). The background EEG signals have magnitudes that are commonly 10 times than that of the ERP. The conventional strategy is to average across trials since this increases the signal-to-noise ratio and makes the ERP more detectable. Many studies have been undertaken on trial-averaged ERPs [52, 53, 73, 125, 18]. Makeig and colleagues attempted to resolve the question of what the ERP actually represented by a detailed analysis of EEG and ERP [73, 71, 72]. They compared the averaged ERP and assessed the inter-trial coherence, and used independent component analysis (ICA) to separate out the different electrical activities. However, there are several limitations associated with trial-averaging. First, the process of averaging filters out much of the

information about cortical dynamics. Second, it requires that each stimulus be presented multiple times, which directly compromises the efficiency. Hence, the trial-averaged approach is not an ideal strategy for time-demanding applications.

A different ERP detection strategy, the single-trial approach, has been developed in recent years and appears to be promising [87, 9, 123, 79, 117]. In the single-trial approach, sensor data are averaged over sensors for improving the signal-to-noise ratio, as opposed to the standard approach, which averages over trials. Parra and his colleagues developed a promising BCI, for visual object recognition, that uses linear spatial integration for single-trial ERP detection [87, 88, 108, 109, 102, 34]. Single-trial analysis of ERPs, which requires to uncover the origin of neural response variability and the precise timing and amplitude of individual ERPs, enables one to study parameters that cannot be controlled across trials, such as reaction time or perception. The single-trial approach, therefore, opens up new opportunities for studying the neurological origin of the ERP and behavioral relevance.

1.2.4 Rapid Serial Visual Presentation

Rapid serial visual presentation (RSVP) is an image presentation paradigm that is a useful tool for exploring the temporal characteristics of visual information processing. The idea of using rapidly-presented visual items was proposed in 1969 to assess visual perception rate [28] and recognition memory [100]. In the RSVP paradigm, visual items, such as images, are displayed sequentially at the same spatial location at a very high rate. In the RSVP paradigm, we can control the time for viewing a given item and the preceding and subsequence demands on the subjects. It has been widely used in behavioral and psychological studies, such as visual perception [28], short time conceptual memory [100, 99], attention [16], and target detection [66, 10].

1.2.5 ERP vs. Behavioral Response

An obvious alternative to using ERP detection in RSVP is to record a behavioral response in RSVP. For example, the user could press a button whenever a target is detected. Previous research indicates that there is no significant difference in detection performance of methods based on the EEG or a button press [35]. However, there are several advantages of ERPs over behavioral measures. First, an ERP does not require a behavioral response. This is one of the greatest advantages of the ERP technique, specifically in BCI applications for people with motor disorders. Second, an ERP provides a continuous (as opposed to a binary) measure of detection, which may provide insight into the time course and spatial distribution, of cognitive processes associated with target detection. An overt response, in contrast, reflects the output of a large number of individual cognitive processes, making it difficult to attribute a variation to a specific cognitive process. Third, an ERP has high temporal resolution and lower timing variance, whereas an overt response has larger variations in reaction time (RT) and accuracy [43]. Fourth, an ERP has a lower latency and can reveal a sign of neural processing well before an overt response [18]. Fifth, ERPs can reduce fatigue relative to physical response mechanisms, such as button presses, the latter of which may cause increased strain over long periods of usage.

Yet another option is to integrate the ERP detection and the behavioral response to bring performance benefits. This approach reduces the false alarms relative to the ERP approach and the temporal variances relative to the behavioral approach [35, 74, 43].

1.2.6 Neural Signatures of Visual Information Processing

Much of the current research on ERPs has been directed toward identifying particular ERP components as markers of specific aspects or stages of human perceptual processing, such as encoding, selecting, memorizing and decision making [40, 94, 95, 53, 52, 96, 2]. Hillyard [40] gave a complete review on ERP components as physiological signs of cognitive and linguistic processes. He pointed out that most of the earlier ERP components varied

as a function of physical stimulus parameters and were relatively insensitive to changes in perceptual processing demands. In contrast, the longer latency components appeared to be asociated with specific perceptual or cognitive processes. A broad negative ERP (latency $150 - 300$ms) was elicited by the attended stimulus. While the early negative ERPs gave some information on the attentional channel, the $P300$ seemed to be a particularly sensitive index of the degree of attention received by deviant stimuli in and out of the designated focus of attention. The amplitude of the $P300$ was generally largest for task-relevant stimuli. Campbell and his colleagues [96] investigated the $P300$ component correlates of human information processing using auditory stimuli. They found that as the probability of the stimuli decreased, the amplitude of the $P300$ became larger, later and more frontal. Johnson et al. [52] studied the human visual system by measuring the time course of neural correlates of object recognition. They found an early ERP component, arising around 135ms post-stimulus, that occurred when there were low-level feature differences between images. Unlike the early component, a late ERP component, arising around $150 - 300$ms, covaried with the subsequent reaction time and was correlated to recognition. Campbell [2] and Johnson [53] had the same finding on the effect of stimulus difficulty on ERPs. They found that the amplitude of the $P300$ component decreased and the latency increased when the difficulty level of stimulus increased. Recently Philiastide and his group [94, 95] characterized the neural correlates of perceptual decision making in the human brain. They used ERPs to assess the human visual perceptual system in a setting where the images are separated in time. They found an ERP component that occurred around $200ms$ after the stimulus onset, the peak value of which was inversely correlated with behavioral performance and stimulus difficulty. None of the above studies used the RSVP paradigms.

Researchers have gained valuable insights into the mechanisms of visual processing by using the RSVP to push the visual system to its temporal limits [122, 125, 58]. Thorpe and his colleagues [122] designed a nature-scene target detection task to assess the speed

of visual processing using the RSVP and the ERP. They demonstrated that the visual processing needed to perform the highly demanding visual task could be achieved in under $150ms$. Their finding suggests that it might be possible to use neural activity associated with perceptual decision making to identify targets within images.

1.2.7 An BCI Application – Visual Object Detection

One BCI application that is the focus of this thesis is detection of targets in satellite images of earth surface. In recent years, the volume of commercial geospatial imagery and other imagery has increased to unprecedented levels. Finding instances of an object (referred to as *targets*) in a large volume of images is an important task in many domains, such as medical archives, web mining, national security, criminal justice and aerial geospatial imagery used for civilian planning. In such complex domains, where there exists a large variability within and across images, human analysts are very skilled at exploiting contextual cues and incorporating prior knowledge, whereas fully-computerized machine learning solutions oftentimes perform poorly. On the one hand, manual search is laborious and there is a shortage of skilled workers. On the other hand, the visual acuity and perceptual capabilities of humans far exceed those of current automated techniques for broad context image search and retrieval tasks [32]. For these reasons, there is a great need to develop an alternative method. Ideally, we would like to build a target detection system that combines the accurate perceptual judgments from humans and the computational power from machines.

BCI technology, aside from assisting people with motor disability, can be used to enhance the cognitive performance for healthy users. When a human detects a target, the brain produces an ERP in the EEG. An automated method is then used to detect the ERP. Part of the motivation behind the application of ERP detection on visual object search is that automated methods are better at detecting a relatively simple and approximately

fixed-shape ERP in EEG data than they are at detecting arbitrarily-shaped, complex targets in images. Recently, a number of researchers have demonstrated the use of ERP-based BCI for visual object detection. Sajda et al. [103, 89, 35] developed a visual object detection system, named cortically-coupled computer vision. The system detected a subject's ERPs, while the subject views a stream of images presented at a high rate (RSVP). They then use a weighted linear logistic classifier (LLC) [91] for single-trial ERP detection. The system enhances the performance of satellite imagery analysts by performing a rapid sorting of image 'chips' drawn from a broad-area satellite image. Shamlo's group developed a target search system that uses the ICA to spatially filter high-density EEG signals [5]. Sheoney's group demonstrated an image classification system using single-trial ERP detection [117, 56].

1.3 Thesis Outline

Chapter 2 describes several experimental factors: subjects, stimuli, image display paradigm, tasks, EEG electrodes and experimental procedures. We design several experiments involving intelligent analysts and naive subjects. We discuss EEG data acquisition and data preparation, which may affect the quality of EEG data and ERP analysis.

Chapter 3 presents several insights on neural correlates of human visual perception. We analyze neural correlates on realistic images in the RSVP image display paradigm. We investigate the relationship between the ERP characteristics and the detection difficulty in terms of visual stimulus complexity and task difficulty. We demonstrate that more difficult trials have smaller ERP magnitudes. The early component appears around 110 millisecond (ms) and the late component appears around 250ms. The ERP pair waveforms for the easier trials are more similar than those for the harder trials. Both target difficulty and task difficulty are correlated with neural activity. ERPs predict behavioral performance. More difficult trials have longer RT and higher RT variance. Smaller ERP magnitude associated with longer RT and higher RT variance. The finding indicates that,

for dynamic tasks in visual information processing, the brain's resource management may trigger allocation of additional cognitive resources resources, such as attention, to a given visual stimulus, under increasingly task or target difficulty conditions.

Chapter 4 reports the design and performance of our BCI system for target image search using single-trial ERP. The feasibility of cross-session training and the efficacy of incremental learning on cross-session data are demonstrated. The results show that high inter-session data variance can be reliably mitigated. The cross-session training method, which uses more data, outperforms the single-session method. The incremental learning is as effective as the batch learning in spite of the fact that it has only $1/3$ of the computational cost. The ERP-based approach is 5.3 times faster than the traditional image viewing approach, and provides equal or higher detection rate. Linear and nonlinear support vector machines (SVMs) significantly outperform the LLC for single-trial ERP detection on our data. Channel dimension reduction using the linear discriminant analysis (LDA) successfully extracts representative features and thus improves the performance, and simultaneously reduces the computational cost.

Chapter 5 reports the design of a generative learning model, mixed-effects models (MEM), which takes advantage of domain knowledge, on modeling and classification for single-trial ERP. The MEM constructs a probability density model from which we can compute decision functions for ERP detection and that provides underlying information about ERP patterns. We propose a hybrid ERP detector, which combines a generative model and a discriminative model. More specifically, the new method combines the MEM and the SVM using the Fisher kernel. The results show that the new method achieves substantial improvement over the generative model (MEM) and marginal improvement over the discriminative model (SVM).

Chapter 6 summaries our findings and contributions, and provides recommendations for future investigation.

Chapter 2

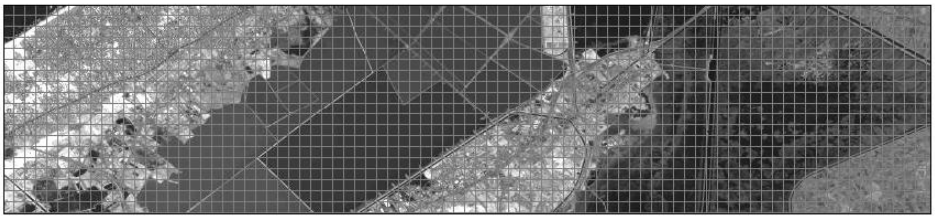
Empirical EEG Data Collection and Preparation

Acquisition and processing of EEG data as well as the experimental design are a fundamental part of this work. Experimental design, data collection, and data preparation are critical procedures in the cognitive study. Picton et al. provided a complete guideline for EEG experimental design and EEG recording standards [97]. The goal at the experimental stage is to record EEG data reliably, measure them accurately, and prepare the data properly. Tasks are designed specifically to elicit the cognitive processes. In this thesis we design several experiments on different study purposes for ERP characteristics in human visual perception and single-trial ERP detection. In our study, subjects perform target detection by clicking on a button as soon as they see target. At the same time, we record their EEG signals.

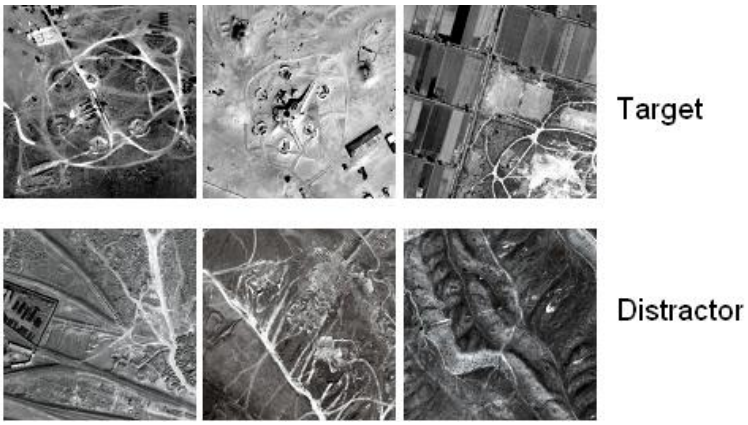
2.1 Subjects

Five groups of human subjects are recruited in the study, including professional intelligence analysts (IAs) and naive subjects. None of these participants have previous experience with the RSVP modality. Written informed consents approved by the internal review board are obtained from all participants. All subjects have normal or corrected visions.

- Group #1 (IA Dataset #1) – three professional IAs are recruited at Honeywell.



(a) Broad-area image



(b) Target and distractor examples

Figure 2.1: (a) A broad-area aerial image is segmented into hundreds of image chips. These chips are rapidly displayed to the subjects one-at-a-time. (b) Examples of target (surface-to-air missile site) and non-target (distractor) chips. Target are encountered rarely.

- Group #2 (IA Dataset #2) – three professional IAs are recruited at Honeywell.
- Group #3 (naive dataset #1) – four graduate students (ages range from $25 - 35$; all males) are recruited at OHSU.
- Group #4 (naive dataset #2) – ten naive subjects are recruited at Honeywell.
- Group #5 (naive dataset #3) – ten graduate students (ages range from $25 - 45$; four females) are recruited at OHSU.

2.2 Stimuli

Large-scale satellite images are decomposed into hundreds of smaller chips, which are labeled according to whether they contained a target. A example of a satellite broad-area image (27000×6500 pixels, representing an area of over $200km^2$) is shown in Figure

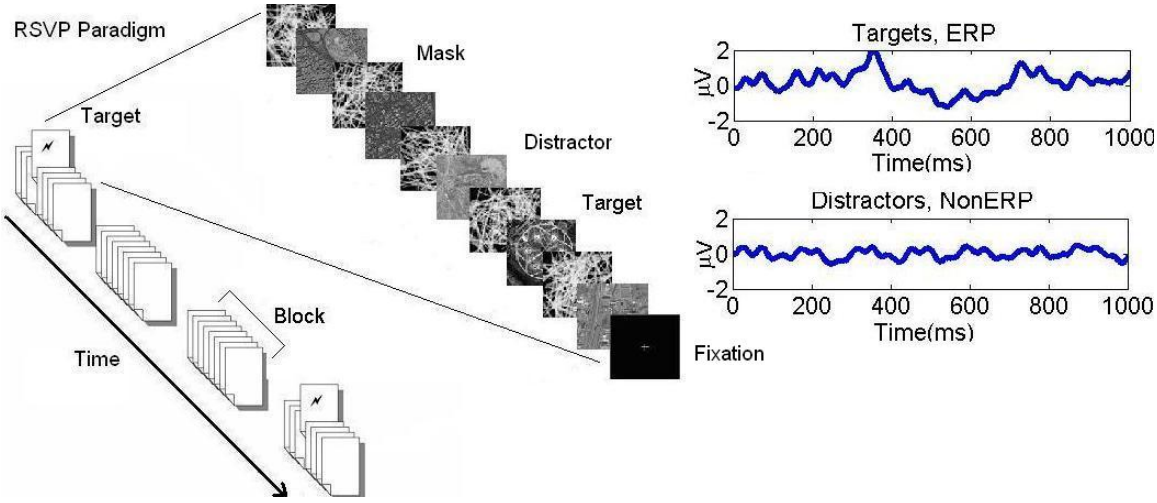


Figure 2.2: An illustration of the rapid serial visual presentation (RSVP) image display modality. Each RSVP block starts with a fixation screen and each image (a target or a distractor) is followed by a single pattered mask (in data collection for naive dataset #2 and #3 only). One half of the blocks contain a single target. During each RSVP block a sequence of images are displayed at the high presentation rates, such as 100ms/image. On the right, the upper trace is the trial-averaged ERP associated with the target stimuli (in one representative channel) and the lower trace is the trial-averaged ERP associated with the distracter stimuli. The stimulus onset corresponds to 0ms.

2.1 (a). Each chip (500×500 pixels) represents an area of $0.09km^2$. The targets are surface-to-air missile (SAM) sites, characterized as star pattern with six circles and in the gray scale satellite imagery. There are 105 different SAM sites. The size of the targets is small and the scale, shape, orientation, and location of the targets naturally varies. Some image chips only contain a portion of the pattern. The three target image examples and the three non-target (distractor) image examples are illuminated in Figure 2.1(b).

2.3 RSVP Image Display Paradigm

The image chips are presented using the RSVP image display paradigm as shown in Figure 2.2. The infrequent targets (SAM sites) are embedded in sequences of distractor images. The chips are presented on a 21 inch CRT monitor using Presentation software (Neurobehavioral Systems, Albany, CA). Each image subtended 22×22 degrees of visual

angle. During each RSVP block, a sequence of images is displayed at a high presentation rate, such as $100ms$/image, i.e. 10 Hz (the image duration is defined as the inter-stimulus duration). Images are presented in short bursts of a few second duration. In data collection for naive dataset #2 and #3, patterned masks are shown after every image to disrupt the accumulation of information from the retinal image representation. In other studies, there is no masks. To break monotony and minimize possible eye strain, consecutive blocks are separated by a fixation screen of user-controlled duration.

2.4 Task

The subjects perform rare target detection. The stimuli are presented to the subjects using the RSVP paradigm as Figure 2.2 in all experiments. The subjects are instructed to keep their eyes on the center of the screen and avoid eye blinks during image viewing, use the black screen to blink or rest, and keep their hands close the space bar so that they can respond as quickly as possible to the presence of targets. Both behavioral and physiological data are collected except that for Group #4 we only collect behavioral data.

2.5 Experimental Procedures

In all experiments, subjects perform SAM-site target detection in the RSVP paradigm. The RSVP sessions are structured in two phases – training phase and test phase. Each subject is trained prior to testing. In the training phase, images are drawn with replacement from the image chip set and shown in a random order. Subjects receive feedbacks on their responses at the end of each block. In the test phase, the chips are presented in the spatial order in which they occur in the broad area image. There is no feedback in the test phase. There are slightly different procedures for each experiment.

Experiment #1 (Group #1, IA Dataset #1) – three professional IAs are recruited. The RSVP rate of $100ms$/image is for subject 1 and subject 3, and the RSVP rate of

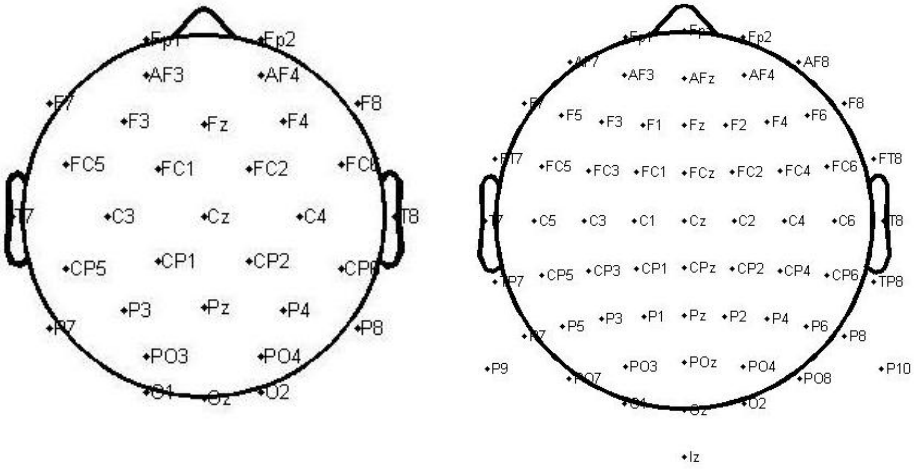
$150ms$/image for subject 2. Each subject is trained on one session and tested on seven sessions. Each session contains dozens of blocks (sequences), and each block is five seconds in length. Half of the sequences contain single target instance. There are around 50 targets in the train session, and only one real target and several synthetic targets in the test sessions for each subject.

Experiment #2 (Group #2, IA Dataset #2) – three professional IAs are recruited to compare the neurophysiologically-driven target image search with the conventional broad-area manual image search. All subjects have experience with a broad range of imagery and target types. They are all trained in the use of geo-spatial analysis tools. None of them are familiar with the RSVP paradigm. The RSVP rate is $100ms$/image for subjects 1 and 2. Subject 1, 2 and 3 are tested on one session, four sessions and seven sessions. The RSVP rates for subject 3 are $60ms$/image for training and four of the test sessions and $100ms$/image for the remaining three test sessions.

Experiment #3 (Group #3, naive dataset #1) – four male graduate students are recruited. To assess cross-session performance, we collect data at different times and under different experimental conditions. Namely, data are collected from each subject during one morning session and one afternoon session each day on five different days. Each session contain 200 blocks. The image duration is $100ms$/image. Each block contain 37 images and is five seconds in length. 75% of the blocks contain a single target instance. Images are drawn with replacement and shown in a random order. Each subject participats in 10 sessions in total.

Experiment #4 (Group #4, naive dataset #2) – Ten naive subjects are recruited. For each subject, our block design consists of five sessions, where the images (target, distractor, and mask) in each session are displayed for one of five different durations, i.e., 25, 50, 100, 150, and 200 ms. Each session consists 210 blocks, each of which lasts for one second. One half of the blocks contain a single target.

Experiment #5 (Group #5, naive dataset #3) – Ten graduate students are recruited. In the training phase, there is no pattern mask. Images are shown in a random order at a



(a) 32 channel (b) 64 channel

Figure 2.3: Layout of the 32 and 64 channel EEG electrode placements. The letter at each electrode identifies the particular subcranial lobe (FP, prefrontal lobe; F, frontal lob; T, temporal lobe; C, central lobe; P, parietal lobe; O, occipital lobe). The number or the second letter identifies its hemispherical location (Z, denoting line zero refers to an electrode placed along the cerebrum's midline; even numbers represent the right hemisphere; odd numbers represent the left hemisphere. The numbers are in ascending order with increasing distance from the midline.)

rate of 100 ms/image. Train session contains 75 blocks. Each block contains around 37 images and is about five seconds. There are 75% of the trials containing a single target instance. In the test phase, patterned masks are used after every image. There are four test sessions per subject, one for each image duration (50, 100, 150, and 200 ms). Each session consists of 210 blocks, each of which lasts for one second. One half of the blocks contain a single target.

2.6 EEG Data Acquisition

In the RSVP condition, EEG data are collected using BioSemi ActiveTwo system (BioSemi, Amsterdam, Netherlands). The 32-channel EEG system is used in experiment #$1-3$ and the 64-channel EEG system is used in experiment #5. Figure 2.3 shows the 32 and 64 channel labels and channel locations. The standardized arrangement of electrodes over

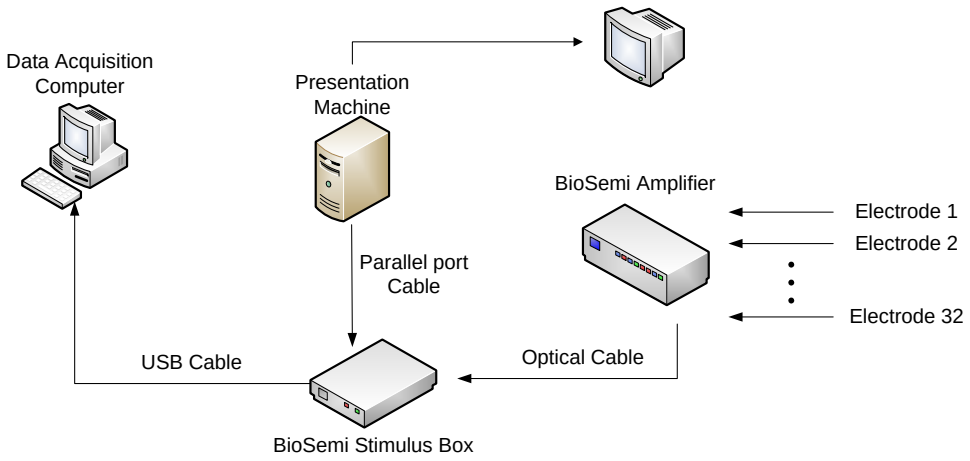


Figure 2.4: Data Collection Scheme. One computer is for image presentation and one computer is for EEG data acquisition.

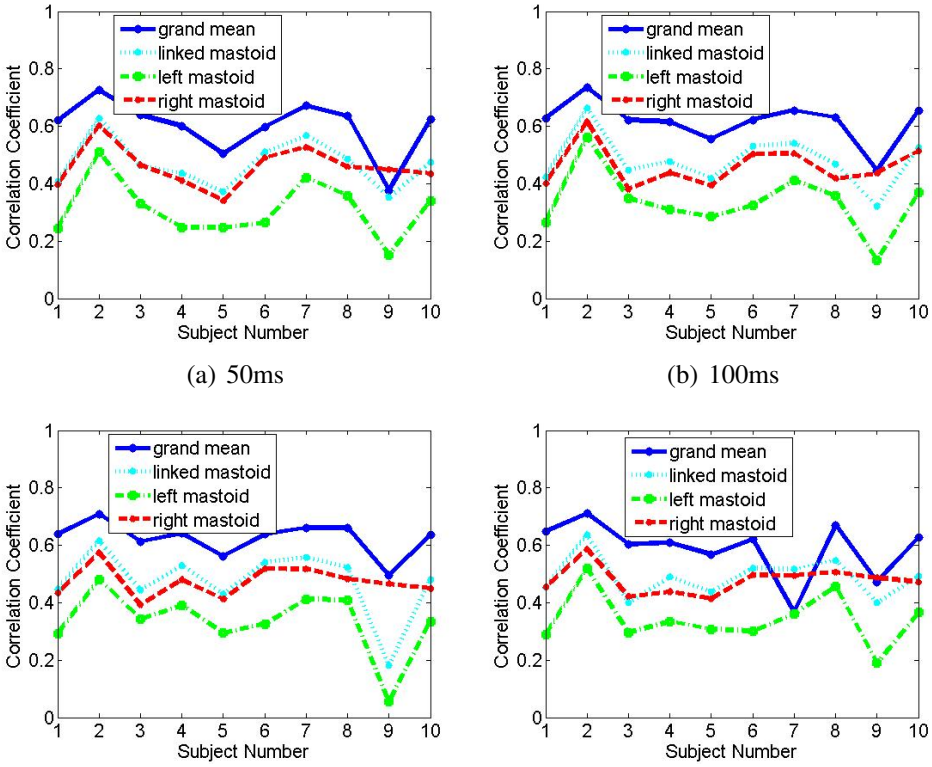
the scalp ensures ample coverage of all parts of the head. Subjects are seated 22 inches from screen. EEG signals are recorded using two computers, one for image display and one for data collection as shown in Figure 2.4. The triggers generated by the Presentation script to mark the onset of target and distractor stimuli are received by the BioSemi system over a parallel port and recorded concurrently with the EEG signals. The user's button presses, indicating the presence of perceived targets, are also recorded by the Biosemi system. The data are sampled at 256 Hz. All channel input impedance is less than $15k\Omega$.

2.7 EEG Data Preparation

2.7.1 Filtering, Calibration and Artifact Removal

EEG signals are bandpass-filtered between $1 - 45$ Hz by a $6th$-order Butterworth filter to correct DC drift and limit the effects of 60 Hz electrical line noise. Due to the gain in Biosemi recording system, consisting of the amplifiers and A/D converters, factor calibration is performed on all channels by multiplying a constant 0.0312 (suggested by Biosemi).

Before the experiment, subjects are requested an eye movement calibration experiment during which they are instructed to blink repeatedly by looking at the fixation cross



(a) 50ms (b) 100ms

(c) 150ms (d) 200ms

Figure 2.5: Comparison of the four referential methods – grand mean reference, linked mastoid reference, left mastoid reference and right mastoid reference. The evaluation is conducted on naive dataset #3. The correlation between the 64 scalp channels and the reference channel is computed. The correlation coefficients are plotted as a function of subject ID. The plots shows the correlation results for four RSVP sessions – (a) $50ms$, (b) $100ms$, (c) $150ms$ and (d) $200ms$. The correlation using the left mastoid reference (green dashdot line) is consistently much lower than the correlation using the other three reference methods across 10 subjects and four RSVP sessions.

in the center of the screen. Channels with bad connections are removed by inspection. Trials with strong eye movements or other movement artifacts are manually removed by inspection. No artifact removal for eye blink is applied because our experimental designs have very short burst in each block (sequence).

2.7.2 Referential Recordings

We apply referential recordings for each active electrode. Referential recordings are recommended for ERP studies [97]. Biosemi ActiveTwo stores a monopolar signal (channel). By subtracting an EEG reference (re-referencing), the common mode interference is reduced and the signal to noise ratio is increased. There are a variety of referential options. Four common referential methods – grand mean reference, linked mastoid reference, left mastoid reference and right mastoid reference – are compared using naive dataset #3. The grand mean reference use the mean of all 64 channels as a reference. The linked mastoid reference use the average of left and right mastoid channels as a reference. The left/right mastoid reference use the left (Tp7) / right (Tp8) channel as a reference. The reference subtraction off-line is performed for each channel to compute a referenced EEG signal.

To select the most uncorrelated reference of the four reference methods, we compute the correlation coefficients between each channel signal and each reference signal for each subject. The final correlation are reported as the mean correlation coefficients across all 64 channels and all trials. Figure 2.5 shows the comparison of the four reference methods. We measure the correlation between all scalp channels and the four different references. The left mastoid reference has the lowest correlation coefficient values across all subjects and all sessions. Therefore we select the left mastoid reference to re-reference each channel EEG signals.

2.7.3 Data Segmentation

In ERP detection study, raw EEG data are segmented into task-relevant epochs. Each epoch is corresponding to a single image and consists of a short segment of EEG (from the stimulus onset to $500ms$ after the stimulus onset). Each epoch is associated with a image chip. Based on the literature, the peak latency of the recognition-related ERPs varies from $250ms$ to $600ms$ post-stimuli depending on stimulus and subject parameters

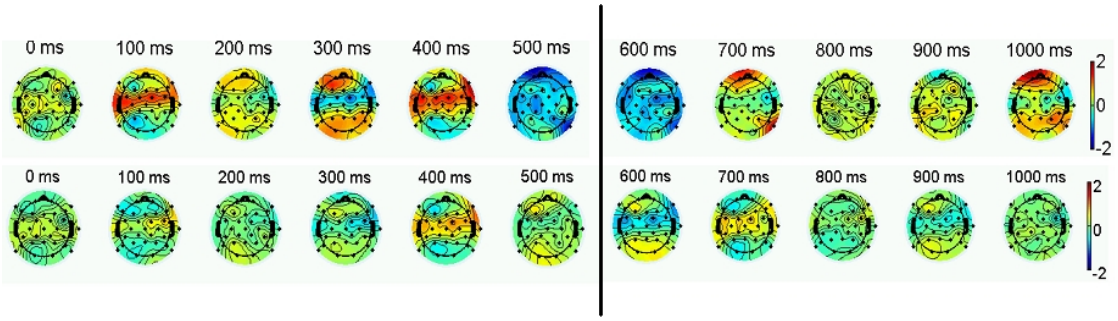


Figure 2.6: Averaged ERP scalp distributions for a subject at $100ms$ intervals following target (top row) and distracter (bottom row) stimulus onsets. Notice that the spatiotemporal activity changes for the target trials and does not change for the distractor trials.

[84]. Figure 2.6 shows the electrical activity over the scalp as a function of time. One clearly discernible feature is a peak in the trial-averaged activity around $300ms$ when target stimuli are present, whereas no amplitude change occurs for distractor stimuli. There are some amplitude changes around $700ms$, which are believed to be mainly caused by motor responses (button click activities). Based on our study on the same set of stimuli, the averaged button response time (RT) across subjects is in the range of 500 to 600ms with RT variance 20 to 50ms depending on the stimuli difficulty for the RSVP rate of 100ms/ image. To extract the neurally-relevant portion of the ERPs (and to avoid EEG signals associated with motor responses, which are presumed to be unavailable in many practical applications), we use only the EEG signals from the stimulus onset to $500ms$ post stimulus. The segment from $100ms$ before the stimulus onset to the stimulus onset is used to normalize the EEG data to have zero mean and unit variance in each channel.

2.7.4 Data Pre-processing

There are three steps of preprocessing for the training data. First, we remove targets that do not have an associated button click within $1.5s$ of the stimulus onset (many studies in the literature have reported that the reaction time of a button press ranges from $300 - 500ms$ depending on the complexity of the visual discrimination task [114]). The

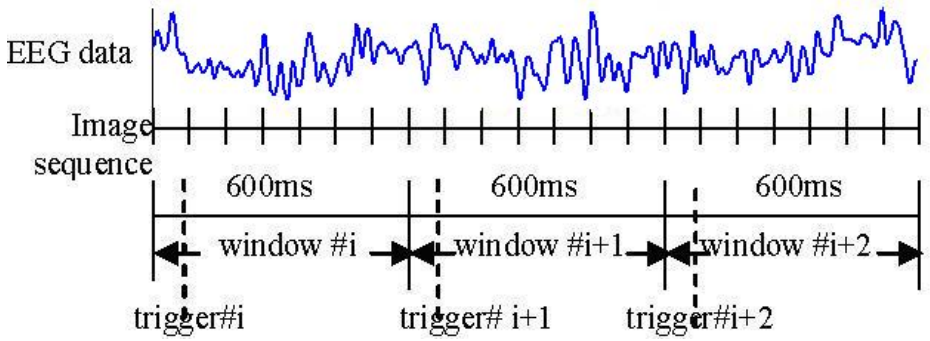


Figure 2.7: Disjoint windowing scheme of continuous EEG data, which is used to extract non-overlapping training samples. Each disjoint window is $600ms$, $100ms$ before the image trigger and $500ms$ after the trigger onset.

assumption is that there is no ERP if there is no button press. Second, we omit all distractor samples within one second before or after the target stimuli so that the distractor windows are not contaminated by ERP leakage (hence, the windows are not contiguous when they are near target stimuli). Third, we create non-overlapping (*disjoint*) windows of EEG activity corresponding to distracters as shown in Figure 2.7. The *disjoint* windowing scheme reduces the temporal correlation between data from two different windows. Each window is $600ms$ in length and extend from $100ms$ before the trigger (normalized window) to $500ms$ after the trigger (epoch window). In non-overlapping windowing scheme, we discard distractor samples occurring within the $600ms$ window.

For test data, on the other hand, we use overlapping windows because we do not know the class label in advance. The 32-channel data in each $500ms$ epoch are eventually concatenated to form a feature vector (32×129 dimension). The processed EEG measurements are then subjected to the classifiers.

Chapter 3

Neural Correlates of Human Visual Perceptual Processes in RSVP Paradigms

3.1 Introduction

Unveiling the neural correlates of visual detection and its underlying component processes is challenging due to the complexity and massively parallel interactions in the human brain. Humans can recognize an object within a fraction of a second. The neural mechanisms underlying this remarkable ability are not fully understood. The human visual information processing includes sensory image analysis, feature extraction, information fusion, decision making and motor response. Feature extraction and information about the object of interest (target) are fused around 100ms to 200ms. There are limits on how much a human brain can process. These limits are generally considered in terms of attentional and cognitive processing resources that can be allocated to different tasks. After the brain recruits the relevant attention and other neuronal resources to accumulate the information, the brain will ultimately determine a decision around 300ms and then control a explicit motor response.

Allocation of attention and cognitive processing resources is a fundamental determinant of performance in most perceptual tasks. With a few exceptions, attentional allocation has been measured by behavioral performance on single trials well separated in time. However behavioral studies do not have high temporal resolution of the spatial distribution of attention, making it hard to determine the timing of rapidly shift attention. In contrast, the ERP methodology provides high resolution measures of the time course of cognitive processing and thus can achieve greater precision than behavioral techniques alone. It provides a second window into attention and cognitive resources allocation in complex informational transactions of the human visual perceptual system. It can reflect the timing, ordering and interactions of the intermediate processes that are engaged in specific cognitive activities and lead to inferences about the mechanisms of parallel, serial and hierarchical processing of the system.

Much of the current research on ERPs has been dedicated to assess specific aspects and stages of human perceptual processing, such as encoding, selecting, memorizing and decision making. Huk [45] investigated neural activity in Macaque parietal cortex and found out that it reflected temporal integration of visual motion signals during perceptual decision making. They showed how the brain accrued and held the sensory signals to guide later actions. Shall [112] examined neural basis of saccade target selection during visual search and demonstrated a mechanism that conspicuous stimuli in complex scenes might attract gaze. Recently Philiastide and his group [95] examined temporal characterization of the neural correlates of perceptual decision making in the human brain. They used single-trial ERP analysis to demonstrate a temporal evolution of component activity indicative of an evidence accumulation process, which began after early visual perception and had a processing time that depended on the strength of the evidence. By pushing the visual system to its temporal limits using RSVP paradigm, recently researchers gained valuable insights into the mechanisms of visual processing using neural signals [122, 58]. Thorpe and his colleagues [122] designed a nature-scene target detection task to assess the speed of visual processing using the RSVP and the ERP. By comparing the difference

between the ERPs associated with targets and the ERPs associated with non-targets in an animal detection task, they showed that there was very sharp diverge at around $150ms$ at frontal sites, which implied a great deal of visual precessing completed before this time. They then concluded that the visual processing needed to perform the highly demanding visual task could be achieved in under $150ms$. Keysers [58] used single-unit recordings and RSVP paradigm in macaques and showed that face-selective activity could be elicited by masked images presented for as little as 14ms. They used functional magnetic resonance imaging on humans and showed that activations in object recognition areas were as brief as 40 ms. Chun [16] proposed a two-stage model for perceptual detection and examined temporal limitations of search (the attentional blink) for multiple targets in an RSVP display paradigm.

Recently the assessment of the covert attentional processes have been intensively studied directed toward the associated particular ERP components. There are a variety of different components showing attention effects in visual selection. Information about different stimulus features can be available to attentional mechanisms at different times depending upon stimulus and task requirements. For example, in a recent study, Philiastides et al. investigated how the decision difficulty affected the allocation of neural resources and the timing of cortical processing using the face/car and the red/green stimuli in a cued paradigm [94]. They found an ERP component, occurring approximately 220 ms after the onset of the stimulus, the peak value of which was correlated with behavioral performance and stimulus difficulty. Their results showed that the early component $N170$ was not directly linked to the actual decision and was initial evaluation of the evidence. The late component, appearing around $300 - 450ms$, was closely linked to decision and highly predictive of behavioral accuracy. The difficulty component, between the early and late components, predicted the onset of the late component and implied the recruitment of relevant attention for a difficulty decision. Hillyard [40] pointed out that most of the earlier ERP components varied as a function of physical stimulus parameters and were relatively insensitive to changes in perceptual processing demands. In contrast, the

longer latency components appeared in conjunction with specific perceptual or cognitive processes. A broad negative ERP (latency $150-300$ms) was elicited by the attended stimulus. While the early negative ERPs gave some information on the attentional channel, the $P300$ seemed to be a particularly sensitive index of the degree of attention received by deviant stimuli in and out of the designated focus of attention. VanRullen [125] continued Thorpe's work [122] and studied the timing of perceptual processing on animal and transportation detection tasks. Their results demonstrated that the perceptual activity occurring at $75ms$ was not accessible to awareness and the one occurring at $150ms$ was decision-related activity. Johnson et al. [52] studied the human visual system by measuring the time course of neural correlates of object recognition using animal and natural stimuli in cued-target paradigm. By subtracting the ERPs associated with targets from the ERPs associated with disctractors, there was a diverge at around 135ms. They found this early ERP component arising when there were low level feature differences between images. Unlike the early component, a late ERP component, arising around $150-300$ms, covaried with the subsequent reaction time and was correlated to recognition. They argued the timing of object recognition was after $150ms$, which was conflict with Thorpe and VanRullen's findings. Picton reported that when the auditory target stimuli became difficult, the amplitude of the $P300$ wave became small and its latency became longer [96]. For the study of the time course of visual processing from early perception to decision making, Johnson's group [53] showed that image stimulus difficulty had the same effects on the amplitude and onset of $P300$ in the visual cued-target paradigms as Picton's finding. Hillyard [39] used ERP and PET to measure the timing and location of visual selective attention. They found that $C1$ occurring at $50-90ms$, $P1$ occurring at $80-130ms$ and $N1$ occurring at $140-200ms$ in study of spatial attention mechanism. For timing of feature selection, they showed that $P1$ and $N1$ components were associated with selection of attended spatial location. The selection negativity component SN, occurring at $150-300ms$, was associated with selection of relevant features. The $N2$ and late positive component LPC, occurring at $250-300$ms were related to disclamation of

targets. They concluded that attention to non-spatial features such as color, motion and shape was manifested by different ERP patterns beginning with latencies of $100-150$ms. The discrimination of infrequent change in speed of movement requiring a manual response was associated with an negative $N2$ component and late positive component. The ERP preceded the motor response times by several hundreds milliseconds. Woodman and Luck [137] investigated ERP measurements of rapid shifts of attention during visual search and found that negative $N2pc$ component around $200-300$ms is related to the covert orienting of visual attention before the object recognition.

In contrast to the previous studies, this thesis uses the ERP as a physiological marker to examine the human visual perception and cognition in a dynamic recognition task. Our target stimuli are rare as opposed to categorizing two equally-prevalent targets. The targets are realistic target (SAM sites) as opposed to faces and cars. We follow each stimulus with a patterned mask to disrupt the accumulation of information from the retinal image representation. To simulate saccadic eye movements, we display visual stimuli at a dynamic RSVP paradigm. In contrast to sequences of well separated trials in experimental settings, the perceptual system in natural situations is confronted by rapidly changing stimuli arising from the dynamics of the environment, e.g. scenery, combined with the organism's own actions, e.g., eye movements. The main difference is that they control only the target difficulty, whereas we control both the target difficulty and the task difficulty. Our experimental design allows us to estimate how physiological and psychophysical performance depends on both the target and task difficulty.

The goal is to assess the dynamics of resource management, such as attention allocation, using brain signatures (ERPs) involved in human visual information system. The question is how attention is allocated if the new stimulus arrives before the observer finishes the analysis of the preceding stimulus? In this study we examine the dynamics of visual detection processes allocation as a function of both task requirements (image duration) and stimulus (target) *difficulty* on realistic target detection tasks. The hypothesis of our research is that there is high correlation between the underlying psychological aspects

of the ERPs and the dynamics of visual detection processes allocation. In this chapter we focus on the ERP characterization correlated with the human information processing. We conduct two experiments. The first experiment is the detectability study to assess the difficulty of realistic stimuli and subject performance using psychological functions. The second experiment is the neural correlate study of human information processing using the ERP temporal characterization on a real target detection task.

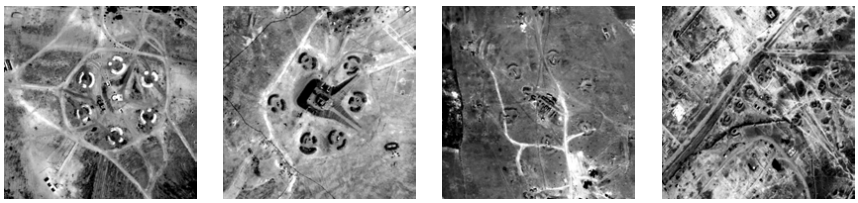
3.2 Experiment 1: Psychometric Function for Stimulus Difficulty and Subject Performance

This experiment is to quantify stimulus difficulty for complex real natural images and subject performance. The stimuli vary in detectability, which is estimated as the minimum exposure duration necessary for subjects to detect a specified target correctly 75% of the time. The detection rate is the probability of correct behavior responses (hits). The image duration threshold is defined as the 75% detection rate of behavior performance. The stimulus difficulty and subject performance are assessed by image duration thresholds using psychometric functions. This study is based on behavioral responses only.

3.2.1 Materials and Methods

Subjects

A total of ten naive subjects participated in the study as described in Chapter 2 Group #4, naive dataset #2. The task is to detect the presence of rare targets in the RSVP paradigm. The subjects perform target detection by clicking on a button as soon as they see a target. The behavior responses are collected.



(a) Easy (b) Median easy (c) Median hard (d) Hard

Figure 3.1: Target image examples – surface-to-air missile (SAM) sites at four difficulty levels – easy, medium easy, medium hard, hard. Target chips are encountered rarely.

Stimuli

The stimuli are real targets in the gray-scale satellite imagery. Specifically they are 500×500 pixel images of 105 different SAM sites. The size of the targets are small, and the scale, orientation, and position of the targets naturally are varied. Four target image examples at different target difficulty levels (easy, medium easy, medium hard, hard) are illustrated in Figure 3.1.

Image Display Paradigm

The image chips are presented using the RSVP paradigm. The chips are presented on a 21 inch CRT monitor using Presentation software (Neurobehavioral Systems, Albany, CA). Figure 3.2 shows the RSVP paradigm in this study. Images are presented in short bursts of approximately one second duration. Patterned masks are shown after every image to disrupt the accumulation of information from the retinal image representation. To break monotony and minimize possible eye strain, consecutive blocks are separated by a fixation screen of user-controlled duration. For each subject, there are five test sessions. The inter-stimulus-interval (ISI) are 25, 50, 100, 150, and 200 for five session respectively. Each session consists 210 blocks, each of which lasts for one second. One half of the blocks contain a single target.

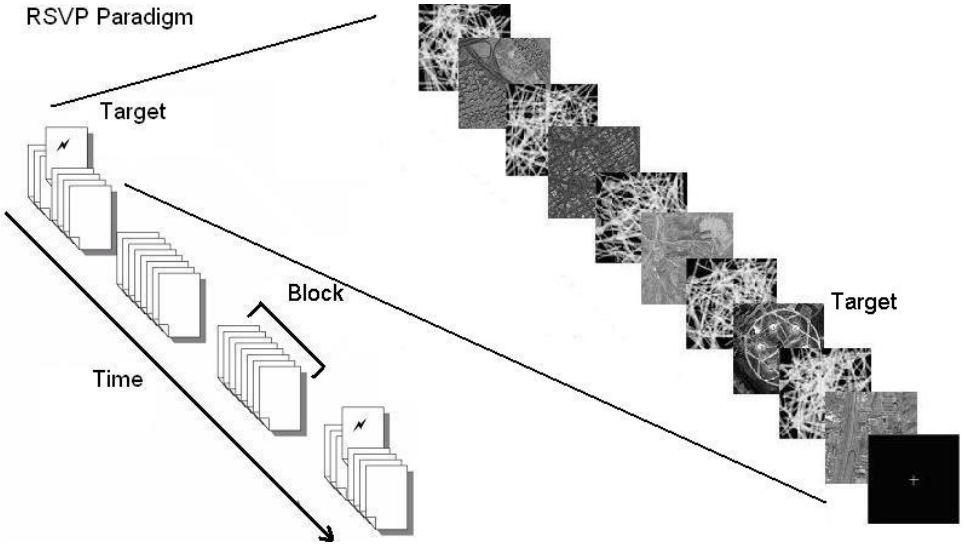


Figure 3.2: An illustration of the RSVP paradigm in this study. Each RSVP block starts with a fixation screen and each image (a target or a distractor) is followed by a single pattered mask. One half of the blocks contain a single target. During each RSVP block a sequence of images are displayed at the rates of 25, 50, 100, 150, and 200 ms / image.

Psychometric Function

To quantify the stimulus difficulty and subject performance, we construct psychological functions to characterize subject behavior performance. We apply a generalized linear models (GLM) [77] to fit behavior responses. The GLM was developed by Nelder and Wedderburn [81] for unifying various statistical models. There are three components in the GLM – random component, systematic component and link function. The random component is the probability distribution of the response variable. The systematic component is a linear function of the independent variables, called the linear predictors. The complete GLM is constructed through a relationship that is assumed between the distribution mean and the linear predictor. In general, the relationship is determined by a link function. For a random response variable $\mathbf{Y}$, the mean of the distribution is $\mu = E(\mathbf{Y})$. The link function is to transform the expectation value of the random component to the systematic component,

$$g(\mu) = \mathbf{X}'\beta, \tag{3.1}$$

where $\mathbf{X}$ is the systematic component, $g(\cdot)$ is the monotonic link function and β are unknown coefficients, which typically estimated with maximum likelihood (ML). The general GLM is,

$$\mu = g^{-1}(\mathbf{X}'\beta). \tag{3.2}$$

Traditionally, the random component is a member of an exponential family, such as Gaussian, binomial, or Poisson distribution. Here we use the binomial distribution because the behavior responses are dichotomous. For the ith systematic condition, the binomial distribution is an exponential family,

$$\begin{aligned} f(y_i; n_i, \mu_i) &= \binom{n_i}{y_i} {\mu_i}^{y_i} (1-\mu_i)^{n_i - y_i} \\ &= \exp\{y_i \ln[\frac{\mu_i}{1-\mu_i}] + n_i \ln[1-\mu_i] + \ln \binom{n_i}{y_i}\}, \end{aligned} \tag{3.3}$$

where n_i is the total number of trials and random variable $y_i \in \{0, 1, 2, ...n_i\}$ is the number of correct behavior responses (hits), and $\mu_i = E(y_i)/n_i$ is the probability of hits (detection rate). One of the important characteristics of the exponential family is the mean of the distribution is related to the natural location parameter, η,

$$\eta = \ln[\frac{\mu_i}{1-\mu_i}]. \tag{3.4}$$

The canonical link for binomial family is the *logit* link, which arises naturally from the general exponential formula (3.3) of binomial distribution. Pairing the binomial family with the logit link produces the logistic regression model. The *loglog* or *comploglog* link may be appropriate when the probability of the responses (detection rate) as a function of the linear predictor (image duration) approaches 0 and 1 asymmetrically. Three typical link functions – *logit*, *loglog comploglog* are as follows,

$$g(\mu) = \log(\frac{\mu}{(1-\mu)}), \tag{3.5}$$

$$g(\mu) = \log(-\log(\mu)), \tag{3.6}$$

$$g(\mu) = \log(-\log(1-\mu)). \tag{3.7}$$

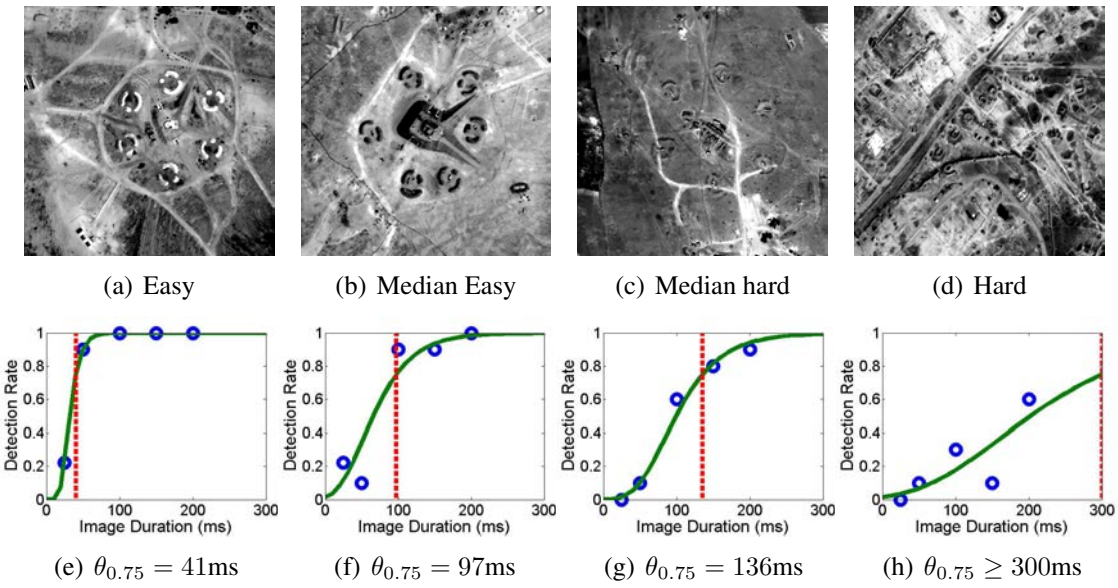
We use the loglog link function for the GLM in this study. The Matlab function *glmfit* is used to fit the GLM for behavior responses using binomial distribution and estimate the coefficients β using the ML. The function of *glmval* is used to compute predicted values for the GLM with the link function and predictor values.

3.2.2 Results

Target Difficulty Metric

We quantify the target difficulty for 105 targets using psychological functions. The target difficulty metric is defined as the image duration thresholds at the 75% detection rate of behavior performance averaged across 10 subjects. We measure the target difficulty as the image duration threshold for each target image using the computational model – the GLM in Section 3.2.1. For each target, The correct behavior responses are averaged across 10 subjects. The mean correct behavior responses at five different image duration are fitted using the GLMs. Figure 3.3 shows the four targets (easy, medium easy, medium hard and hard) and their corresponding psychological functions. These four targets are specifically selected at the median difficulty thresholds from four difficulty groups respectively. For each target, the difficulty thresholds are measured at the 75% detection rate. The image duration threshold values for these four images are 41ms, 97ms, 136ms and 300ms respectively.

Figure 3.4 shows target difficulty distribution in terms of the image duration threshold for 105 targets. The target difficulty thresholds are then sorted. The small thresholds are associated with easy targets and the large thresholds are associated with hard targets. The easier targets can be detected at 75% detection rate for shorter image durations whereas



(a) Easy (b) Median Easy (c) Median hard (d) Hard

(e) $\theta_{0.75} = 41$ms (f) $\theta_{0.75} = 97$ms (g) $\theta_{0.75} = 136$ms (h) $\theta_{0.75} \geq 300$ms

Figure 3.3: Psychometric functions of four target examples. The four targets (top row) are selected from four difficulty groups (easy, medium easy, medium hard and hard). For each image, the averaged correct behavior responses at different image duration are fitted with the GLM (bottom row). The blue circles are the correct response for each image duration. The green curve is the GLM fit with the binomial distribution and the loglog link function. The red dash line shows the image duration threshold at the 75% detection rate. The target difficulty metric is measured as the image duration threshold at the 75% detection rate of behavior performance. The target difficulty in terms of image duration thresholds for the four targets are $41ms$, $97ms$, $136ms$, $300ms$ respectively.

the harder targets can be detected at the 75% detection rate for longer image durations. It is clear that the harder images require more time to process to achieve the same correct detection level than the easier ones. Based on the target distribution, We divide the 105 targets into four groups – easy (26 targets), medium easy (26 targets), medium hard (26 targets) and hard (27 targets) based on the image duration threshold values on the psychological functions.

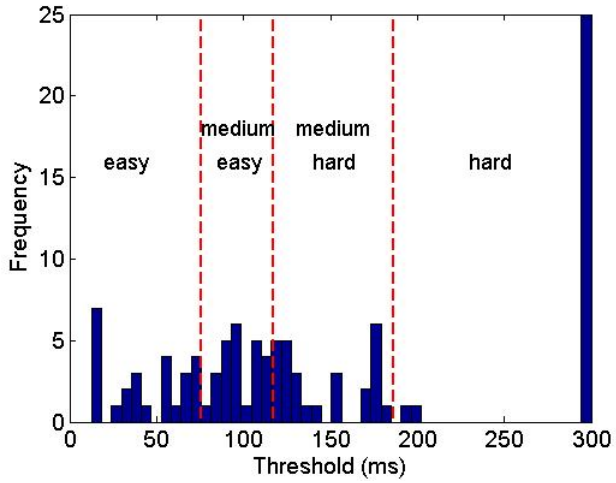
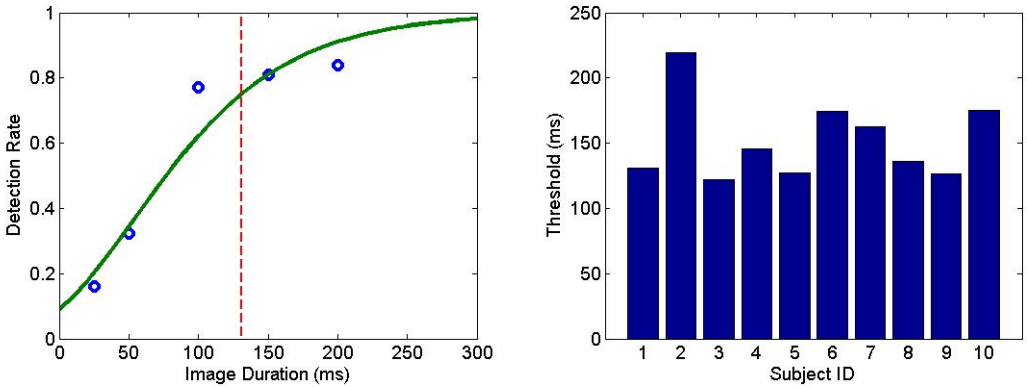


Figure 3.4: Target difficulty distribution in terms of the image duration thresholds for 105 targets. The image duration threshold is defined as the 75% detection rate of behavior performance. The target difficulty thresholds are sorted. The small thresholds are associated with easy targets and the large thresholds are associated with hard targets. The total 105 targets are divided into four groups – easy, medium easy, medium hard and hard based on the threshold. The number of targets in the four groups are $26, 26, 26, 27$ respectively

Subject Performance Metric

We use the psychological function to quantify subject performance by the image duration threshold for each subject. The subject performance metric is defined as the image duration thresholds at the 75% detection rate of behavior performance averaged across 105 targets. We measure subject performance as the image duration threshold for each subject using the GLM. There are five image durations – 25, 50, 100, 150, and 200 ms per image in the experiment. We average the correct behavior responses across 105 targets for each subject. The GLMs are used to fit the mean correct behavior responses at different image duration. Figure 3.5 shows Subject performance in terms of image duration thresholds. Figure 3.5(a) is the psychological function for subject 1. The image duration threshold at 75% detection rate for subject 1 is $131ms$, which is corresponding to the first histgram in Figure 3.5(b). Figure 3.5(b) is subject performance in terms of image duration threshold as a function of subject ID. The averaged image duration threshold is $152ms$ with standard deviation of $31ms$. The range of image duration threshold for ten subjects is from $122ms$ to $220ms$. The results show that most subjects achieve 75% detection rate at the



(a) $\theta_{0.75} = 131$ms (b) Subject performance

Figure 3.5: Subject performance in terms of image duration thresholds. The image duration threshold is defined as the 75% detection rate of behavior performance averaged across 105 targets. (a) Psychological function for subject 1. The blue circles are the correct response for each image duration. The green curve is the GLM fit with binomial distribution and loglog link function. The red dash line is the image duration threshold at the 75% detection rate. The image duration threshold for subject 1 is 131ms. (b) Subject performance in terms of image duration thresholds as a function of subject ID.

range of image duration threshold from $100ms$ to $250ms$, which suggests that the image duration from $100ms$ to $250ms$ is a good range for the task.

3.3 Experiment 2: Neural Correlates of Visual Perception

This experiment is to assess neural correlates of human visual information system on real target detection tasks in the RSVP paradigms . We examine the dynamics of attention allocation as a function of both stimulus (target) *difficulty* and task requirements (image duration). Target difficulty is defined by image duration thresholds and task difficulty is defined by image durations. Our experimental design allows us to estimate how behavioral and physiological performance depends on both target and task difficulty. We investigate the psychological aspects of the ERPs. Specifically we assess the effects of target difficulty and task difficulty on ERP characteristics and the timing of these effects on the ERP components. We use both behavioral and physiological data in this study.

3.3.1 Materials and Methods

EEG Data Acquisition and Pre-processing

Ten naive subjects (ages range from $25-45$; 4 females) participate in the study as group #6 in Chapter 2. The stimuli and procedures are the same as Section 3.2.1. The subjects perform the same target detection tasks as the experiment 1 by pushing the space bar as soon as they see a target. Both behavioral and physiological data are collected.

The same target chips (105 SAM sites) as shown in Figure 3.1 are used in this study. The stimuli are presented to the subjects using the RSVP paradigm as Figure 3.2. The stimuli are infrequent targets embedded in sequences of distractor images. Each target and distractor image is followed by a single patterned mask. There are four test sessions per subject, one for each image duration (50, 100, 150, and 200 ms). Each session consists of 210 blocks, each of which lasts for one second. One half of the blocks contain a single target. There is no feedback.

The 64-channel EEG data are sampled at 256 Hz. The data are bandpass filtered between $1-45$ Hz. and performed factor calibration on all channels. We perform factor calibration artifact removal and offline re-referencing using left mastoid reference. The detailed EEG acquisition and pre-processing procedures are described in Section 2.6 and Section 2.7. The bad connection channels are manually removed for each subject (channel#28, channel #1, #33 and #34, channel #58 and #63, channel#15, channel#57, channel #16, #62, #27 and #63 for subject3, 5, 6, 7, 8 and 9).

Target Difficulty Metric

Target difficulty metric is assessed by the image duration thresholds based on the detectability study as discussed in Section 3.2.1. The image duration thresholds estimated from the group #6 are used to divide the total of 105 targets into four difficulty groups – easy, medium easy, medium hard and hard targets – as shown in Figure 3.4. We use the target image duration thresholds based on Group #6 throughout this study.

To examine the dynamics of neural signatures as a function of target difficulty, we compute the grand averaged ERPs for the hit trials over ten subjects at each target difficulty level for each image duration. To visualize the spacial extent of the ERP activity across time, we compute the averaged ERP scalp maps by interpolating the ERP activity across all electrode locations. All scalp maps are plotted using EEGLAB [22].

Task Difficulty Metric

Task difficulty metric is the image durations in the RSVP condition. We have four image durations (RSVP sessions) – 50ms, 100ms, 150ms, and 200ms, which are associated with hard, medium hard, medium easy and easy tasks.

To examine the dynamics of neural signatures as a function of task difficulty, we compute the grand averaged ERPs for the hit trials over ten subjects at each task difficulty condition for each target difficulty level. We also use ERP scalp maps to visualize the spacial ERP activity.

Early and Late Difficulty Component Correlates

To determine whether the neural signatures – an early component and a late component – are indeed associated with target difficulty or task difficulty and determine the timing of these components, we perform a correlation study on ERP amplitudes with target and task difficulty metrics. For target difficulty, the correlation is measured by the ERP amplitudes and the target difficulty in terms of image duration thresholds for each image duration. For task difficulty, the correlation is measured by the ERP amplitudes and the task difficulty in terms of image durations for each target difficulty level.

We use the traditional ERP analysis technique – trial averaging – across hit trials, across selected electrodes and across 10 subjects in this ERP correlation study. Based on the scalp image maps, we select frontal and centrofrontal sites in this study. The analysis temporal range is 40ms to 500ms after the stimulus onset. We average ERP activity across short-length non-overlapping temporal windows (typically 20ms in width). For

each short-length temporal window, the stimulus locked time is the median of the window. To make our estimates more robust, we apply overlapping windows (typically 40ms in width) and shift 10ms each time on image duration thresholds to obtain small ranges of target difficulty. The mean image duration thresholds are measured. For each image duration, the averaged ERP is computed across fontal and centrofrontal sites, across trials associated with specific stimulus difficulty range and across 10 subjects. The ERP amplitude is measured by the mean amplitude of the averaged ERPs in each 20ms-length window. For target difficulty correlation analysis, we correlate the mean amplitudes at several time intervals with the target difficulty in terms of mean image duration thresholds. For task difficulty correlation analysis, we correlate the mean amplitudes at several time intervals with the task difficulty in terms of image durations.

Late Component Correlates using Single-trial ERP Detection

We use a single-trial analysis of the EEG to discriminate among the four target difficulty levels at each image duration. We apply linear support vector machine (SVM) [126] as the ERP detector and use the 10-fold cross-validation approach [26] on single-trial ERP detction. We quantify the single-trial ERP detection performance by the area under the receiver operator characteristic (ROC) curve, referred as AUC. To make our estimates more robust, we use short-length temporal window ($40ms$ in width) and the analysis temporal range is 170ms to 500ms after the stimulus onset. For each short-length temporal window, the stimulus locked time is the median of the window. Given a fixed window width, we apply the SVM classifier while sweeping the window from 170ms to 500ms after the stimulus onset for each subject and each RSVP image duration. The AUC metric is computed to characterize the single-trial ERP detection performance at several time intervals. For target difficulty correlation analysis, we correlate these AUC values from 10 subjects with the corresponding target difficulty levels for each image duration condition.

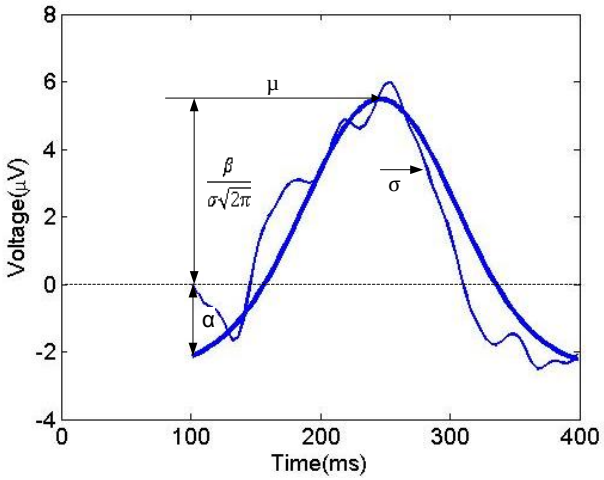


Figure 3.6: Illustration of Gaussian fitting procedure. Fit $\hat{y}(t)$ (solid curve) for component activity $y(t)$ (dash curve). Parameters of a fit include the following: ERP component peak magnitude ($\alpha + \frac{\beta}{\sigma\sqrt{2\pi}}$), peak latency relative to the onset of visual stimuli (μ) and component spread (σ).

Late Component Peak Detection

To quantify the ERP peak components – peak magnitude, peak duration and peak latency, we perform peak detection by fitting a parametric function to the spatially integrated discriminating component. For simplicity, we use a shifted Gaussian,

$$\hat{y}(t) = \alpha + \frac{\beta}{\sigma\sqrt{2\pi}} \exp^{-\frac{(t-\mu)^2}{2\sigma^2}} . \tag{3.8}$$

The Gaussian profile is parameterized by its height β, width σ, delay μ, and baseline offset α . We compute the optimal parameters by the least square error.

We analyze the EEG signals of two midline channels – Fz and Cz – at the time range of $100 - 400$ms after the stimuli. We apply the fitting procedures for two RSVP rates - 100ms and 200ms to investigate the extent to which the ERP components varies with the target difficulty and task difficulty. We align all ERP waveforms associated with different image durations to the same starting point and perform manual inspection for the fits.

Pairwise Correlates to Target and Task Difficulty

To investigate the effects of the combinational information from the target and task difficulty, which accounts for both the psychological and physiological responses, we conduct pairwise-correlation analysis. We measure the ERP waveform similarity as a function of behavioral performance level. The ERP waveform similarity is measured by the correlation between the ERP waveform pairs at the same behavioral performance. The analysis window is 0ms to 500ms after the stimulus onset.

We use four categories of behavioral performance, each of which contains a subset of the ERPs from 16 joint categories of four target difficulty and four task difficulty in Table 3.1. To obtain roughly equal size of numbers of members, the four behavioral performance levels are selected as the hit rate at range of $0.20-0.50$, $0.50-0.90$, $0.90-0.96$, $0.96-1$. To quantify the pairwise similarity, we measure the averaged pairwise correlation coefficients for all the combinations of the ERP waveform pairs in the set associated with a given behavioral performance.

To further confirm whether neural activity is strongly related with target difficulty, task difficulty independently or both, we measure the ERP waveform similarity in terms of pairwise correlation as a function of target difficulty, task difficulty and behavioral performance respectively. Here we define three classes – target difficulty, task difficulty and behavioral performance. Referring to Table 3.1, we have four levels for the target difficulty (each column in the table), four levels for the task difficulty (each row in the table), and four levels for the behavioral performance (the hit rate at range of $0.20-0.60$, $0.60-0.90$, $0.90-0.95$, $0.95-1$). We compute the within-class correlations, between-class correlations and all-class correlations for each class and compare the within-class correlations with between-class ones.

Relations to Behavioral Responses

To assess the relations of target and difficulty and behavioral responses, we measure the averaged behavioral response time (RT) and RT variance across 10 subjects at each target difficulty level at each task difficulty condition. To visualize the profile of the components (stimulus or response locked) across all trials, we construct discriminant maps. We align all trials of an experimental condition of interest to the onset of visual stimulation and sorted them by their corresponding RTs. Each row of the discriminant component map represents a single trial across time.

3.3.2 Results

Effect of Target Difficulty

We access the relation between the ERP components and target difficulty and examine the dynamics of the neural signatures as a function of target difficulty. The target difficulty metric is described in 3.3.1. We compute the grand averaged ERP at each image duration for each target difficulty condition. Figure 3.7 shows the averaged ERPs at four different target difficulty levels for $200ms$ image duration. Figure 3.7(a) and 3.7(b) illustrates the results for two sensors of interest (FCz and O2, respectively). One can see that there is an early negative component around 100ms after stimulus and a late positive component around 250ms after stimulus across target difficulty. In addition, we construct averaged ERP scalp maps to visualize the spatial distribution of these components. The scalp maps in Figure 3.7 (c) demonstrates these results. There are high negative activations appears around $100ms$ and high positive activations appear around 250ms after the stimulus onset for four target difficulty conditions. The activation decreases when stimuli become harder for both the early negative component and late positive component. The results indicates that the magnitudes of which are inversely proportional to the stimuli difficulty. The harder the target, the smaller the magnitude for both the early negative component and the late positive component.

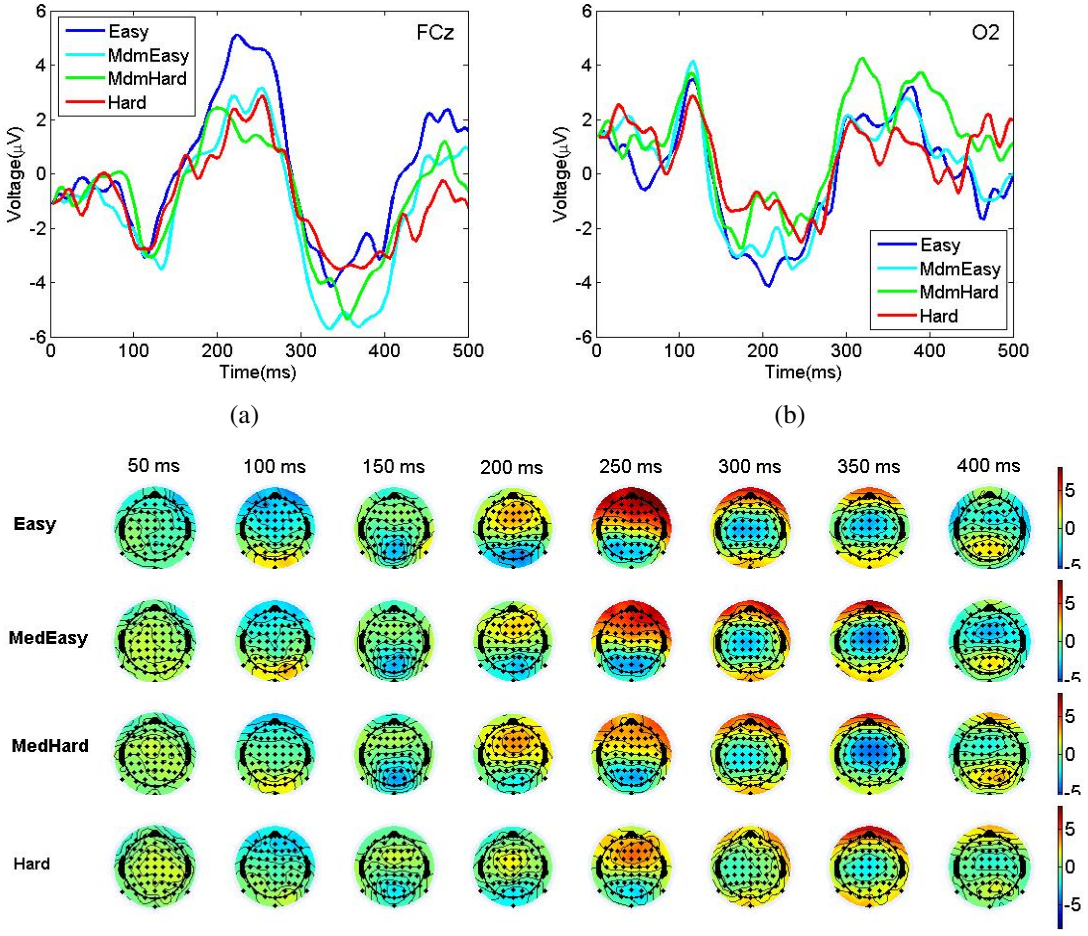
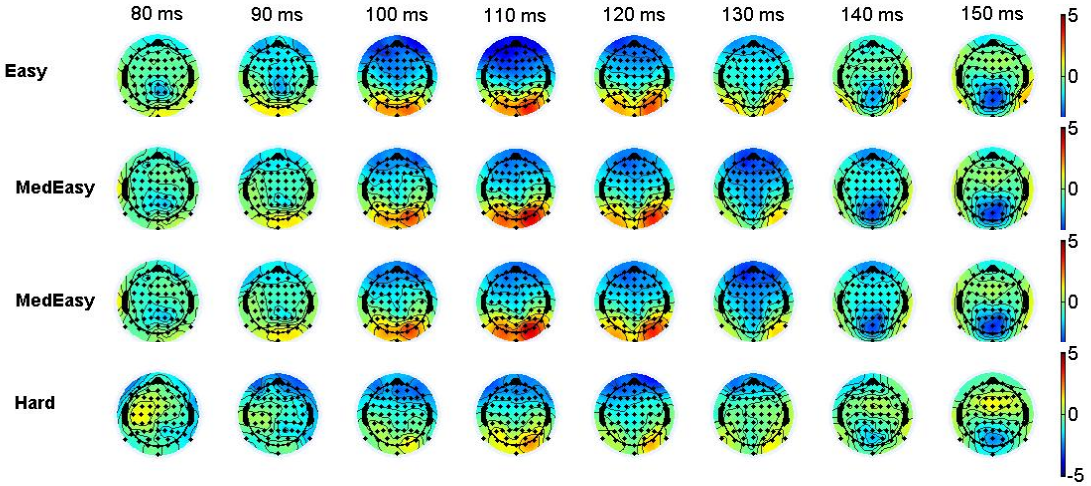
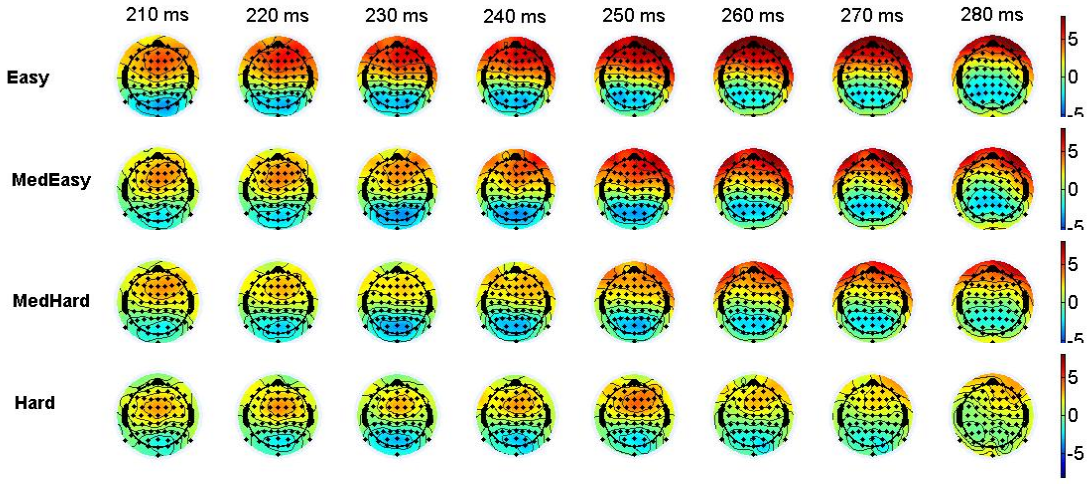


Figure 3.7: Averaged ERPs for target difficulty analysis. The ERPs are averaged across 10 subjects and all hit trials at 200ms image duration session. (a) Averaged ERPs for each target difficulty condition (easy, medium easy, medium hard and hard) for a centrofrontal site (FCz). (b) Averaged ERPs for an occipitoparietal electrode site (O2). (c) Averaged ERP scalp maps at 200ms image duration for four target difficulty levels (easy, medium easy, medium hard and hard). The scalp distribution of ERPs for hit trials are averaged across 10 subjects at eight specific time instants from $50ms$ to $400ms$ after the stimulus onset. Red corresponds to high positive activation and blue corresponds to high negative activation.



(a) Early Component



(b) Late Component

Figure 3.8: Averaged ERP scalp maps of early component and late component at 200ms image duration for four target difficulty levels (easy, medium easy, medium hard and hard). The ERPs are averaged across hit trials for 10 subjects. (a) Averaged ERP scalp maps of the negative peak at around 110ms for four target difficulty levels. The scalp distribution of averaged ERPs are measured at eight specific time instants from $80ms$ to $150ms$ after the stimulus onset. (b) Averaged ERP scalp maps of the positive peak at around 250ms for four target difficulty levels. The scalp distribution of averaged ERPs are measured at eight specific time instants from $210ms$ to $280ms$ after the stimulus onset.

To further determine the timing of the early and late components for target difficulty, we compute the ERP scalp maps centered around 120ms and 250ms for the early and the late components respectively. Figure 3.8 shows the results. In Figure 3.8 (a) the scalp maps for the early components show that the component is distributed over a range of electrode locations as indicated by the decreased activations at several centrofrontal (high negative activations) and occipitoparietal (high positive activations) sites for the harder target conditions. It is clear that the early component appears around 110 ms. One interesting finding is that the peak latency shifts forward in time (latency becomes larger) as targets become harder. From Figure 3.8 (a) it is noticed that the timing shift effect slightly deteriorates in the hard target difficulty condition. The reason could be the random guessing in the hard target difficulty condition and the averaged ERP at the hard target condition is noisy due to less hit trials in the hard condition. In Figure 3.8 (b) The scalp maps for the late components show that the amount of high positive activation is decreased as the target difficulty becomes progressively harder. It is clear that the late component appears around 250 ms. The peak latency for the late component also shifts forward in time (latency becomes larger) for as targets become progressively harder. There is also a similar slightly-deteriorated effect in the hard target difficulty condition.

Effect of Task Difficulty

We examine the dynamics of neural signatures and behavioral performance as a function of task difficulty in this study. Task difficulty is based on the image duration. We have four different task difficulty levels – $50ms$(hard), $100ms$(medium hard), $150ms$(medium easy), and $200ms$(hard). The ERPs for the hit trials are averaged over 10 subjects at each target difficulty condition for each image duration. Figure 3.9 shows the averaged ERPs at each of four different task difficulty levels for the easy target difficulty condition. Figure 3.9(a) and 3.9(b) illustrates the results for two sensors of interest (FCz and O2, respectively). One can see that there is an early negative component around 100ms after stimulus and a late positive component around 250ms after stimulus, the magnitudes of

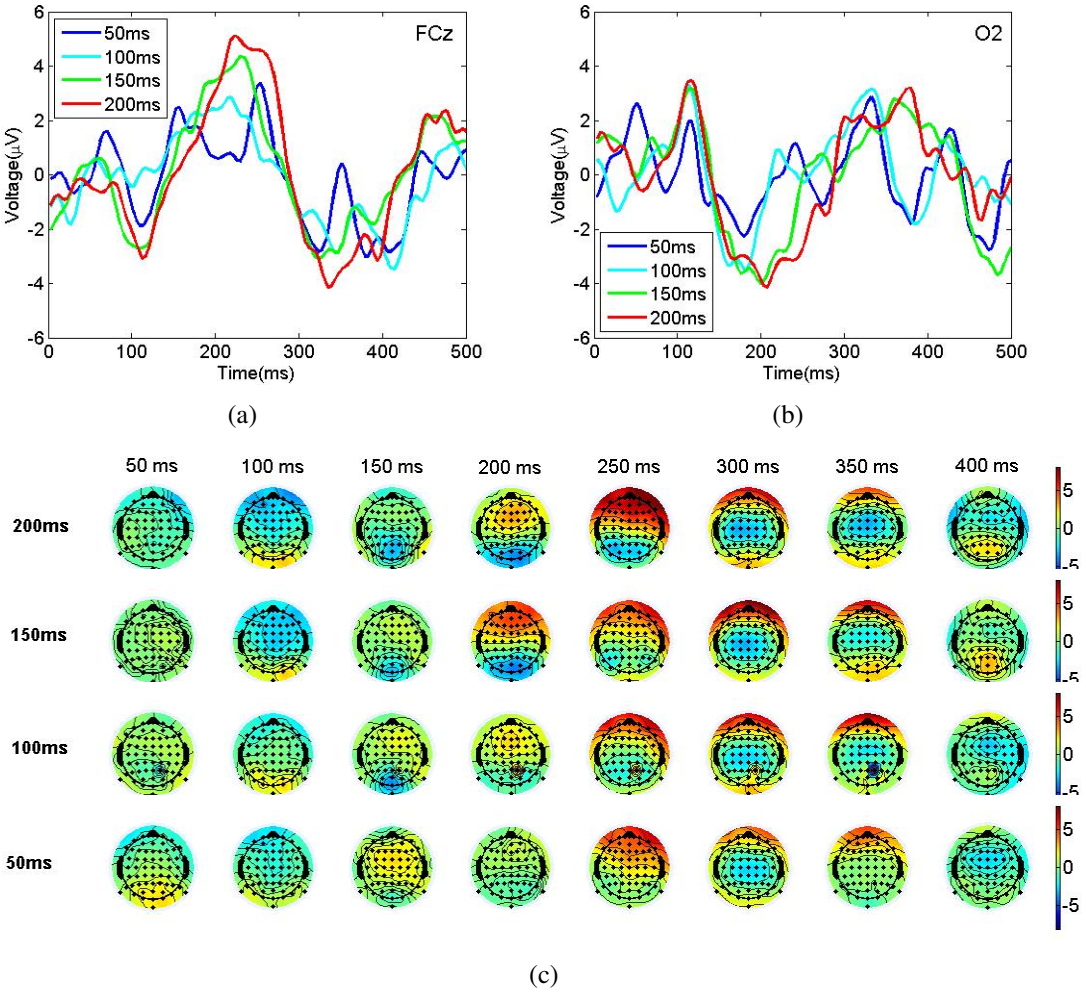


Figure 3.9: Averaged ERPs for task difficulty analysis. The ERPs are averaged across 10 subjects and all hit trials at easy target difficulty condition. (a) Averaged ERPs for each task difficulty condition in term of image duration ($200ms$, $150ms$, $100ms$ and $50ms$) for a centrofrontal site (FCz). (b) Averaged ERPs for an occipitoparietal electrode site (O2). (c) Averaged ERP scalp maps at easy target difficulty condition for four task difficulty levels ($200ms$, $150ms$, $100ms$ and $50ms$). The scalp distribution of ERPs for hit trials are averaged across 10 subjects at eight specific time instants from $50ms$ to $400ms$ after the stimulus onset.

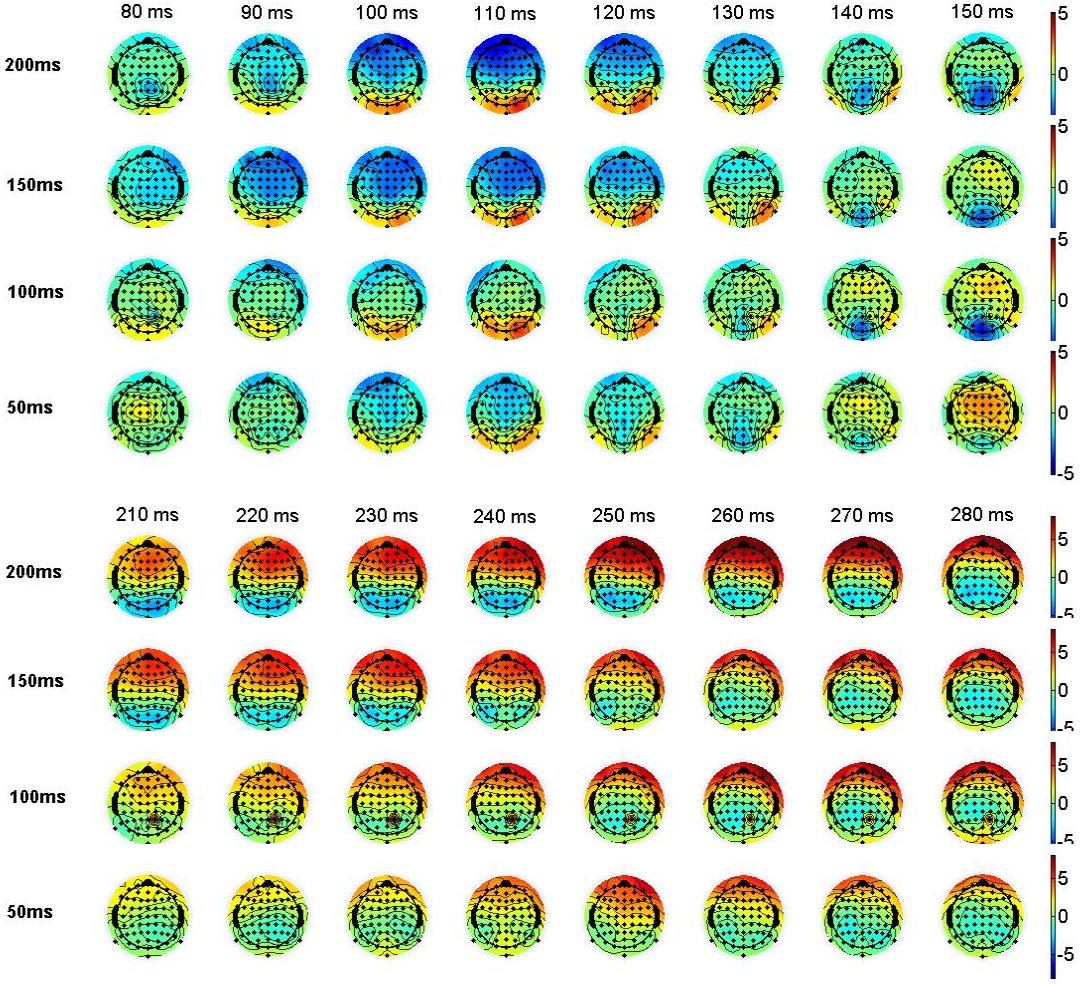


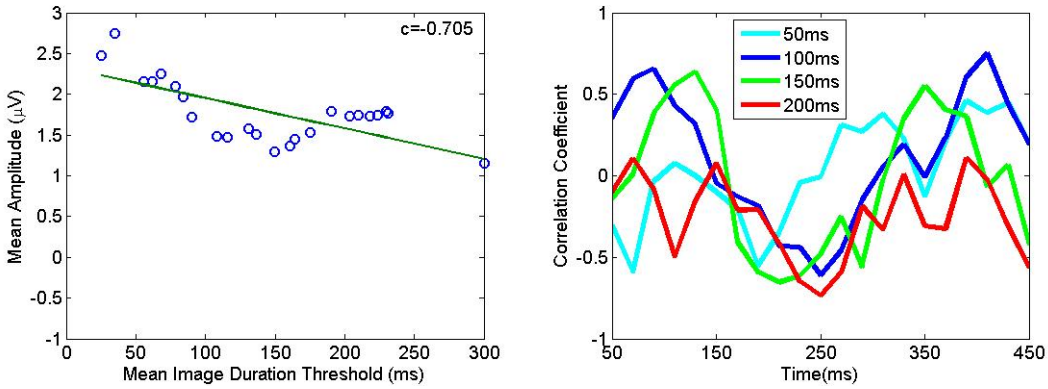
Figure 3.10: Averaged ERP scalp maps of early component and late component at easy target difficulty condition for four task difficulty levels ($200ms$, $150ms$, $100ms$ and $50ms$). The ERPs are averaged across hit trials for 10 subjects. (a) Averaged ERP scalp maps of the negative peak at around 110ms for four task difficulty levels. The scalp distribution of averaged ERPs are measured at eight specific time instants from $80ms$ to $150ms$ after the stimulus onset. (b) Averaged ERP scalp maps of the positive peak at around 250ms for four task difficulty levels. The scalp distribution of averaged ERPs are measured at eight specific time instants from $210ms$ to $280ms$ after the stimulus onset.

which are inversely proportional to the strength of the stimuli. The harder the task, the smaller the magnitudes of the early negative component and the late positive component. It is noticed that the ERPs at hard condition (50ms) is noisy. In addition, we construct averaged ERP scalp maps to visualize the spatial distribution of these components. The scalp maps in Figure 3.9 demonstrates these results. There are high negative activations appears around $100ms$ and high positive activations appear around 250ms after the stimulus onset for four task difficulty conditions. The trend of the activation decreases when stimuli become harder. This effect slightly deteriorates in the $50ms$ condition due to less hit trials to be averaged.

To further determine the timing of the early and late components for the task difficulty, we compute the ERP scalp maps centered around 120ms and 250ms for the early and the late components respectively. In Figure 3.10 (a) the scalp maps for the early components show that the component is distributed over a range of electrode locations as indicated by the decreased activations at several centrofrontal (high negative activations) and occipitoparietal (high positive activations) sites. The early component appears around 110 ms. There is much higher negative activation for the easy (200ms) and medium easy (150ms) task conditions comparing to the hard (50ms)and medium hard (150ms)task conditions. There is no peak latency shift effect associated with task difficulty. In Figure 3.10 (b) The scalp maps for the late components show that the amount of positive activation for the easy and medium easy task is much higher than that for the hard and medium hard task. The late component appears around 250 ms. There is no peak latency shift.

Early and Late Difficulty Component Correlates

To further investigate whether only the total amount of information from both task and target difficulty matters or task and target difficulty contribute independently to neural responses, we resort the physiology signals (EEG) to examine the timing of these effects. Specifically we examine correlate the amplitude of the early component and the late component with target difficulty in terms of image thresholds and task difficulty in terms of



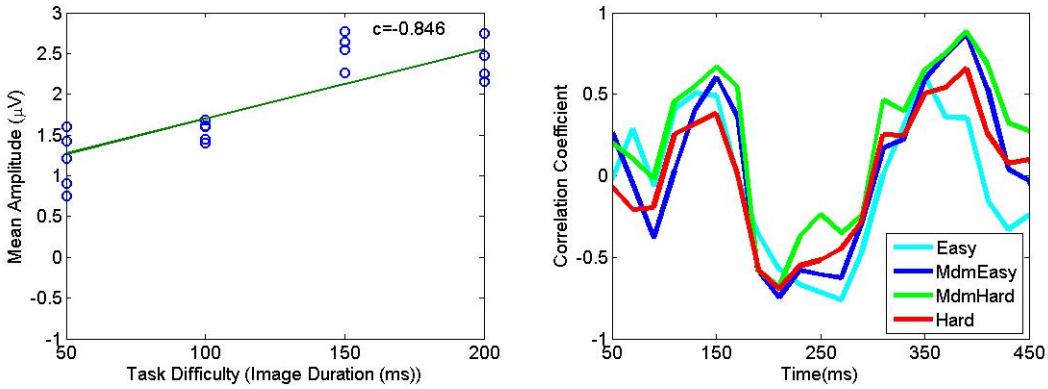
(a) 250ms latency at 200ms duration (b) Target difficulty correlation

Figure 3.11: The correlation between the target difficulty and the amplitude of the ERP component as a function of time. (a) ERP amplitude correlation of the late component at $250ms$ latency at $200ms$ image duration with correlation coefficient $c = -0.705$. (b)Target difficulty correlation as a function of time.

image durations as described in Section 3.3.1. The mean image duration thresholds are estimated in overlapping short target difficulty windows. The ERP amplitude is measured by the mean amplitude of the averaged ERPs in each 20ms-length temporal window. For the target difficulty correlation analysis, the mean amplitudes in each window are correlated with target difficulty in terms of the mean image duration thresholds. For the task difficulty correlation analysis, the mean amplitudes in each window are correlated with four task difficulty in terms of image durations.

Figure 3.11 shows the ERP components correlates to target difficulty. The correlation curve for the 50ms is noisy due to less trial to be averaged. The early component have positive correlates around $110ms$ across $100ms$ and $150ms$ image durations. It indicates that the amplitude of the early component increases as the target difficulty increases. The late component has consistent negative correlation around $250ms$ across the other three image durations. It indicates that the amplitude of the late component decreases as the target difficulty increases across image durations.

Figure 3.12 shows the ERP components correlates to task difficulty. Both the early and late components have consistent correlates around $150ms$ and around $250ms$ across different target difficulty conditions. The early component has positive correlations across



(a) 250ms latency for easy target (b) Task difficulty correlation

Figure 3.12: The correlation between the task difficulty and the amplitude of the ERP component as a function of time. (a) ERP amplitude correlation of the late component at $250ms$ at at easy target condition. (b)Task difficulty correlation as a function of time.

different target difficulty levels. The late component has highly significant negative correlations across different target difficulty levels. The task difficulty effect is consistent across different target conditions.

Late Component Correlates to Single-trial ERP Detection

To determine whether the neural signature around 250ms after stimulus – is indeed associated with target difficulty, we conduct a single-trial analysis of the ERP to distinguish four target difficulty levels as described in Section 3.3.1. We compare the correlation of target difficulty and single-trial ERP detection at each time intervals for four RSVP image durations for each subject. During each comparison, we apply the classifier at several time intervals and detect ERPs on single-trial basis. Each time we compute AUC values as a metric for the single-trial ERP detection performance. The correlation is measured between the target difficulty levels and AUCs for each subject. We find a significant correlation at around $270 - 310$ms after stimulus consistently across four image durations, as can be seen in Figure 3.13. The harder the targets, the lower AUCs for four image durations. It demonstrate that the late component is strongly related to the separation of the ERPs associated with target from the ERPs associated with distractors. This indicated that the late component is strongly correlated with target difficulty. In Figure 3.13 the

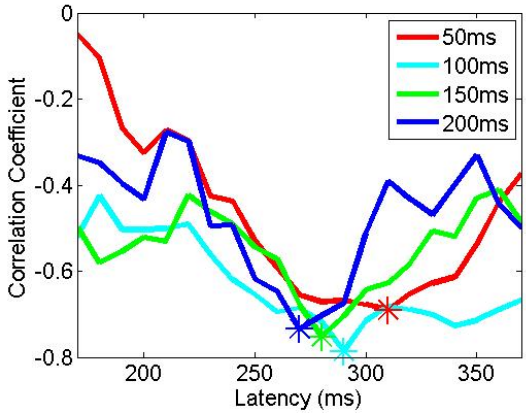
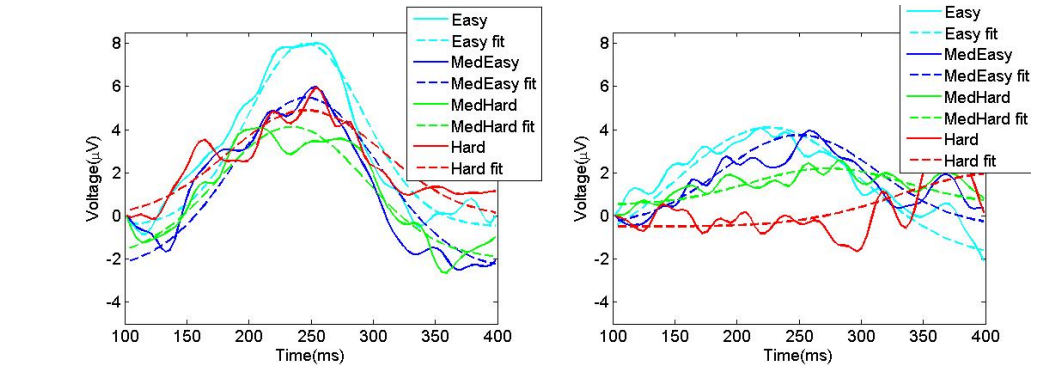


Figure 3.13: Correlation analysis between the single-trial ERP detection performance in terms of area under ROC curves (AUC) for easy-vs-hard discrimination. The correlation coefficients are plotted as a function of latency(time). A significant correlation between AUC and four target difficulty levels – easy, medium easy, medium hard and hard, can be seen in the time interval between $270 - 310$ms after stimulus (asterisks) across task difficulty conditions – $50ms$(hard), $100ms$(medium hard), $150ms$(medium easy), and $200ms$(hard). The time shift forward as the tasks become harder.

highest correlation points are marked as the asterisks for each image duration. It is clear that the latency shift forward (become larger) as the tasks become harder (image duration become shorter).

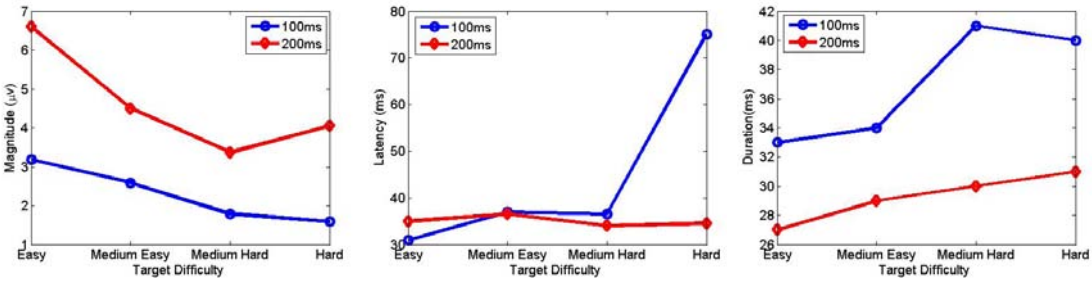
Discriminant Component Peak Detection

To quantify the characteristics of ERP peak components – peak magnitude, peak latency and peak duration, we perform peak detection by fitting a shifted Gaussian as describe in Section 3.3.1 to the spatially integrated discriminating component for the two RSVP rates - 100ms and 200ms. We analyze the relation between the ERP components and target and task difficulty. Figure 3.14 (a)(b) show the Gaussian profile fits of the late discriminating component at a fontal site for four target difficulty conditions at 100ms and 200ms image durations. Figure 3.14 (c) shows that Estimation of the late component peak magnitude, latency and duration as a function of target difficulty levels for two task difficulty conditions (100ms and 200ms) using Gaussian fits. Results shows that there is strong relation between the ERP peak components, and the target difficulty and task difficulty. The magnitude of the ERP decreases when target becomes harder for both image



(a) 200ms, Fz

(b) 100ms, Fz



(c) relations to target difficulty

Figure 3.14: Illustration of Gaussian peak fitting procedures at 200ms image duration (a) and 100ms image duration (b). The plots shows the Gaussian fittings (dash curves) for averaged component activity (solid curves) at electrode Fz for four target difficulty conditions (easy, medium easy, medium hard and hard). Parameters of each fit – peak magnitude, peak latency relative to the stimulus onset and component spread are described in 3.3.1. (c) Estimation of the late component peak magnitude, latency and duration as a function of target difficulty levels for two task difficulty conditions (100ms and 200ms) using Gaussian fits. Results indicate decreased peak magnitude, increased duration and increased latency of the late component as target or task becomes harder.

durations. The peak latency duration increases when target becomes harder for the 100ms image duration and keeps consistent across target difficulty for the 100ms image duration. These results are consistent with the findings in the previous results. There is an increasing trend of peak duration when target becomes harder for both image duration. There is slightly-deteriorated effects in the hard target difficulty condition for peak magnitude at the 200ms image duration and for peak latency at 100ms image duration due to the noisy ERP signals in the hard target condition. It is also noticed that the magnitudes have significant decrease and the duration have significant increase comparing the $200ms$ to

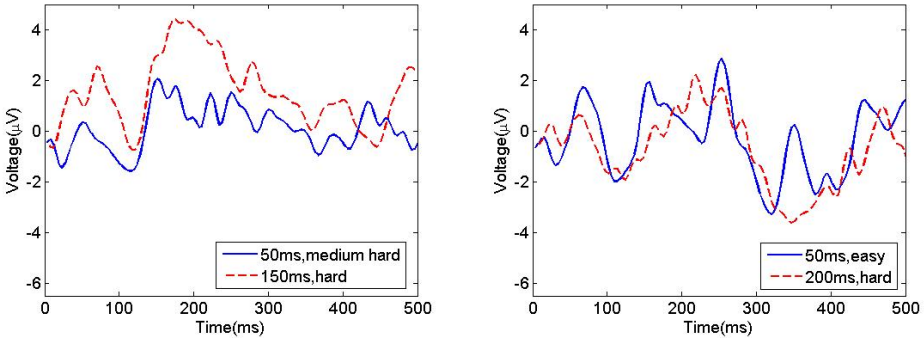
Table 3.1: Hits (Hit rate) of each target difficulty level for ten subjects at each image duration

Image Duration (Task Difficulty)	Easy	Medium Easy	Medium Hard	Hard
50ms(hard)	234 (90%)	161 (62%)	109 (42%)	62 (23%)
100ms(medium hard)	251 (97%)	240 (92%)	195 (75%)	95 (35%)
150ms(medium easy)	254 (98%)	250 (96%)	217 (83%)	132 (49%)
200ms(easy)	255 (98%)	259 (100%)	245 (94%)	157 (58%)

$100ms$ conditions. The latency at the 100ms image duration is much larger in hard task.

Pairwise-ERP Correlates to Combinations of Target and Task Difficulty

To investigate the effects of the combinational information from the task requirement and target difficulty, which accounts for both the psychological and physiological responses, we measure the ERP waveform similarity as a function of behavioral performance level as described in Section 3.3.1. The ERP waveform similarity is measured using correlation between the waveform pairs at the same behavioral performance level. Table 3.1 shows the correct detections (hits) and the hit rate of each target difficulty level at each image duration session. One can see the hit rate progressively decreases as the targets become harder or the tasks become harder. Figure 3.15 shows the ERP waveform pairs for the same behavioral performance from four different ranges of behavioral performance – $0.20 - 0.50$, $0.50 - 0.90$, $0.90 - 0.96$, $0.96 - 1$ (the four behavioral performance levels are selected to obtain roughly equal size of numbers of members in each level). One can see that the ERP waveform pairs have very similar profiles at the same behavioral performance level. To quantify the pairwise similarity, we measure the averaged pairwise correlations for all combinations of waveform pairs in the set associated with a given behavioral performance. Figure 3.16 shows the correlation analysis between the same behavior performance pairs as a function of hit rate for the three midline electrodes. There is a increasing trend of the correlation coefficients for all of the selected electrodes when behavioral performance increase.



(a) 40% vs 47%

(b) 87% vs 60%

(c) 94% vs 92%

(d) 98% vs 97%

Figure 3.15: (a)-(d) Grand Averaged ERP waveform pairs for the same behavioral performance at electrode Cz. (a) ERP pairs at the hit rates of $0.20 - 0.50$, (b) ERP pairs at the hit rates of $0.50 - 0.90$, (c) ERP pairs at the hit rates of $0.90 - 0.96$, and (d) ERP pairs at the hit rates of $0.96 - 1$. The ERP waveform similarity is measured using correlation between the waveform pairs. The correlation coefficients for these four pairs are $0.34, 0.48, 0.72, 0.88$. There are four categories of behavioral performance – the hit rate of $0.20 - 0.50$, $0.50 - 0.90$, $0.90 - 0.96$, $0.96 - 1$, each of which contains a subset of the ERPs from 16 joint categories of target and task difficulty in Table 3.1.

To further confirm whether neural activity is strongly related with target difficulty, task difficulty independently or both, we measure the ERP waveform similarity in terms of ERP pairwise correlation as a function of target difficulty, task difficulty and behavioral performance respectively. Figure 3.17 shows that the within-class correlations for target difficulty, task difficulty and behavioral performance are much higher than the between-class correlations. It indicates that both target and task difficulty are strongly correlated with neural activity. The dynamic neural signatures reflect the behavioral performance.

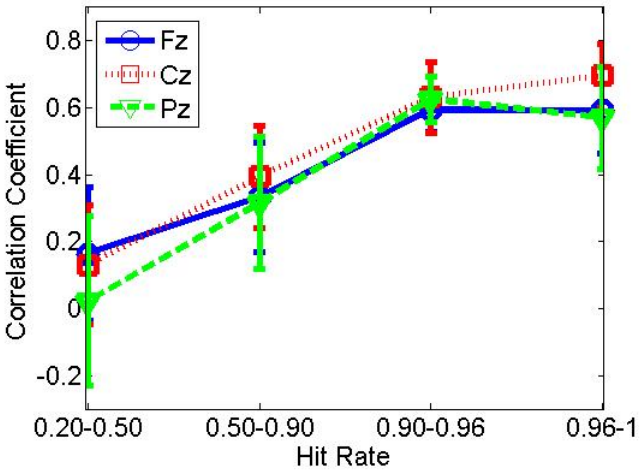


Figure 3.16: Correlation analysis between the ERP pair at the same behavioral performance for the midline electrodes. The correlation coefficients are plotted as a function of behavioral performance levels. There are four categories of behavioral performance – the hit rate of $0.20-0.50$, $0.50-0.90$, $0.90-0.96$, $0.96-1$, each of which contains a subset of the ERPs from 16 joint categories of target and task difficulty in Table 3.1.

3.3.3 Relations to Behavioral Response Time

We assess the relations between the target and task difficulty (associated with late components) and behavioral response time (RT). In Figure 3.18(a), there is an increasing trend of the mean RT as targets become progressively more difficult across different task difficulty. Surprisingly the mean RTs are almost the same across task difficulty. We suspect that the RT at hard task conditions may not be accurate because the subjects could response by chance. In Figure 3.18(b), there is significant bigger variance for both hard target condition and hard task condition compared to the easy ones. It is noticed that the hard task condition (the $50ms$ image duration has much bigger RT variance than other conditions. Figure 3.18(a) shows discriminant component maps for easy task at different target difficulty levels and for easy target condition at different task difficulty levels. It shows single-trial dynamic range of the ERP late components and RT. The results indicate that the late component is strongly correlated with the RT and the RT variance. The less strong the ERP (hard condition), the longer the RT and the bigger RT variance.

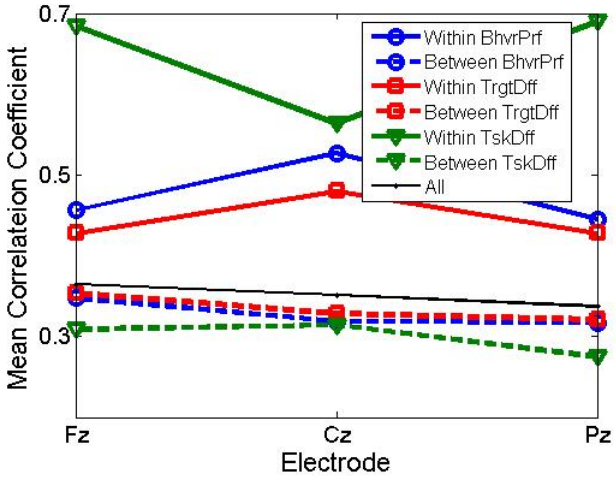
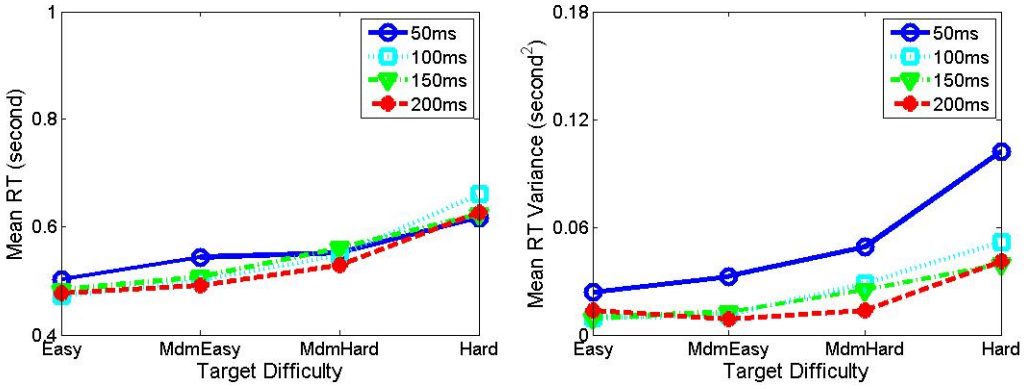


Figure 3.17: ERP pairwise correlation as a function of target difficulty, task difficulty and behavioral performance. As in Table 3.1, we have four classes for the target difficulty (each column in the table), four classes for the task difficulty (each row in the table), and four classes for the behavioral performance (the hit rate at range of $0.20 - 0.50$, $0.50 - 0.90$, $0.90 - 0.96$, $0.96 - 1$).

3.3.4 Discussions

We develop a computational model to quantify the difficulty of realistic stimuli and subject performance in terms of image duration thresholds – the minimum exposure image duration necessary for subject to detect a specific target correctly at a fixed rate. The model can be generalized to asses the difficulty of other uncontrolled stimuli.

Although the RSVP paradigm presents different challenges to the human visual perceptual system than the well-separated paradigm used in the conventional visual information processing study, we also observe the effect of target difficulty. We find the magnitude of the late component, occurring approximately 250 ms after the stimulus onset as shown in Figure 3.8 (b), is inversely proportional to the strength of the stimulus. The harder the target (the larger image threshold), the lower the magnitude of this component. This finding show that in RSVP condition, the effect of target difficulty on the late component is consistent with Picton's finding [96] and Johnson's finding [53] in well separated image display condition. Another finding is that an early component occurring approximately 110 ms after the stimulus onset that depends on target difficulty as shown in Figure 3.8



(a) (b)

Figure 3.18: Averaged behavioral response time (RT) (a) and the variance of averaged behavioral RT (b) as a function of target difficulty for four different image durations.

(a). The magnitude of this early component is also inversely proportional to the strength of the stimulus – the hard the target (the larger image threshold), the lower the magnitude of this component. In RSVP condition the early negative component, elicited by the attended stimulus, indicates that attention allocation at a given image may be triggered as early as around 110ms. The component is more likely related to low level feature difference of images [52]. It could be related with $N2$ [39] and $N2pc$ [137] but with less latency in RSVP condition. This early component has different property from the early component found by Philiastides et al. [94] in a well-separated paradigm. They found an early component occurring approximately 220 ms that depends on target difficulty and the magnitude of the difficulty component is proportional to the strength of the stimulus. Another interesting finding is that the peak latency of the early and late component shifts forward in time (latency becomes larger) as targets become harder. It indicates that subjects may engage more attention or other perception resources when targets become harder in the RSVP condition the same as in the well-separated condition [94].

Our experimental design allows us also estimate the effect of task difficulty in addition to the effect of target difficulty. The task difficulty manipulates the limits of the time that subjects can extract information from the images. Our finding is that the neural activity is correlated with task difficulty. We also find the early negative component around 110ms and late positive component around 250ms. The harder the task difficulty, the lower the

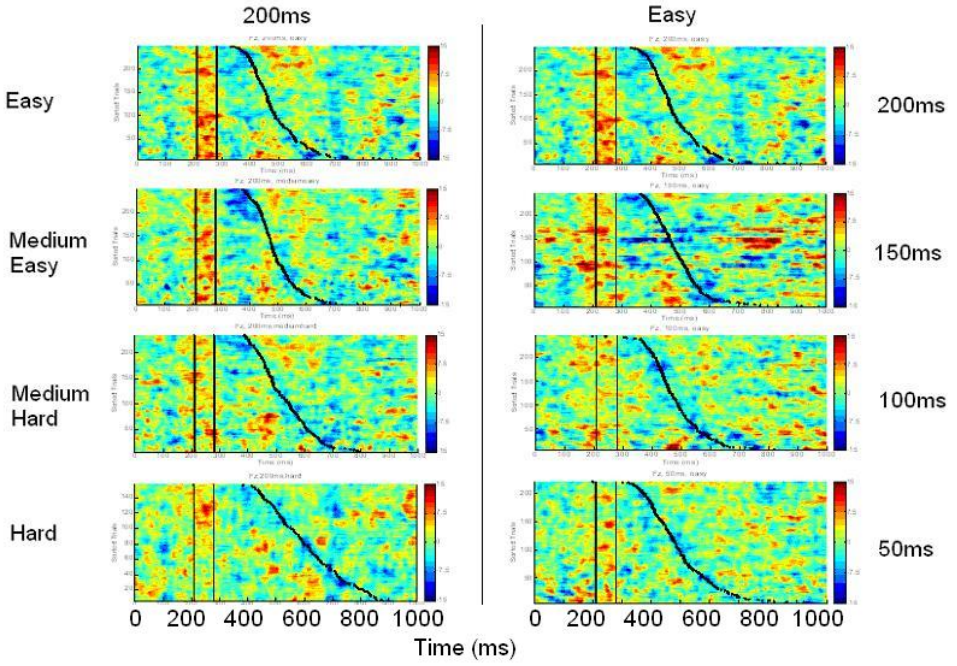


Figure 3.19: Discriminant component maps averaged cross 10 subjects at electrode Fz. In each subplot, the y-axis is sorted trials by response time (RT). All trials are aligned to the onset of stimuli. The late ERP components are shown within two vertical lines. The black sigmoidal curves represent the behavioral RTs for each condition. The left panel shows the image maps for different target difficulty levels at easy task (200ms). The right panel shows the image maps for different task difficulty levels at easy target condition.

magnitudes of the ERPs for both early and late components. Comparing the task difficulty effect with the target difficulty effect, both of them have lower magnitude of the early negative component around 110ms and lower magnitude of the late positive component around 250ms in harder target or harder task conditions. Target difficulty effect is more progressively across the difficulty level comparing to task difficulty effect because the target difficulty is a continuous variable and the task difficulty is not. There may be a threshold effect of the target difficulty, which cause much higher negative activation for easy (200ms) and medium easy (150ms) task conditions comparing to hard (50ms)and medium hard (150ms)task conditions in Figure 3.10 (a) and (b). There is latency shift trend in the target difficulty effect, but there is no this trend in the task difficulty effect. One possible reason for deteriorated effects in the hard target or task difficulty conditions is that subject barely response by chance and there are less hit trials to be averaged.

To further examine the timing of the target difficulty and task difficulty effects, we correlate the amplitude of the early component and the late component with target difficulty and task difficulty. Figure 3.11 and Figure 3.12 shows the timing of both difficulty effects in the RSVP conditions. The results indicate that the late ERP component is correlates with both target and task difficulty. The late neural responses reflects the overall information. The early ERP component is also correlates with the target and task difficulty. The early component reflects a process that may trigger allocation of more resources to a given image. In RSVP condition, we also find that the early component is not particularly sensitive as the late component to the degree of the designated focus of the attention [40].

We conduct a single-trial ERP detection to determine whether the neural correlate around 250ms after stimulus is indeed associated with target difficulty in RSVP condition. The results across four image durations show that the late component is strongly correlated to the separation of the ERPs associated with target from the ERPs associated with distractors, which imply that the late component is strongly correlated with target difficulty at the timing around 250ms to 300ms. Another interesting finding is that the latency of neural correlates associated with target difficulty shift progressively forward (latency becomes larger) as the task difficulty becomes harder. It confirms that subjects need to recruit more attention in hard task condition.

Additional peak detection analysis in Figure 3.14 allows us to relate the late component with other parameters. The peak magnitude gradually decreases as the target difficulty increases for both medium hard (100 ms) and easy tasks (200 ms). This confirmed our previous results. The peak magnitudes for 100 ms condition is significant larger than that for 200 ms condition. The peak latency has more sensitivity of target difficulty in 100 ms condition than in 200 ms condition. The results confirm the threshold effect of the task difficulty. The peak duration gradually increases as the target difficulty increase for both medium hard (100 ms) and easy tasks (200 ms). The peak duration for 100 ms condition is significant larger than that for 200 ms condition. We speculate that the longer latency and larger duration for the hard target condition and hard task condition could

be implicated in the recruitment of the relevant attentional and other cognitive resources required for the hard targets or hard tasks. These findings in the RSVP conditions are consistent with Philiastide's findings in the well-separated condition [94]. Different with their approach, our experimental designs allow our findings in two dimensions in terms of target difficulty and task difficulty.

To investigate the effects of the combinational information from the task requirement and target difficulty , which accounts for both the psychological and physiological responses, we equalize the behavioral performance on easy and difficult tasks. This is accomplished by adjusting the stimulus duration by an amount that is proportional to the stimulus difficulty. This allows us to decouple the stimulus difficulty and the task difficulty and the behavioral performance by measuring the pairwise-ERP waveform similarity. The exciting finding is that the different ERP waveform pairs become more similar at higher behavioral performance level in Figure 3.16. It is interesting that target and task difficulty manipulations affect more on the ERP characteristics at lower behavioral performance than at higher behavioral performance. The within-class and between-class correlation for target difficulty, task difficulty and behavioral performance in Figure 3.17 indicates that both target and task difficulty are strongly correlated with neural activity. The dynamic neural signatures is highly predictive of behavioral performance.

In the analysis of the target and task difficulty (associated with late components) and behavioral RT, we find that the mean RT and mean RT variance gradually increase as target difficulty increases. The mean RT variance increases as task difficulty increases. Our results in RSVP conditions are consistent with Philiastides et al's and Picton's findings in well-separate conditions. We speculate that in the difficult condition (either hard target condition or hard task condition), subjects need more time to recruit relevant attention and other neuronal resources to make a difficulty decision, by explicitly pressing a button. Surprisingly the mean RT is consistent across task difficulty. We suspect that the RT is synchronize to the onset of the image more than the offset. Another intuitive thought is that subjects may response more promptly in more demanding RSVP tasks.

In summary, opposed to the previous study, this thesis uses the ERP as a physiological marker to to assess the dynamics of resource management, such as attention allocation, in a dynamic recognition task. In contrast to sequences of well separated trials in experimental settings, the perceptual system in natural situations is confronted by rapidly changing stimuli arising from the dynamics of the environment, e.g. scenery, combined with the organism's own actions, e.g., eye movements. Our experimental design allows us to estimate how physiological and psychophysical performance depends on both target and task difficulty. This allows us to decouple the stimulus difficulty and the behavioral performance. This is accomplished by equalizing the behavioral performance on easy and difficult tasks and adjusting the stimulus duration by an amount that is proportional to the stimulus difficulty. Our results demonstrated that there is high correlation between the underlying psychological aspects of the ERPs and the dynamics of visual detection processes allocation.

We first propose an approach to assess complex realistic stimuli and subject performance. We then examine the dynamics of attention allocation as a function of the task requirement and the complexity of visual stimuli on a real target detection task. We assess the dynamics of attention allocation from a rapidly changing environment to establish a better understanding of the properties of neural correlates of human cognitive processes in the RSVP paradigm. We speculate that the attention allocation might be happened in parallel in RSVP to deal with the new stimulus before finishing the analysis of the preceding stimulus. By challenging the visual system, the process may trigger allocation of more cognitive resources, such as attention to a given image as early as 110 ms. The late component around 250ms is more sensitive to the degree of the designated focus of the attention. The early dynamic neural signature is more related to feature extraction and the late dynamic neural signature is highly predictive of behavioral performance. By controlling the limits of the time that subjects can extract information from the images in addition to the target complexity. We demonstrate that the neural correlats are associated with both target difficulty and task difficulty. The target and task difficulty effects are

more sensitive at the high end (associated lower behavioral performance). Subjects need to allocate additional neural resources and attention to evaluate harder targets or in harder tasks. Based on these findings we further surmise that, during visual information processing, the brain dynamically allocates additional cognitive resources under increasingly difficult conditions.

Chapter 4

A Framework of Target Detection using Single-trial ERPs

[1]

4.1 Introduction

We design a brain computer interface for single-trial detection of viewed images based on human dynamic brain response signatures in 32-channel EEG acquired during a RSVP condition. The system explores the feasibility of speeding up image analysis by tapping into split-second perceptual judgments of humans. The brain's novelty response – ERP associated with human perceptual judgments is used to indicate a target (object of interest) from an image set. Specifically, the task of searching for unusual or rare targets in a sequence of visual stimuli results in a reasonably strong ERP that can be detected using classifiers. Figure 4.1 shows the images of ERP signals corresponding to target and distracter stimuli for one subject at channel Cz rendered using EEGLAB [22]. The images show electrical activity following target and non-target (distractor) images for hundreds of trials. The bottom traces are the EEG signals averaged across trials. One can observe a clear ERP pattern corresponding to target images, whereas no pattern is observed in

[1] This work has been submitted for publication.

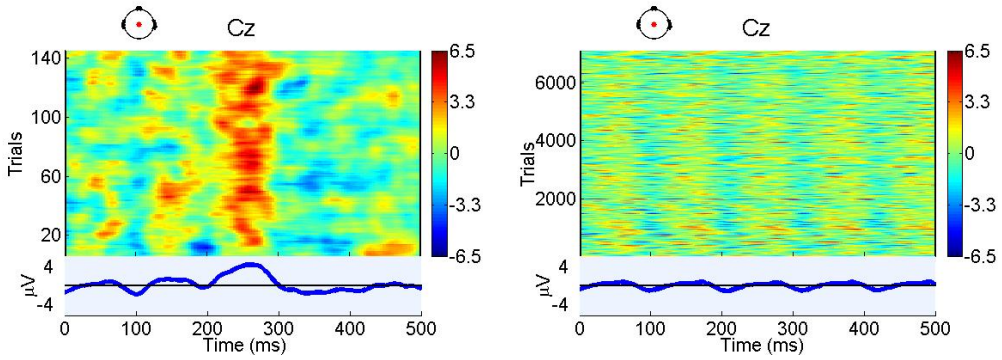


Figure 4.1: Images of ERPs associated with targets (left) and ERPs associated with distractors (right). The stimulus onset in each trial corresponds to $0ms$. The bottom traces are the EEG signals averaged over trials.

response to distractor images. The evidence indicates that the task of distinguishing the target images from distractor images can be achieved via the ERP detection.

One critical focus is to develop a neurally driven image search system using pattern recognition techniques for detecting ERPs on a single-trial basis. The goal is to develop an effective framework to detect ERPs associated with target stimuli efficiently based on the visual and cognitive systems of the human. The ERP-based image search system collects continuous EEG signals when a subject performs visual target detection from a large number of images, and then detects the ERPs associated with target stimuli. The system uses EEG as the indicator to see if an image seen briefly by the expert contains a target or a distractor. Using this technique, a human expert is able to rapidly screen a high volume of images, and identify a subset of images deserving careful inspection. The system sorts a huge stack of images by the estimated likelihood of each image being a target (referred to as image triage) and selects a subset of the images with the highest likelihood values. In Figure 4.2 the upper-half portion shows the framework of the ERP-based image search system. We use the RSVP paradigm described in Section 2.3 for image presentation. During an RSVP session, a target image, embedded in a sequence of distracter images, elicits an ERP as shown in Figure 2.2. EEG data are collected and extracted as described in Section 2.6 and Section 2.7. The raw EEG features or processed features are then subjected to classifiers, such as support vector machines (SVMs). The classifiers perform

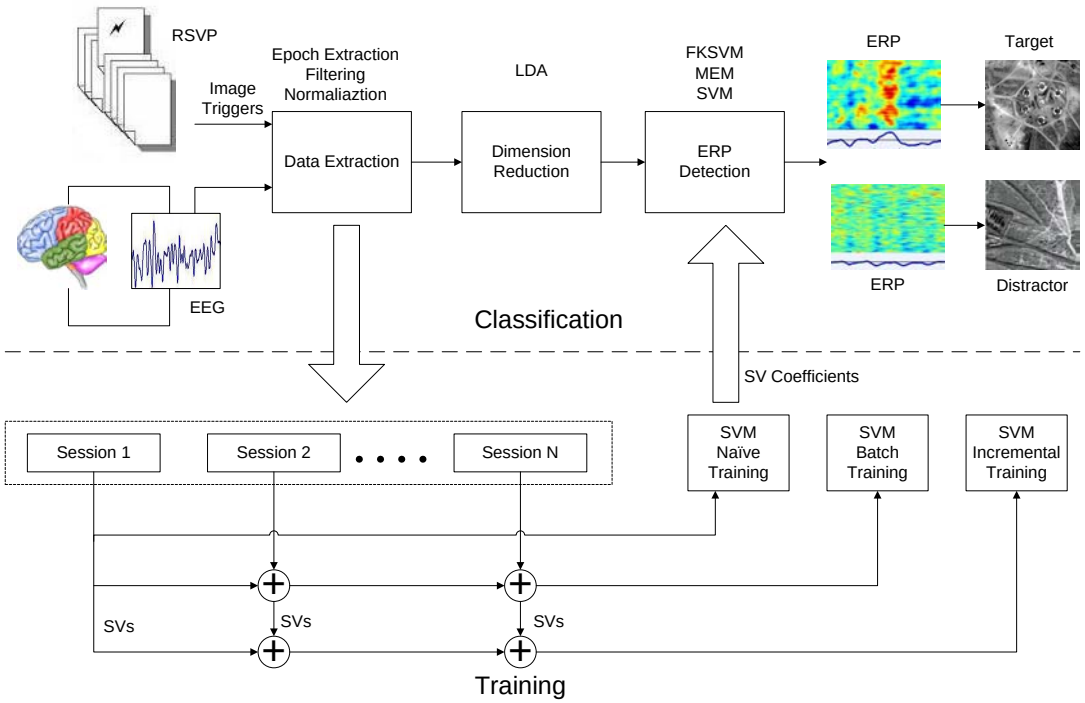


Figure 4.2: The framework of the ERP-based image search system. The upper half portion is the classification scheme. The major stages include data collection, data extraction, dimension reduction (optional), ERP detection and image triage. The lower half portion is the training schemes, which include the SVM naive training using single-session data, the SVM batch learning using cross-session data and the SVM incremental learning using only support vectors.

pattern detection to distinguish the ERPs associated with targets from the ERPs associated with distractors. In training phase, the input of classifiers is the extracted features and the output is the estimated classifier parameters, such as support vector (SV) coefficients of SVMs. The lower half portion of Figure 4.2 shows the training schemes, including the SVM naive training using single-session data, the SVM batch learning using cross-session data and the SVM incremental learning using only SVs.

The main challenges of single-trial ERP detection are the high data dimensionality and the scarcity of labeled EEG data. Although the existing methods on ERP detection [103, 70, 52, 5] have been successful, they have several shortcomings. In particular, the algorithm for each observer must be trained anew for each session, and the system does not benefit from adding other observers. Most of the existing methods for single-trial

ERP detection are trained using within-session data. The problem with using within-session data is that we may not record enough ERPs from a single subject in one sitting to sufficiently train the classifiers, at least in part because the amplitude of the ERP reduces for closely-spaced targets. The natural trade-off is that cross-session ERPs are expected to have considerably higher variation than within-session ERPs. Here we develop an ERP-based rapid visual image search system based on a novel use of incremental learning and cross-session training. We introduce an adaptive training method that uses cross-session data and propose to use incremental learning as an alternative to batch learning [44]. The impetus for using incremental learning is to reduce the computational load, which is critical for real-time implementations.

4.2 ERP Detector – Support Vector Machine

Our goal in classification is to build a reliable ERP detector to accurately detect the brain responses associated with target stimuli. The inputs of the classifiers are raw EEG measurements or processed data. The outputs of the classifier are the estimated likelihood values, which are used to label the EEG epochs according to whether or not they contain an ERP pattern associated with a target.

Traditional classification approaches are not generally capable to provide good generalization when the number of features is higher than the training set size. The SVM technique can avoid the pitfalls of very high dimensional representations because the learning algorithm is able to take advantage of the benign relation between the decision function and the distribution. The SVM learning system uses a hypothesis space of linear function in a high dimensional feature space and trains the learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. This learning strategy, introduced by Vapnick and co-workers [126, 19], has been studied intensively [20, 85, 12, 113] and widely applied to many real-world problems. Many researchers have reported that SVMs deliver state-of-art performance in a wide variety

of applications, especially for large data sets with high dimensionality, such as text categorization [51, 116], hand-written character recognition [126, 19], image classification [86] and biosequence analysis [48].

To explore more flexible classification strategy for high dimensional EEG data, we apply SVMs on single-trial ERP detection. The SVM algorithm is to map input observations to a high dimensional feature space via kernel tricks and then optimize the decision boundary by constructing a maximum-margin hyperplane. A hyperplane is an affine subspace which divides the high-dimensional feature space into two half spaces, each of which is associated with one of the two classes. The optimization is a convex quadratic programming problem. After training, the optimal Lagrange multipliers and weights for each sample are obtained. The SVs, which are the data points lying at the border of the margin, have non-zero optimal solutions for their coefficients in the final discriminant, whereas the coefficients for the other data points converge to zero. Thus the training leads to a sparse nonparametric forward discriminant function. Once we train the SVM, we simply determine on which side of the decision boundary a given test ERP pattern lies and assign the corresponding class label to it.

For a detection problem, given n data samples $\mathbf{x}_i$ and class labels $c_i \in \{-1, 1\}$, where $i = 1, ..., n$, the hyperplane is defined as

$$\mathbf{w}^T\mathbf{x} + b = 0, \tag{4.1}$$

where $\mathbf{w}$ is the weight vector, which defines a direction perpendicular to the hyperplane, and b is the bias, which defines the distance of the hyperplane from the origin. The geometric margin of the hyperplanes equals to $\frac{1}{\|\mathbf{w}\|^2}$ by normalizing the weight vector $\mathbf{w}$. To maximize the margin, the optimization problem can be formulated by minimizing

$$f(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2, \tag{4.2}$$

subject to the constraints,

$$c_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0 \quad \forall i. \tag{4.3}$$

In this constrained optimization problem [83, 113], we define the Lagrangian function as the objective function (4.2) plus a linear combination of the equality constraints in (4.3). The coefficients of the combination are the Lagrange multipliers, α_i. We extend the method by allowing inequality constraints in (4.3), and then we apply the Karush-Kuhn-Tucker theory to the above convex optimization problem in (4.2). This dual formulation requires that the following criterion be maximized

$$L_D = \sum_{i}^{n} \alpha_i - \frac{1}{2} \sum_{i,j}^{n} \alpha_i \alpha_j c_i c_j \mathbf{x}_i \mathbf{x}_j, \tag{4.4}$$

subject to the constraints,

$$\sum_{i}^{n} \alpha_i c_i = 0 \qquad \alpha_i \geq 0 \quad \forall i. \tag{4.5}$$

By solving the quadratic programming optimization problem, the linear SVM solutions are computed.

To solve a non-linear discrimination problem, we introduce a non-linear kernel function $K(\mathbf{x}, \mathbf{x}_i)$. Instead of the explicit dot products of samples in (4.4), we use a kernel function to transform the data to a high-dimensional feature space in hopes that the data are linearly separable in the feature space. One of the popular kernels is Gaussian kernel,

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\frac{\| \mathbf{x} - \mathbf{x}_i \|^2}{2\sigma^2}), \tag{4.6}$$

where σ^2 is the kernel width. The decision function is as follows,

$$f(\mathbf{x}) = sgn[\sum_{i=1}^{m} c_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b]. \tag{4.7}$$

where m is the number of SVs and $sgn()$ is the signum function. Usually the number of SVs is only a small fraction of the original training samples. The kernel parameters, such as kernel width σ^2, in the Gaussian kernel can be chosen by the users.

For the non-separable case, we introduce positive slack variables ξ_i to obtain a soft margin by minimizing

$$f(\mathbf{w}, \xi) \quad = \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i, \tag{4.8}$$

subject to the constraints,

$$\begin{aligned} c_i(\mathbf{w}^T\mathbf{x}_i + b) \quad &\geq \quad 1 - \xi_i \\ \xi_i \geq 0 \quad \forall i, \end{aligned} \tag{4.9}$$

where ξ_i are positive slack variables and the cost parameter C can be chosen by the users. The soft margin optimization can tolerate noise and outliers. The cost parameter C is a tradeoff parameter determining the relative weight of the penalty compared to the fit of the data. It controls the number of SVs and the trade-off between learning error (margin) and model complexity (the size of the slack variables). A larger C corresponds to assigning a higher penalty to errors (when the classes are not separable by a hyperplane in the feature space). Non-zero values of C are needed whenever the data are not linearly separable in the the high-dimensional feature space.

Applying the SVM approach to ERP detction involves resolving a number of design questions. The first design problem is to choose an appropriate kernel. Typically the choice of kernel will be a family of kernels parameterized by some hyperparameters. We select the linear SVM (LinearSVM) and the Gaussian kernel SVM (GKSVM) on single-trial ERP detection. The second decision is to determine the kernel parameters. We use cross-validation to set the parameters. Finally the choices of optimization criterion has to be made on whether to use the maximal margin or the soft margin approach. We use soft margin approach to solve the non-separable condition based on our high dimensionality of data and high level of noise present in our data. We use 10-fold cross validation on the training session to adjust two regularization parameters of the GKSVM: the kernel width of Gaussian kernels, σ^2, and the cost parameter, C of soft margin optimization criterion. We let the kernel size, σ^2, range from 0.01 to 500 and we let the cost parameter C range

from 1 to 10^6. We vary σ^2 and C over the logarithmic grid formed by the selected values above. The GKSVM classifier is trained using the σ^2 and C giving the best validation performance. We implement the SVM algorithm using a SVM toolbox, developed by Schwaighofer and freely available over the Internet [115].

4.3 ERP Detector – Linear Logistic Classifier

To determine whether the linear projections of the EEG sensor measurements can maximize discriminability, we adopt a linear discrimination approach based on a logistic regression model for ERP detection. Linear classifiers have been proposed on analyzing multivariate EEG signals by linearly combining channels to generate an aggregate representation of the data [91, 88, 71, 22]. The linear logistic classifier (LLC) algorithm is based on the assumption that the EEG signals are a linear combination of distributed source activity and zero-mean white Gaussian measurement noise. For sample $i = 1, ..., n$, where n is the sample number, the linear discriminant function is defined as

$$y_i = \mathbf{w}^T \mathbf{x}_i + b. \tag{4.10}$$

where $\mathbf{x}_i$ is a $(32 * 129 \times 1)$ data vector, and b is the threshold. To optimize the projection we assume that the probability of belonging to a class is represented by a logistic function. In particular the likelihood of sample $\mathbf{x}_i$ belonging to class c_i, where $c_i \in [0, 1]$, follows a logistic model. The likelihood can be parameterized as follows,

$$f(\mathbf{x}_i) = \frac{1}{1 + \exp^{-(\mathbf{w}^T \mathbf{x}_i + b)}}. \tag{4.11}$$

To obtain the optimal weights, we use the batch gradient-descent algorithm and the least-square criterion. The objective function is

$$J(\mathbf{w}) = \sum_{i=1}^{n} (c_i - f(\mathbf{x}_i))^2. \tag{4.12}$$

We calculate the gradient and update the weight vector as follows,

$$\nabla J = -2\sum_{i=1}^{n}(c_i - f(\mathbf{x}_i))f^{'}(\mathbf{x}_i), \tag{4.13}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu\nabla J, \tag{4.14}$$

where μ is a constant learning rate and k is the iteration number. Given the linear projection and the corresponding class, we use the minimum overall Bayes risk as a criterion to determine the optimal threshold. Once we obtain the weights and the threshold from the training data, we can apply the linear detector to the test data for ERP detection. Note that the classifier in (4.11) can also be trained using the maximum likelihood.

4.4 Dimension Reduction - Linear Discriminant Analysis

Given limited training examples and high dimensionality in single-trial ERP detction, it is hard to expect good generalization for classifiers. For the spatio-temporal EEG measurements, we combine multiple channels to generate an aggregate representation of the data as described in Section 2.7. For instance, if a 32-channel EEG system sampling at 256Hz rate, a $500ms$ epoch will produce a feature with 32×129 dimension. We perform channel dimension reduction to improve the ERP detection performance.

We apply linear discriminant analysis (LDA) [26] to project multichannel EEG data to lower dimensions and congregate the selected channel projections to form a feature vector as the basis for classification. We employ a linear channel dimensionality reduction method inspired by the LDA approach in classifier design. Specifically, we seek to identify a set of channel linear combination coefficients that keep the average EEG responses between the two classes as separated as possible, and simultaneously attempt to minimize the total variance in the projected data. For each ERP sample $\mathbf{y_i^1}$ associated with a target and each ERP sample $\mathbf{y_i^0}$ associated with a disctractor (which are $C \times T$ matrices where C is the number of channels and T is the number of temporal samples following stimulus

onset), we obtain the trial-average and trial-covariation matrices as follows:

$$\begin{aligned} \mathbf{M}^1 &= \frac{1}{N^1}\sum_{i=1}^{N^1}\mathbf{y}_i^1, \\ \mathbf{M}^0 &= \frac{1}{N^0}\sum_{i=1}^{N^0}\mathbf{y}_i^0, \end{aligned} \tag{4.15}$$

$$\begin{aligned} \mathbf{C}^1 &= \frac{1}{N^1}\sum_{i=1}^{N^1}(\mathbf{y}_i^1 - \mathbf{M}^1)(\mathbf{y}_i^1 - \mathbf{M}^1)^T, \\ \mathbf{C}^0 &= \frac{1}{N^0}\sum_{i=1}^{N^0}(\mathbf{y}_i^0 - \mathbf{M}^0)(\mathbf{y}_i^0 - \mathbf{M}^0)^T. \end{aligned} \tag{4.16}$$

For a single dimensional linear channel projection of the form $\mathbf{w}^T\mathbf{y}_i^1$ and $\mathbf{w}^T\mathbf{y}_i^0$, the linear projection direction $\mathbf{w}$ is identified by maximizing the Fisher discriminant

$$J(\mathbf{w}) = \mathbf{w}^T\mathbf{S}_b\mathbf{w}/\mathbf{w}^T\mathbf{S}_w\mathbf{w}, \tag{4.17}$$

where the between cluster scatter matrix is $\mathbf{S}_b = (\mathbf{M}^1 - \mathbf{M}^0)(\mathbf{M}^1 - \mathbf{M}^0)^T$ and the within cluster scatter matrix is $\mathbf{S}_w = (\mathbf{C}^1 + \mathbf{C}^0)$. The solution to this is given by the generalized eigen decomposition of this symmetric matrix pair, which can also be obtained as the largest eigenvector of a nonsymmetric matrix as follows:

$$\mathbf{w} = eig(\mathbf{S}_w^{-1}\mathbf{S}_b). \tag{4.18}$$

For projections to higher-than-one dimension, we select the subset of largest eigenvectors with cardinality matching the desired reduced channel dimensionality. The number of eigenvectors to be retained are determined using cross-validation.

4.5 Cross-session Training

The main challenges of single-trial ERP detection are the high data dimensionality and the scarcity of labeled EEG data. Ideally, we would collect large amounts of data from each

subject during a single protracted session. However, this is both monotonous and time consuming. Besides, given the low prior probability of a target in this task, we obtain very few ERP samples associated with targets within a long sequence of ERP samples associated with distractors because the amplitude of the $P300$ reduces for closely-spaced targets. Since it is impractical to train a subject extensively in one session to obtain the sufficient ERP training samples, we propose the cross-session training, which train the classifiers using data aggregated across multiple sessions. However, when multiple EEG measurements are obtained from each individual at different times and possibly under changing experimental conditions, we cannot perfectly duplicate the conditions under which previous measurements were taken. Hence, considerable variations of the measurements from session to session could be a problem.

Instead of conventional single-session training, we explore the feasibility of cross-session training on ERP detection to improve the generalization performance of classifiers. To assess cross-session performance, data were collected at different times and under different experimental conditions from four naive subjects performing image search tasks. The naive dataset #4 was described in Section 2.1 and Section 2.5. Each subject had 10 sessions performed at different time and different days. The EEG data were segmented and pre-processed using the procedures described in Section 2.7.

4.5.1 Naive Learning

We use the term 'naive learning" to refer to single-session training. To simulate a realistic scenario, we use only the current session as the test set and the previous one as the training set. In naive learning (single-session training), we train on $session\ 1(S_1)$ and test on S_2.

4.5.2 Batch Learning

We use the term "batch learning" to refer to cross-session training using all data from multiple sessions. In batch learning (cross-session training), we aggregate the data across

all previous sessions for training. For instance, We train on $S_1 \cup S_2$ and test on S_3; and so on until we train on $S_1 \cup ... \cup S_9$ and test on S_{10}. The aggregated data are subjected to a classifier to evaluate the cross-session performance. We use the GKSVM with kernel size $k = 1$ and cost parameter $C = 10$ as the ERP detector in cross-session training.

We use Monte Carlo method to repeat pseudorandom sampling to compute the averaged cross-session performance. We use only the current session as the test set and the previous sessions as the training set to create Monte Carlo pseudorandom sessions. In naive learning, we train on train on S_1 and test on S_2; train on S_2 and test on S_3; and so on until we train on S_9 and test on S_{10}. We compute the mean and the standard deviation across all trials for the single-session performance. In batch learning, we train on $S_1 \cup S_2$ and test on S_3; train $S_2 \cup S_3$ and test on S_4; and so on until we train on $S_8 \cup S_9$ and test on S_{10}. We compute the mean and the standard deviation across trials for the double-session performance. Similarly we use the Monte Carlo method to create pseudorandom sessions for multiple-session training and computer the averaged performance.

4.6 Incremental Learning

The cross-session training in batch mode may produce higher performance than the single-session training due to more training samples. However, such batch training is computationally intensive and thus infeasible for real-time systems. Incremental learning paradigm, as opposed to the batch learning paradigm, in which all training examples are provided at once for optimization, is a training mode where only a few training examples are added at a time to update model parameters. The motivation of incremental learning is to deal with very large training sets or non-stationary data. An important advantage of incremental learning is that it allows the algorithm to combine additional available training examples without having to retrain classifiers from scratch. This has numerous benefits, including saving a substantial amount of storage space and speeding the computation up. Therefore, incremental learning algorithms have been investigated in many applications,

such as intrusion detection [65] and blind separation task of acoustic signals [80]. One of the main difficulties with using incremental learning methods is the sensitivity of choosing training parameters.

Since the emergence of SVM in the 1990s, incremental learning of SVMs has been investigated intensively [121, 14]. Early work provided only approximate solutions [121, 59, 106]. In [121], Syed proposed an approximate solution to the problem of incremental SVM learning. An SVM was trained on new data by discarding all previous data except the support vectors, which were combined with the new subset of data. Cauwenberghs and Poggio first proposed an exact SVM incremental algorithm [14]. It used an online recursive algorithm for SVM training and updated an optimal solution of the training one vector at a time. To our knowledge there are not very many follow-up publications, possibly because it is not easy to implement. In Ho's work [41], they applied incremental SVM to stream-based active learning. Laskov et al. focused on the design and analysis of efficient incremental SVM learning and proposed a new design of storage and numerical operation to speed up the training of incremental SVM [64]. Compared with Cauwengerghs's work, Syed's method is more straightforward and easier to implement.

The essential property of the SVM algorithm is that only the SVs contribute to the decision boundary so that the remaining training examples may be regarded as redundant. Based on this property, Syed et al. proposed an incremental learning for the SVM to deal with large datasets [121]. They segmented a huge dataset into small partitions to avoid problems associated with limited available memory, and incrementally trained the SVM with the small partitions. Their results demonstrated that the SVs selected by the SVM algorithm was a minimal set. Any further removal of data samples significantly deteriorated the performance because the loss of SVs led to loss of vital information about the class distribution.

Motivated by Syed's method, we develop an incremental learning scheme for cross-session ERP detection to solve the problem of data growing over time. The incremental

learning ERP detection is to train an SVM on the previous session of EEG data. The SVs found during training are preserved and combined with the training samples from the current session. For the cross-session EEG data (naive dataset #2) in Section 2.5, there are 10 datasets, S_1 to S_{10}. Instead of training on all previous data as S_1, $S_1 \cup S_2$, ..., $S_1 \cup ... \cup S_9$ as in the batch learning, we preserve the SVs from the previous training sets and discard the redundant data. Let V_i represent the SVs in session S_i. For the proposed incremental learning scheme, we train using S_1, $V_1 \cup S_2$, $V_1, \cup V_2 \cup S_3$,..., $V_1 \cup ... V_8 \cup S_9$. Since it compacts the previous training data to the SVs and then combines only the SVs with the new dataset, it is more computationally efficient. Based on the fact that only the support vectors contribute to the essential class boundary in the SVM algorithm, we would argue that the model obtained from the incremental SVM should provide the same or similar performance as the original method using all the data together to train.

4.7 Performance Evaluation

4.7.1 Receiver Operating Characteristic Curve

We adopt the well-developed statistical tool – the receiver operating characteristic (ROC) curve to depict the ERP performance and assess the quantitative efficacy. The ROC curve [92, 31] describes the relationship between the false positive fraction (FPF) and the true positive fraction (TPF) as the threshold for discrimination between two classes is varied. The most widely used summary measure is the area under the ROC curve (AUC). A perfect test has an AUC of 1.0, whereas random chance gives an AUC of 0.5. We use AUC to evaluate the single-trial ERP performance.

4.7.2 Compare Correlated AUCs

There are many ways to analyze correlated ROCs [92, 63, 37, 3]. We use Delong's non-parametric approach [21] to evaluate correlated AUCs by generating an estimated covariance matrix. Suppose the total test sample number is N and the numbers of targets and

distractors are m and n where $n = N - m$. Let $\mathbf{X}_i, i = 1, 2, ..., m$ and $\mathbf{Y}_j, j = 1, 2, ..., n$ be the output values of the variables from two classifiers respectively. These outcome values are used to construct an empirical ROC curves. The AUC is the probability that a randomly selected observation from the distractor population will be less than or equal to a randomly selected observation from the target population. It can be computed as the average for all possible pairs over a comparison function, ψ, as

$$\hat{AUC} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} \psi(\mathbf{X}_i, \mathbf{Y}_j), \tag{4.19}$$

where

$$\psi(\mathbf{X}_i, \mathbf{Y}_j) = \begin{cases} 1 & \mathbf{Y} < \mathbf{X} \\ \frac{1}{2} & Y = \mathbf{X} \\ 0 & \mathbf{Y} > \mathbf{X} \end{cases}$$

In terms of probabilities, the expected value of the AUC $E(\hat{AUC}) = Pr(\mathbf{Y} < \mathbf{X}) + \frac{1}{2}Pr(\mathbf{Y} = \mathbf{X})$. For the estimated AUC from classifier 1, the X-component and Y-component are defined for the target class and the distractor class, respectively, as

$$\mathbf{V}_1^1(\mathbf{X}_i) = \frac{1}{n} \sum_{j=1}^{n} \psi(\mathbf{X}_i^1, \mathbf{Y}_j^1) \quad (i = 1, 2, ..., m), \tag{4.20}$$

and

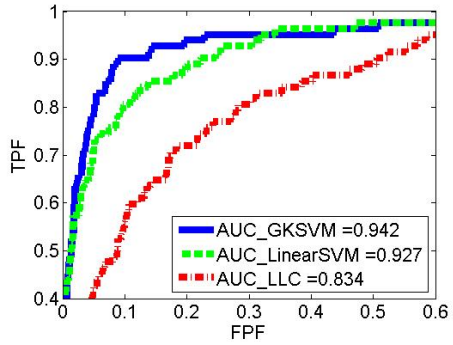
$$\mathbf{V}_0^1(\mathbf{Y}_i) = \frac{1}{m} \sum_{i=1}^{m} \psi(\mathbf{X}_i^1, \mathbf{Y}_j^1) \quad (i = 1, 2, ..., n). \tag{4.21}$$

where $\mathbf{V}_1^1(\mathbf{X}_i)$ and $\mathbf{V}_0^1(\mathbf{Y}_i)$ are the $m \times 1$ and $n \times 1$ vectors. Similarly, we can obtain $\mathbf{V}_1^2(\mathbf{X}_i)$ and $\mathbf{V}_0^2(\mathbf{Y}_i)$ for the two classes of classifier 2. Let $\mathbf{V}_1 = [\mathbf{V}_1^1(\mathbf{X}_i), \mathbf{V}_1^2(\mathbf{X}_i)]$ and $\mathbf{V}_0 = [\mathbf{V}_0^1(\mathbf{X}_i), \mathbf{V}_0^2(\mathbf{X}_i)]$. The vector of estimated AUCs for two classifiers is $\mathbf{\hat{AUC}} = [\hat{AUC}^1, \hat{AUC}^2]'$. The estimated covariance matrix for the two classifiers is

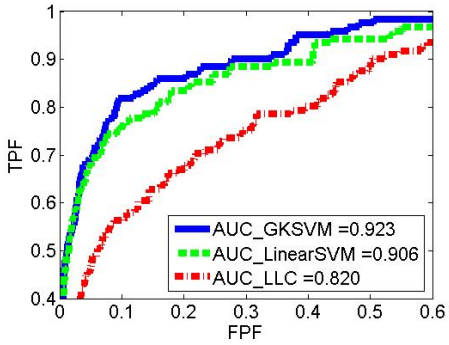
$$\mathbf{S} = \frac{1}{m}\mathbf{S}_1 + \frac{1}{n}\mathbf{S}_0, \tag{4.22}$$

where

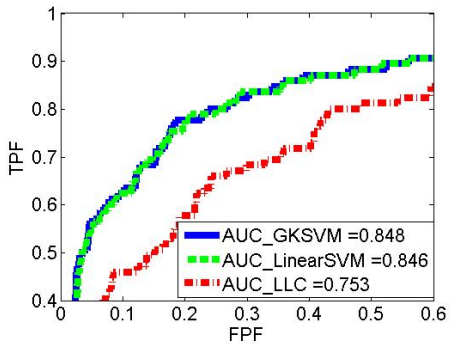
$$\mathbf{S}_1 = \frac{1}{m-1}\mathbf{V}_1{}'\mathbf{V}_1 - m\mathbf{\hat{AUC}}'\mathbf{\hat{AUC}}, \tag{4.23}$$



(a) Subject 1



(b) Subject 2



(c) Subject 3

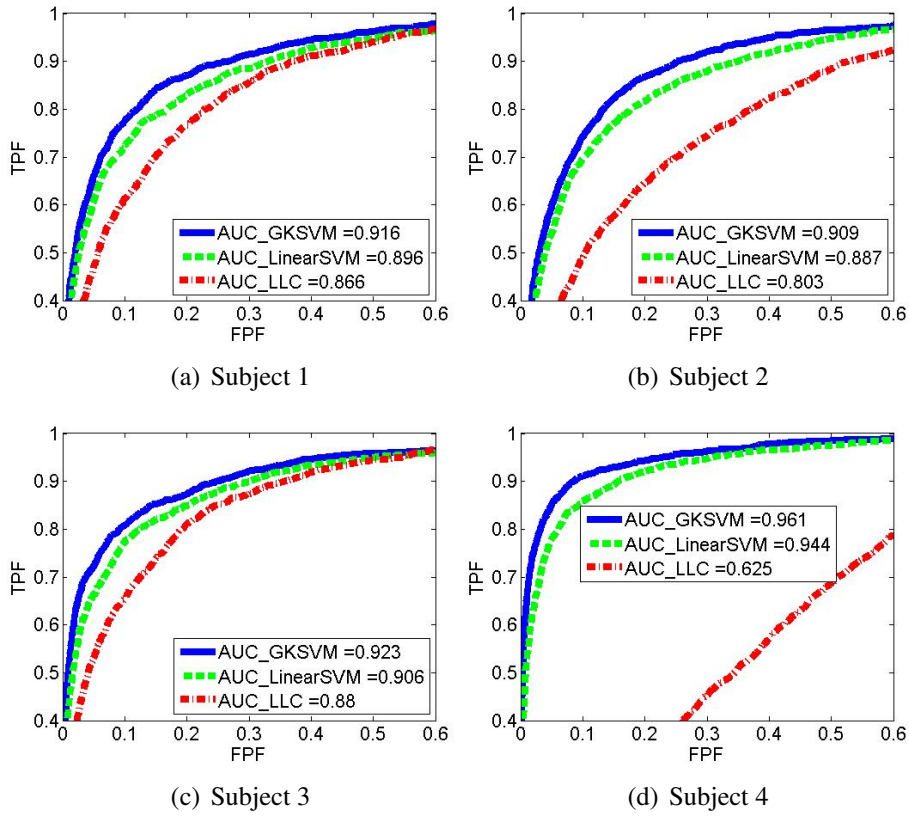
Figure 4.3: The ROC curves of the GKSVM (solid), linearSVM (dashed), and LLC (dash-dotted) for the three IA subjects. ROC curve depicts the relationship between the false positive fraction (FPF) and the true positive fraction (TPF). The performance in term of the area under the ROC curves (AUC) for each classifier is shown. The performance of the linear and nonlinear SVMs is significantly higher than the performance of the LLC.

$$\mathbf{S}_0 = \frac{1}{n-1}\mathbf{V}_0{}'\mathbf{V}_0 - n\mathbf{A}\hat{\mathbf{U}}\mathbf{C}'\mathbf{A}\hat{\mathbf{U}}\mathbf{C}. \tag{4.24}$$

The $\mathbf{S}_1$ and $\mathbf{S}_0$ are 2×2 covariance matrices for the $\mathbf{V}_1$ of target class and the $\mathbf{V}_0$ of distractor class. The standard deviation of the estimated AUC is $\sqrt{\mathbf{LSL}'}$, where $\mathbf{L} = [1, -1]$ is the contrast. The covariance matrix and the standard deviation of the estimates may be used to construct the confidence regions.

4.8 ERP Detection Performance: SVM vs. LLC

The purpose of this experiment is to select a classifier on single-trial ERP detection for incremental adaption process. We evaluate the detection performance of single-trial ERP



(a) Subject 1 (b) Subject 2

(c) Subject 3 (d) Subject 4

Figure 4.4: The ROC curves of the GKSVM (blue solid), linearSVM (green dashed), and LLC (red dash-dotted) for four naive subjects. ROC curve depicts the relationship between the false positive fraction (FPF) and the true positive fraction (TPF). The performance in term of the area under the ROC curves (AUC) for each classifier is shown. The performance of the linear and nonlinear SVMs is significantly higher than the performance of the LLC.

over three classifiers – the GKSVM, the linearSVM and the LLC on both professional image analysts (IAs) and naive subjects (IA Dataset #1 and the naive dataset #1 described Section 2.1 and Section 2.5). For the IA dataset, we have three subjects. Each one has one training session and seven test sessions. For each subject, the processed data have around 50 target samples and 500 distractor samples in the training set. The total test samples contain around 100 targets and 20000 distractor samples. For the naive dataset, we have four subjects and each one has ten sessions. We select the first session as the training set and the other nine sessions as the test sets. For each subject, the processed data have around 140 target samples and 1000 distractor samples in the training set. The

total test samples contain around 1400 targets and 65000 distractor samples.

We use the non-parametric method to evaluated the difference of two correlated AUCs. Our null Hypothesis, H_0, is that there is no statistically-significant difference between the two correlated AUCs (the AUC of the SVM and the AUC of the LLC). We use 10-fold cross validation to select the parameters for GKSVMs. The optimal parameters for the three IA subjects are $\sigma^2 = 0.05$ and $C = 100$; $\sigma^2 = 0.1$ and $C = 100$; $\sigma^2 = 10$ and $C = 10000$, respectively. The optimal parameters for the four naive subjects are $\sigma^2 = 1$ and $C = 10$; $\sigma^2 = 1$ and $C = 10$; $\sigma^2 = 1$ and $C = 100$; $\sigma^2 = 1$ and $C = 1000$, respectively. We use $C = 1$ for the linearSVMs (in our experience the performance of the linearSVMs is not sensitive to the C value).

In the comparison of the GKSVM and the LLC, the GKSVM achieves significant higher ERP detection performance than the LLC in terms of the AUC for both the IA dataset and the naive dataset. For the three IA subjects the AUCs of the GKSVM is higher than the LLC by 0.108, 0.103 and 0.095, respectively, as shown in Figure 4.3. The two-tailed p values for the three IA subjects are $<< 0.0004$ ($z = 5.179$, 5.734 and 4.556, respectively). For the four naive subjects the AUCs of the GKSVM is higher than the LLC by 0.050, 0.106, 0.043 and 0.336, respectively, as shown in Figure 4.4. The two-tailed p values for four naive subjects are $<< 0.0004$ ($z = 14.529$, 21.063, 12.008 and 47.420, respectively). From the results of both the IA dataset and the naive dataset, one can conclude that the difference between the AUCs from the GKSVM and the LLC are highly statistically significant and thus we can reject H_0.

Similarly, in the comparison of the linearSVM and the LLC, the linearSVM achieves significant higher ERP detection performance than the LLC in terms of the AUC for both IA subjects and naive subjects. For the three IA subjects the AUCs of the linearSVM is higher than the LLC by 0.093, 0.086 and 0.093, respectively, as shown in Figure 4.3. For the four naive subjects the AUCs of the linearSVM is higher than the LLC by 0.030, 0.084, 0.026 and 0.319 respectively, as shown in Figure 4.4. The two-tailed p values for the three IA subjects and the four naive subjects are $<< 0.0004$ ($z = 5.260$, 4.773, 4.383, 8.708,

17.560, 7.413, and 14.271, respectively). One can conclude that the difference between the AUCs from the linear SVM and the LLC are also highly statistically significant and thus reject H_0.

Comparing the GKSVM with the linearSVM, the GKSVM achieves significant higher ERP detection performance than the linearSVM in terms of the AUC for six out of seven subjects. For the three IA subjects the AUCs of the GKSVM is higher than the linearSVM by 0.015, 0.017 and 0.002, respectively, as shown in Figure 4.3. For the four naive subjects the AUCs of the GKSVM is higher than the linearSVM by 0.020, 0.022, 0.017 and 0.017 respectively, as shown in Figure 4.4. The two-tailed p values for the IA dataset are 0.215, 0.021, 0.029 (z = 1.245, 2.307, 2.186, respectively) and the two-tailed p values for the naive dataset are $<< 0.0004$ (z = 9.420,10.237, 9.848,and 44.137, respectively). The difference between the AUCs from the GKSVM and the LinearSVM are statistically significant for all subjects except one IA subject.

The non-parametric test for correlated AUC measurements indicates that the linear and nonlinear SVM achieves significantly better performance, in a statistical sense, than the LLC for single-trial ERP detection on both IA and naive datasets. Likewise, the GKSVM, at the cost of additional computational complexity, performs better than the linearSVM for ERP detection. The result indicates that the data may not be linearly separable in the original feature space so that the nonlinear projection to the high-dimensional feature space helps the discrimination of the two classes. However the tradeoff is the computation efficiency.

For high dimensional data (32×129 dimension and sparse training samples (for IA dataset, roughly 50 positive samples and 500 negative samples; for naive dataset, roughtly 140 positive samples and 950 negative samples), the SVMs are capable to capture nonlinear class separation boundary. The SVMs map input data to a high dimensional feature space via kernel tricks and optimize the linear separating hyper-plane in feature space. However, the LLC is simply a linear hyper-plane boundary. This could be the main reason

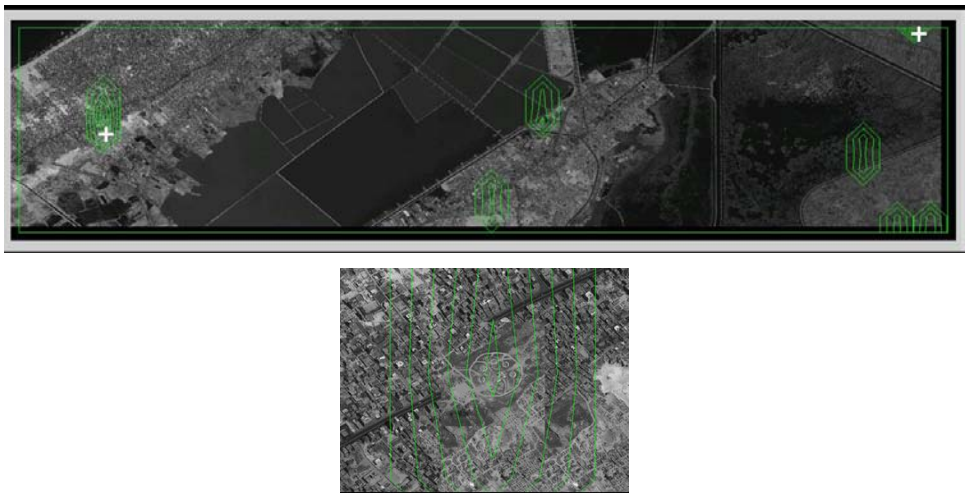


Figure 4.5: Broad area image with overlaid contour maps (upper). The crosses indicate true target locations. Users can zoom into the contour hotspot to confirm the presence of a target (lower).

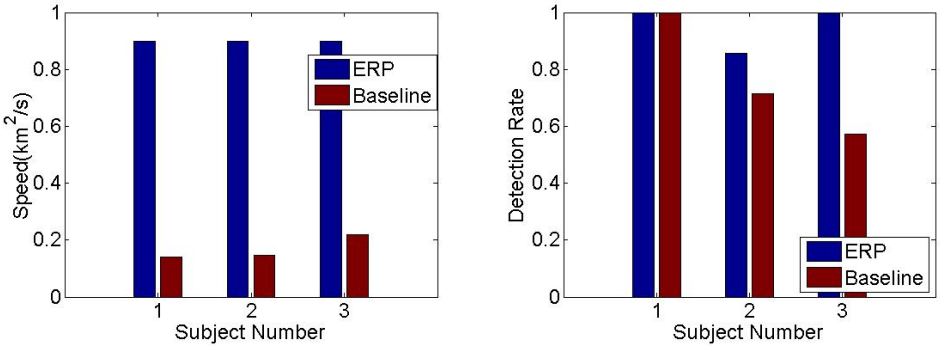


Figure 4.6: The detection speed (left) and detection rate (right) averaged across test sessions for the three IA subjects.

for performance discrepancy. Some regularization could be made to the LLC to approximate the SVMs, for instance, by modified iteratively re-weighted least squares estimation procedure or a modified penalized log likelihood function. Consequently, based on the results, we chose the GKSVM as the classifier for the incremental adaptation process.

4.9 Efficiency: ERP Approach vs. Tradition Approach

We conduct a experiment to explore the efficiency of neurophysiologically-driven image search. More specifically, we compare the ERP approach with the tradition image viewing

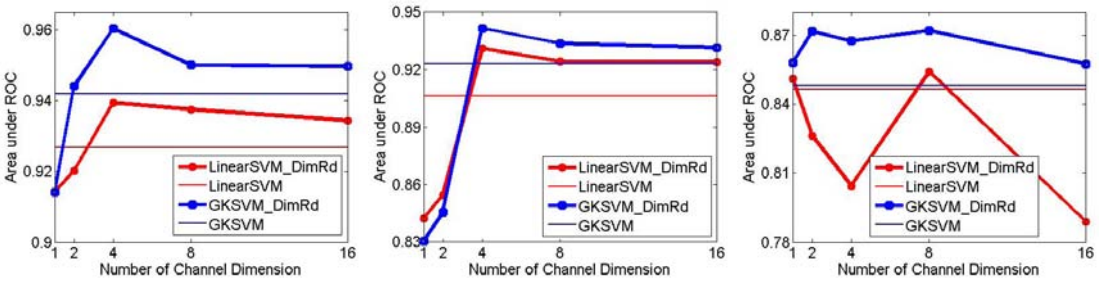
approach on the detection rate (the ratio of the number of detected targets and the number of total targets) and detection speed (the ratio of the total search time and the total image area, which is measured in units of $seconds/km^2$). The study was carried on a group of professional image analysts as described in Section 2.1.

The ERP-based target detection system collects EEG signals as a subject observes a high-speed scan of the thousands of chips extracted from several broad-area images and then analyzes the data to identify ERPs. We use IA Dataset #2 in this experiment. We employ the GKSVM as the ERP detector in this study. Based on the estimated likelihood values from the GKSVM, we can construct a contour plot of the target likelihood and overlay it on the associated broad-area image. The human experts use this plot to do a final confirmation.

In the tradition image viewing approach, participants use a geo-spatial analysis tool called GlobalMapper (Global Mapper Software). It provides zoom and pan controls and allows high resolution satellite imagery to be efficiently searched and annotated. Participants are allowed as much time as they wished to search the targets in a broad area image. A set of prototype images depicting the targets are shown to each participant.

Figure 4.5 shows the target likelihood contour maps (for all contours above a fixed threshold) for one of the broad-area images used in the test. One can see that, in this case, the ERP-based approach accurately detects one target and has a few false alarms. This visualization technique allows efficient post-processing of the triage outputs.

Figure 4.6 shows the detection rate and the detection speed averaged across test sessions for each subject. We can see that the averaged detection rates of the ERP system are equal to or higher than those of the manual method. The detection speeds of the ERP system are much faster than those of the manual method for all subjects. The overall detection rate across subjects for the ERP system is 93% compared to 67% for the manual approach. The averaged AUC across seven test sessions for the three subjects is 0.82. The overall false alarm rate for the ERP system is higher than the manual method, which had



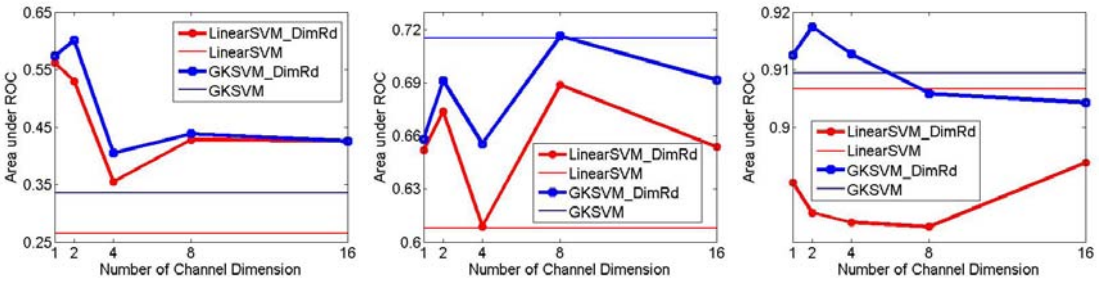
(a) Subject 1 (b) Subject 2 (c) Subject 3

Figure 4.7: Dimension reduction performance on ERP detection using LDA on linearSVM and GKSVM for three IA subjects from IA Dataset #1. The red and blue curves represent the linearSVM and the GKSVM detection performance respectively. The red and blue thin lines represent the performance using the original features on the linearSVM and the GKSVM respectively. One can see the dimension reduction performance using projection features is higher than the performance using original features on both classifiers for all subjects.

zero false alarms. The precision for the ERP system is 78% and the precision for the manual method is 100%. The overall detection speed for the ERP system is 5.3 times faster than the manual method. The results demonstrate that the ERP-based image search system is more efficient than the tradition image viewing paradigm in terms of the detection speed and detection rate.

4.10 Dimension Reduction Performance

The purpose of this study is to compare the ERP detection performance with and without dimension reduction. We use the LDA approach for reduce channel dimensionality on two IA datasets (IA Dataset #1 and IA Dataset #2 as described in Section 2.1 and Section 2.5). The dimension reduction performance is tested on both linearSVM and GKSVM. The original channel number is 32. The number of reduced dimension are set as $2^0, 2^1, 2^2, 2^3, 2^4$ to evaluation the channel dimension reduction performance. Figure 4.7 and Figure 4.8 shows the dimension reduction detection performance using LDA on the linearSVM and the GKSVM for six IA subjects from two datasets. One can see the dimension reduction performance using projection features is higher than the performance using original features on both classifiers for all subjects. It indicates that the LDA



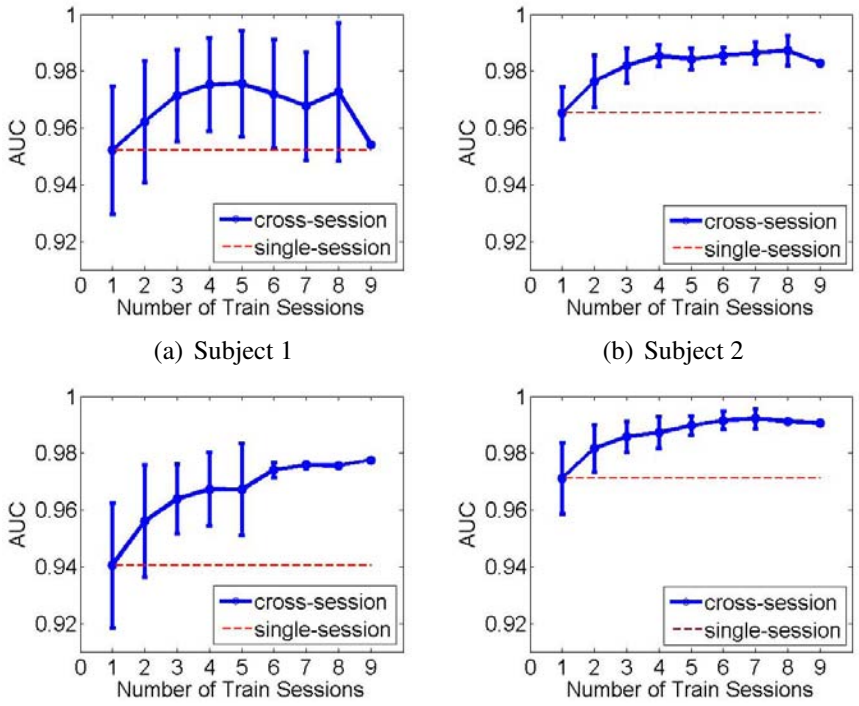
(a) Subject 1 (b) Subject 2 (c) Subject 3

Figure 4.8: Dimension reduction performance on ERP detection using LDA on linearSVM and GKSVM for three IA subjects from A Dataset #2. The red and blue curves represent the linearSVM and the GKSVM detection performance respectively. The red and blue thin lines represent the performance using the original features on the linearSVM and the GKSVM respectively. One can see the dimension reduction performance using projection features is higher than the performance using original features on both classifiers for all subjects.

feature projections capture the distinguishable characteristics and thus produce high detection performance with less computational complexity. For the linearSVM the optimal numbers of projections are $4, 4, 8, 1, 8, 16$ and for the GKSVM the optimal numbers of projections are $4, 4, 8, 2, 8, 2$. The numbers of dimension for the highest performance are much less than the original dimension for both datasets.

4.11 Cross-session Performance

We conduct an experiment to extend the results from our ERP detection framework (Section 4.8) and explore the robustness of the ERP-detection approach to changes of EEG signals over time. Here we evaluate the viability of cross-session batch learning on ERP detection by comparing it with naive learning on naive dataset #4. Figure 4.10 shows the ERP detection performance in term of AUC using the batch learning on cross-session data and the naive learning using single-session data. We use only the current session as the test set and the previous sessions as the training set to create Monte Carlo pseudorandom sessions. The averaged performance are computed across Monte Carlo trials. The GKSVM achieves high cross-session generalization performance for all subjects. The



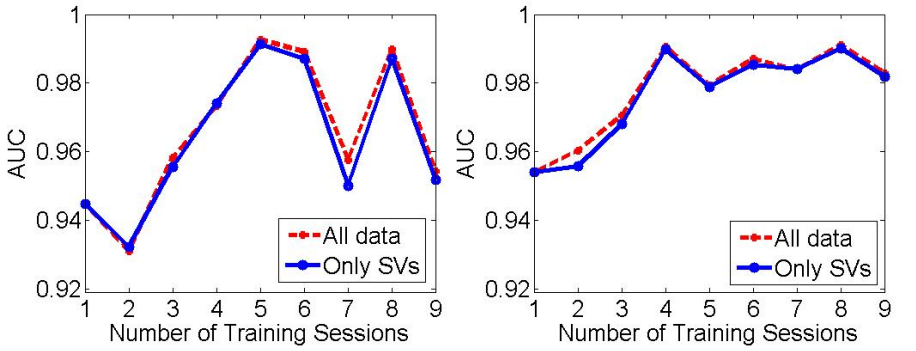
(a) Subject 1 (b) Subject 2

(c) Subject 3 (d) Subject 4

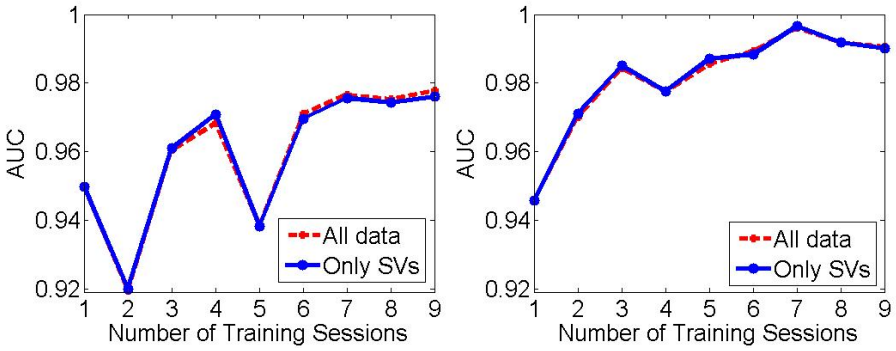
Figure 4.9: The cross-session performance in term of the area under ROC (AUC) on different number of training sessions for four subject using the GKSVM. We use only the current session as the test set and the previous sessions as the training set to create Monte Carlo pseudorandom sessions. In naive learning using single-session data (red dashed line), we train on the single previous session and test on the current session. In batch learning using cross-session data (blue solid curve), we train on all previous multi-session data and test on the current session. The batch learning performance using cross-session training is compared with the naive learning performance using single-session training. One can see the increasing trend after aggregated more training data for all subjects.

averaged AUC exhibits a generally increasing trend with the inclusion of additional training data from subsequent sessions for four subjects. The averaged performance across subjects increases around 5%.

Our results demonstrate the viability of cross-session single-trial ERP detection. Given a reasonable amount of cross-session training data, the SVM achieves excellent generalization performance, as indicated by the high AUC values. The results suggest that the classifier is able to capture much of the range of variation in our EEG data and approach the problem of characterizing between session variation in signal statistics. The results show that intra-session variances do not significantly deteriorate detection performance.



(a) Subject 1 (b) Subject 2



(c) Subject 3 (d) Subject 4

Figure 4.10: The Incremental learning performance in term of the area under ROC (AUC) using the GKSVM. The AUC as a function of the number of training sessions for four subjects for batch learning (red dashed; using all previous data for training), and incremental learning (blue solid; only the SVs are propagated). The incremental learning performance is identical to the batch learning performance across different number of training sessions for all subjects.

4.12 Incremental Learning vs. Batch Learning

We evaluate the efficiency of incremental learning on ERP detection by comparing it with batch learning on both detection performance and computational load. We examine the method on cross-day data from the same naive subject group as previous section. The aggregated SVs from the previous sessions combined with the current session data were input to the SVM to evaluate the cross-session performance for incremental learning.

We compare the single-trial ERP detection performance on the batch learning method and the incremental learning method. Figure 4.10 shows the cross-session ERP detection performance in terms of the AUCs for four subjects. The incremental learning SVM

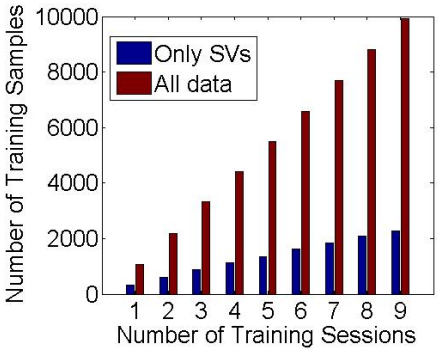


Figure 4.11: The number of training samples for different number of training sessions for subject 1 using batch learning (uses all previous data for training) and incremental learning (only the SVs are propagated).

achieves identical AUCs as the batch approach for all subjects. The AUCs of both incremental learning and batch learning exhibit a generally increasing trend with the inclusion of additional training data from subsequent sessions for all four subjects.

We compare the computational load, in terms of the number of training samples, of the incremental learning is compared with that of the batch mode. Figure 4.11 shows the number of training samples for one subject on both incremental learning and batch learning. The other three subjects have similar results. As one can be seen in this plot the incremental learning using only SVs, which is a small fraction of all data, substantially decreases the computational load after the aggregated data grow over time. The number of training samples for the incremental learning is less than $1/3$ that for the batch learning.

We implement the incremental learning SVM by keeping only the support vectors, instead of all the data, from the previous sessions and incorporating them with the data of the current session. Thus the incremental learning dramatically reduces the computational load. Since only the SVs contribute to the decision boundary, the incremental learning and the batch learning produce similar SV coefficients. The results demonstrate that the incremental learning ERP detection system performs as well as the batch mode, which uses all training data. Furthermore, it is more computationally efficient, which allows it to better cope with a continuous stream of EEG data.

4.13 Summary

We report the design and performance of a brain computer interface framework for visual target image search using single-trial ERP associated with human perceptual judgements. We demonstrate an ERP-based target detection system for speeding up visual target search by tapping into the split-second perceptual judgments of humans. Our ERP-based approach provides a higher detection rate and speed, albeit with more false alarms, than the tradition image viewing approach. It should be noted that many of the false positives can be removed with limited effort by manual confirmation of the prospective targets. The linear and nonlinear SVM classifiers significantly outperforms the LLC for single-trial ERP detection due to good generalization capabilities of SVM for high dimensional data. Therefore we select SVM as one of major ERP detectors. The channel dimension reduction using LDA successfully extracts representative features and thus improves the ERP detection performance with less computational cost.

We propose cross-session training for single-trial ERP detection and demonstrate the efficacy of incremental learning on cross-session data. Our cross-session results show that the high inter-session variance can be reliably mitigated. With more training samples, the cross session methods outperform the single-session method. Our motivation of using incremental learning is based on the fact that the SVM algorithm is able to summarize the data in the compact form of SVs. The incremental learning method using only SVs performs better than the naive method and it achieves, with a substantially lower computational load, a performance similar to the batch method for cross-session ERP detection. The incremental learning performs as well as the batch mode due to the fact that only the SVs contribute to the decision boundary. Since the incremental learning compacts the previous training data to the SVs and incorporates only the SVs with the new dataset, it is more computationally efficient than the batch learning method. We show that even though incremental learning is as effective as batch learning, the computational complexity is only $1/3$ that of batch learning (measured in terms of number of training samples).

This work represents a first step towards our vision of neurophysiologically-driven image triage system. The SVM as a discriminative learning approach optimizes a mapping from the inputs to the desired outputs by adjusting the classification boundary and thus demonstrate superior performance in the ERP detection task. However the SVM approach does not take advantage of the domain knowledge. Next we will investigate the characteristics of ERP patterns and explore some statistical models. As opposed to SVM, generative learning models , such as the mixed-effects models [62], are capable to model the underlying distributions of the variables and manipulate the probability density model to compute decision functions and thus provide a rich framework to learn useful and detailed modeling of ERP phenomenon from the prior knowledge. Besides, The use of more sophisticated kernels will be explored in the following chapter to deliver higher performance.

Chapter 5

A Hybrid Generative/Distriminative Model for Single-trial ERP Detection

5.1 Introduction

[1]Single-trial ERP detection is a fundamental signal processing problem for noninvasive BCI operating synchronously with some stimulus sequence. The problem needs to be solved effectively and efficiently so that the technique of single-trial ERP detection can become a practical solution to assistive and augmentative communication needs of persons with disabilities. In most assistive communication device applications the role of the BCI is to act as a keyboard-substitute. Recently, applications of noninvasive BCIs has been extended to utilize similar visual presentation paradigms where the displayed letters are replaced by images of interest to achieve goals related to particular tasks involved. For instance, image retrieval from large databases for medical and civilian research, and planning purposes, as well as for image mining in the web for recreation, are potential applications that will create a broader impact on the society.

The research that utilize the visual evoked potentials in various visual presentation

[1]This work was accepted as a book chapter, and will be originally published in the Recent Advances in Biomedical Signal Processing (eds. J. Ramirez, J.M. Gorriz, E. Lang) in July 2010. Copyright 2010 Bentham Scientific Publishers. Permission was granted.

paradigms have primarily relied on multi-trial ERP detection to achieve practically acceptable accuracy levels. For instance, the well known P300 Speller used $8-16$ repetitions typically [136], while G.tec claimed accurate-enough detection for brain-controlled typing using only 2-repetitions in well-trained subjects [36]. It is clear that, under the assumption of statistically independent EEG measurements in response to different presentations of the same stimulus, an exponentially diminishing ERP-detection error probability can be attained by simply multiplying the class likelihoods obtained from a single-trial ERP detector as in independent Bernoulli trials.

In statistical inference theory, the problem of signal detection in noise, when framed as a hypothesis test, one usually constructs a generative probabilistic model of the measured data, which then forms the basis for Bayesian inference [8] - optimal in terms of expected risk minimization. Deciding the class label (a target ERP (or 1) or a non-target ERP(or 0)) of a novel EEG waveform obtained in response to a particular visual stimulus could be achieved using the likelihood ratio test procedure. For instance: specifically, the ratio of likelihoods of the new data under the two competing generative models compared with a threshold reveals the optimal decision in this minimum expected risk framework (note that one might be interested in other statistics of the risk as well, in which case, different optimal decision rules would have been obtained) [26]. One particular strength of generative probabilistic models is that it provides an understanding of the underlying mechanisms for the process. Therefore, they contain more information that one can extract in addition to the optimal decision for the signal detection problem. On the negative side, this could also be interpreted as a weakness of the generative approach - the model attempts to capture more information than needed for the purpose, thus generalization might fail for complex classification boundaries with small amount of training data.

Discriminative learning in machine learning is an approach that focuses on learning only the decision boundary in a classification problem, in an attempt to avoid the shortcomings of a restrictive parametric generative model that tries to fit to the data throughout

the whole space. Linear discriminants, multi-layer perceptrons, SVM are among the examples of this approach [126]. They are trained to estimate a scalar discriminant index (which essentially corresponds to identifying a linear or nonlinear data projection function, and after which simple thresholding can reveal optimal decisions under the assumed model and criterion). Note that the Bayes minimum risk detector using true underlying class distributions is also effectively a nonlinear dimensionality reduction that allows for threshold comparison. While generative models attempt to approximate the individual class distributions, discriminative models attempt to model the surface in the feature space which will be mapped to the threshold value. Due to the reduced complexity of the function to be modeled, these latter models generally outperform the former approach in real world problems [50] where true distributions of underlying generative mechanisms are difficult to formulate (usually due to lack of understanding at the fundamental level or due to prohibitive computational complexity of such forward models in dynamic systems – both are influential in the case with brain signals).

The Fisher kernel was proposed for data with variable lengths where generative models can be formulated but discriminative learning is desired due to the inherent penalization of longer data [48, 47]. This technique provides a link between generative models and discriminative methods (although the technique refers to the use of a kernel, in fact, it is mathematically an alternative model-based distance metric, thus can be used in various discriminative methods if employed properly). Specifically, the intuition is that distances between pairs of data points should be measured using geodesics on the manifold induced by the generative probability density model. This approach is consistent with the information geometry of statistical models and mathematically superior to techniques that utilize Euclidean distances or other linear algebraic variations. The Fisher information matrix is known to form a natural Riemannian metric for a given parametric probability density model in its parameter space [1]. The Fisher kernel exploits this property of the Fisher information matrix and employs the Fisher score to construct an inner product that measures distances between datum pairs informed by the underlying generative model.

We propose a new hybrid generative/discriminative model using the Fisher kernel, in the area of single-trial ERP detection for the purpose of stimulus-synchronous BCI design. Specifically, we will utilize the mixed effects model (MEM) approach [62] (a graphical hierarchical Bayesian special case) to develop generative models for multichannel EEG signals, and use the Fisher kernel derived from the generative model in a discriminative model – SVM. The motivation of applying the hybrid model for single-trial ERP detection is to exploit the detailed modeling information encoded in the MEM and apply it in the design of a similarity measure between input samples via the Fisher kernel to improve the performance. Recently, researchers have investigated the connection of the generative and discriminative learning paradigms and combined their complementary strengths [50]. Jaakkola et al. first introduced the Fisher kernel to create a generic mechanism for incorporating generative probability models, such as hidden Markov models, into discriminative models, such as SVMs, to solve the variable length problem of feature vectors [48]. My colleagues, Lu and Leen [69] developed hierarchical Fisher kernels using the MEMs for longitudinal time series analysis to deal with varying sequence lengths and varying sampling intervals for predicting cognitive decline in aging.

In this chapter, we develop a hybrid model combining the generative mixed models with the SVM using the Fisher kernel for single-trial ERP detection. The challenge of this domain is high dimensionality and scarcity of training data. The first section starts with the model description and parameter estimation for the MEM. To avoid numerical problems in high dimensional matrix computations, we design low-dimensional calculations utilizing low-rank matrix properties. This significantly reduces computational complexity. We also employ the channel dimension reduction for EEG data (the same strategy as described in Chapter 4 Section 4.4, which is critical for applying the model on high-dimensional EEG data. To take advantage of the underly information about the target ERP and non-target ERP, We construct the mixed effects evoked response model using principal component analysis (PCA) [26] for the population model and linear discriminant analysis (LDA) [26] for the individual models. The second section introduces the Fisher

kernel and Fisher scored, derived from the MEM, and the construction of the SVM using Fisher kernel. In the result section we provide the performance evaluation of likelihood ratio MEM test and Fisher kernel SVM (FKSVM) detectors in comparison to the linear SVM (LinearSVM) and Gaussian-kernel SVM (GKSVM) detectors for single-trial ERP detection on two datasets.

5.2 MEM for Stimulus-Synchronized EEG

A linear mixed effects model (MEM) [23] is a statistical hierarchical model with a complex, multilevel and hierarchical structure. It was first proposed for the analysis of longitudinal time-series data [62]. More detailed discussion of model fitting and checking was provided on analysis of unbalanced repeated measures and longitudinal data in the guideline paper [17]. It has been widely used on the analysis of longitudinal sequence data in biostatistics [23, 69], where each subject/sample yields a sequence of measurements. These measurement sequences are assumed to follow a temporal structure that consists of three components: (i) population contribution, which corresponds to the population average; (ii) individual variability component, which determines the random variation of an individual from the population mean; (iii) stationary measurement noise. The population and individual components are assumed to be linear combinations of basis functions of time and the measurement noise is usually assumed to be temporally white. There are two sources of variation in the MEM: between-individual variations and within-individual variation. It assumes that observations between levels or clusters are independent, but observations within each cluster are dependent. An advantage of the MEM is the ability to genuinely combine the data by introducing multilevel random effects. Therefore it is well suited for the analysis of complex clustered or longitudinal data, the data with multiple sources of variation, and biological variety and heterogeneity.

In BCI applications, for each visual stimulus, the corresponding ERP waveforms do not necessarily have exactly the same shape. Specifically, ERP components, such as

P300, may have variations in amplitude, duration, or latency from trial to trial. These variations might arise from a variety of factors, including fatigue and attention, task difficulty and stimulus complexity, as discussed in Chapter 3. In this study, we employ the MEM framework to develop two generative models for EEG waveforms in response to *target* and *non-target* visual stimuli. It is assumed in the BCI design that target stimuli results in active ERP generation and non-target images are largely ignored by the brain. The mixed-effects ERP detector models single-trial ERP waveforms as varying individuals from a population; thus the classifier attempts to explain fluctuations in the baseline ERP waveform via a hierarchical Bayesian topology. By introducing multilevel random effects, the MEM easily handles data with multiple sources of variation, such as EEG data. Specifically for designs of data across trials/sesseions/subjects, we can easily use the population-averaged parameters to specify the common EEG signal type (consistent pattern across trials/sesseions/subjects), and the specific parameters to specify individuality (individual variety with the within- and between- trial/sesseion/subject variance). Therefore the MEM provides principled basis for combining population and individual characteristics of ERP signals.

5.2.1 Model Description

In the following, we will refer to the measured (vectorized) multichannel EEG response for a particular visual stimulus as an *individual* and the group of individuals that come from the same type of stimulus (i.e., target or non-target) as a *population*. For each individual i of N from population *c* (where $c \in \{1, 0\}$ denotes class membership: a target ERP or a non-target ERP), the MEM is written as:

$$\mathbf{y}_i^c = \mathbf{X}^c \boldsymbol{\alpha}^c + \mathbf{Z}\mathbf{b}_i^c + \boldsymbol{\varepsilon}^c, \tag{5.1}$$

where $i = 1, \ldots, N$, N is the number of target or non-target samples. In the MEM expression:

- $\mathbf{y}_i^c$ is an $n^c \times 1$ vector of observations for the ith individual and n_i is the number of observations for the ith individual.
- $\boldsymbol{\alpha}^c$ is an $p^c \times 1$ population effects coefficient vector.
- $\mathbf{X}^c$ is an $n^c \times p^c$ population design matrix (basis vectors for fixed effects).
- $\mathbf{b}_i^c$ is an $k^c \times 1$ individual random effects vector. These random effects vectors are assumed to have a hyper-distribution (e.g., zero-mean multivariate Gaussian with covariance $\mathbf{D}^c$: $\mathbf{b}_i^c \sim N(\mathbf{0}, \mathbf{D}^c)$).
- $\mathbf{Z}$ is an $n^c \times k^c$ individual design matrix (basis vectors for random effects).
- $\boldsymbol{\varepsilon}^c$ is an $n^c \times 1$ vector of independent and identically distributed (iid) noise with zero mean and positive definite within-individual covariance (typically, $\boldsymbol{\varepsilon}^c \sim N(\mathbf{0}, \sigma^{c2}\mathbf{I})$, where I denotes an $n^c \times n^c$ identity matrix.).

Thus, assuming that all distributions involved are Gaussian, the density model corresponding to (5.1) can be written as $\mathbf{y}_i^c \sim N(\mathbf{X^c}\boldsymbol{\alpha}^c, \sigma^{c^2}\mathbf{I} + \mathbf{Z}\mathbf{D}^c\mathbf{Z}^T)$. In (5.1), $\mathbf{y}_i^c$ is the vectorized spatiotemporal stimulus-time-locked EEG measurement (for instance from 32 channels over the duration $0 - 500$ms following stimulus onset). In test mode, when class labels are not known, the superscript indicating class label is to be determined. In the same equation, $\mathbf{X}^c$ and $\mathbf{Z}$ are known design matrices (consisting of preselected basis vectors for population and individual effects in their columns). The parameters to be determined via model fitting using maximum likelihood estimation, for instance, are population-averaged parameter $\boldsymbol{\alpha}^c$, the covariance $\mathbf{D}^c$ of the random vectors $\mathbf{b}_i^c$, and the covariance of the additive background noise component $\boldsymbol{\varepsilon}^c$, specifically σ^{c2}, if the noise is assumed to be spatiotemporally white for each class (a target ERP and a non-target ERP).

5.2.2 Model Parameter Estimation

The maximum likelihood (ML) estimates of the MEM parameters ($\boldsymbol{\alpha}^c$, $\mathbf{b}_i^c$, σ^{c2} and $\mathbf{D}^c$) are assessed using available data with the Expectation-Maximization (EM) algorithm [24].

ML Estimation of $\boldsymbol{\alpha}^c$ and $\mathbf{b}_i^c$

For a given class, let $\mathbf{V}_i^c = \sigma^{c2}\mathbf{I} + \mathbf{Z}\mathbf{D}^c\mathbf{Z}^T$ denote $Cov(\mathbf{y}_i^c)$, the covariance of the measurement vectors from this class. If $\mathbf{V}_i^c$ was known ($\hat{\sigma^{c2}}$ and $\hat{\mathbf{D}}^c$ were known), we could estimate $\boldsymbol{\alpha}^c$ and $\mathbf{b}_i^c$. Assuming that the measured vector, $\mathbf{y}_i^c$, is independent and identically distributed according to the Gaussian model prescribed by MEM, the joint data likelihood would be given by

$$p(\mathbf{y}^c;\boldsymbol{\theta}^c) = \prod_{i=1}^{N} \frac{\exp[-\frac{1}{2}(\mathbf{y}_i^c - \mathbf{X}^c\boldsymbol{\alpha}^c)^T\mathbf{V}_i^{c-1}(\mathbf{y}_i^c - \mathbf{X}^c\boldsymbol{\alpha}^c)]}{(2\pi)^{\frac{n^c}{2}}|\mathbf{V}_i^c|^{\frac{1}{2}}}, \tag{5.2}$$

where $\boldsymbol{\theta}^c = (\boldsymbol{\alpha}^c; vec(\mathbf{D}^c); \sigma^c)$. For simplicity of model, $\mathbf{D}^c$ could be assumed to be diagonal, in which case, the parameter vector would only include the individual variances of the individual random effect coefficients. From this expression, the log-likelihood as a function of the parameter vector is obtained as

$$\begin{aligned} l(\boldsymbol{\theta}^c) &= -\frac{1}{2}\{Nn^c ln(2\pi) + \sum_{i=1}^{N}[ln|\mathbf{V}_i^c| \\ &\quad +(\mathbf{y}_i^c - \mathbf{X}^c\boldsymbol{\alpha}^c)^T\mathbf{V}_i^{c-1}(\mathbf{y}_i^c - \mathbf{X}^c\boldsymbol{\alpha}^c)]\}. \end{aligned} \tag{5.3}$$

If the covariance parameter estimates $\hat{\sigma}^{c2}$ and $\hat{\mathbf{D}}^c$ were available, then the log-likelihood function could be maximized by the generalized least squares estimator. Specifically, taking the derivative of $l(\boldsymbol{\theta}^c)$ with respect to $\boldsymbol{\alpha}^c$ and equating to zero, we get

$$\hat{\boldsymbol{\alpha}}^c = (\sum_{i=1}^{N}\mathbf{X}^{cT}\mathbf{V}_i^{c-1}\mathbf{X}^c)^{-1}\sum_{i=1}^{N}\mathbf{X}^{cT}\mathbf{V}_i^{c-1}\mathbf{y}_i^c. \tag{5.4}$$

Once an estimate for $\boldsymbol{\alpha}^c$ is available, we can treat $\mathbf{b}_i^c$ as fixed effects and obtain the estimate from (5.1)using least square estimation as follows:

$$\hat{\mathbf{b}_i^c} = \mathbf{D}^c\mathbf{Z}^T\mathbf{V}_i^{c-1}(\mathbf{y}_i^c - \mathbf{X}^c\hat{\boldsymbol{\alpha}}^c). \tag{5.5}$$

EM Estimates for σ^{c2} and $\mathbf{D}^c$

M-step: If we were to observe $\mathbf{b}_i^c$ and $\boldsymbol{\varepsilon}^c$, we could easily obtain a simple closed-form solution using ML estimates of variances,

$$\hat{\sigma^{c}}^2 = \frac{1}{Nn^c}\sum_{i=1}^{N}\boldsymbol{\varepsilon}^{cT}\boldsymbol{\varepsilon}^c, \tag{5.6}$$

$$\hat{\mathbf{D}^c} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{b}_i^c\mathbf{b}_i^{cT}. \tag{5.7}$$

E-step: If σ^{c2} and $\mathbf{D}^c$ estimates were available, we could calculate the sufficient statistics as follows:

$$\sum_{i=1}^{N}\boldsymbol{\varepsilon}^{cT}\boldsymbol{\varepsilon}^c = \sum_{i=1}^{N}\hat{\boldsymbol{\varepsilon}}^c(\hat{\boldsymbol{\theta}})^T\hat{\boldsymbol{\varepsilon}}^c(\hat{\boldsymbol{\theta}}) + \sum_{i=1}^{N}tr\{Cov[\boldsymbol{\varepsilon}^c|\mathbf{y}_i^c, \hat{\boldsymbol{\alpha}}^c(\hat{\boldsymbol{\theta}}^c), \hat{\boldsymbol{\theta}}^c]\}, \tag{5.8}$$

$$\sum_{i=1}^{N}\mathbf{b}_i^{cT}\mathbf{b}_i^c = \sum_{i=1}^{N}\{\hat{\mathbf{b}}_i^c(\hat{\boldsymbol{\theta}})^T\hat{\mathbf{b}}_i^c(\hat{\boldsymbol{\theta}}) + Cov[\mathbf{b}_i^c|\mathbf{y}_i^c, \hat{\boldsymbol{\alpha}}^c(\hat{\boldsymbol{\theta}}^c), \hat{\boldsymbol{\theta}}^c]\}, \tag{5.9}$$

where $\hat{\boldsymbol{\varepsilon}}^c(\hat{\boldsymbol{\theta}}^c) = \mathbf{y}_i^c - \mathbf{X}^c\hat{\boldsymbol{\alpha}}_i^c(\hat{\boldsymbol{\theta}}^c) - \mathbf{Z}\hat{\mathbf{b}}_i^c(\hat{\boldsymbol{\theta}}^c)$ and $\hat{\mathbf{b}}_i^c(\hat{\boldsymbol{\theta}}^c)$ were obtained from ML estimation. Based on $\boldsymbol{\varepsilon}^c|\boldsymbol{\theta}^c \sim N(\mathbf{0}, \sigma^{c2}\mathbf{I})$, $\mathbf{y}_i^c|\boldsymbol{\varepsilon}^c; \boldsymbol{\theta}^c \sim N(\mathbf{X}^c\boldsymbol{\alpha}^c, \mathbf{Z}\mathbf{D}^c\mathbf{Z}^T)$, and $\mathbf{y}_i^c|\boldsymbol{\theta}^c \sim N(\mathbf{X}^c\boldsymbol{\alpha}^c, \sigma^{c2}\mathbf{I} + \mathbf{Z}\mathbf{D}^c\mathbf{Z}^T)$, we can derive

$$Cov[\boldsymbol{\varepsilon}^c|\mathbf{y}_i^c, \hat{\boldsymbol{\alpha}}^c(\hat{\boldsymbol{\theta}}^c), \hat{\boldsymbol{\theta}}^c] = [(\mathbf{Z}\mathbf{D}^c\mathbf{Z}^T)^{-1} + (\sigma^{c2}\mathbf{I})^{-1}]^{-1}. \tag{5.10}$$

Similarly, based on $\mathbf{y}_i^c|\mathbf{b}_i^c; \boldsymbol{\theta}^c \sim N(\mathbf{X}^c\boldsymbol{\alpha}^c + \mathbf{Z}\mathbf{b}_i^c, \sigma^{c2}\mathbf{I})$, $\mathbf{y}_i^c|\boldsymbol{\theta}^c \sim N(\mathbf{X}^c\boldsymbol{\alpha}^c, \sigma^{c2}\mathbf{I} + \mathbf{Z}\mathbf{D}^c\mathbf{Z}^T)$, and $\mathbf{b}_i^c|\boldsymbol{\theta}^c \sim N(\mathbf{0}, \mathbf{D}^c)$ we can calculate

$$Cov[\mathbf{b}_i^c|\mathbf{y}_i^c, \hat{\boldsymbol{\alpha}}^c(\hat{\boldsymbol{\theta}}^c), \hat{\boldsymbol{\theta}}^c] = (\mathbf{Z}^T\mathbf{Z}/\sigma^{c2} + \mathbf{D}^{c-1})^{-1}. \tag{5.11}$$

Thus from (5.6)-(5.11), we obtain the variance parameter estimates as:

$$\begin{aligned}\hat{\sigma^{c}}^2 &= Nn^c\sum_{i=1}^{N}\hat{\boldsymbol{\varepsilon}}^c(\hat{\boldsymbol{\theta}}^c)^T\hat{\boldsymbol{\varepsilon}}^c(\hat{\boldsymbol{\theta}}^c) \\ &\quad + \frac{1}{Nn^c}\sum_{i=1}^{N}tr\{[(\mathbf{Z}\mathbf{D}^c\mathbf{Z}^T)^{-1} + (\sigma^{c2}\mathbf{I})^{-1}]^{-1}\},\end{aligned} \tag{5.12}$$

$$\hat{\mathbf{D}}^c = \frac{1}{N}\sum_{i=1}^{N}\{\hat{\mathbf{b}}_i^c(\hat{\boldsymbol{\theta}}^c)^T\hat{\mathbf{b}}_i^c(\hat{\boldsymbol{\theta}}^c) + (\frac{\mathbf{Z}^T\mathbf{Z}}{\sigma^{c2}} + \mathbf{D}^{c-1})^{-1}\}. \tag{5.13}$$

Upon convergence of the EM iterations, we obtain $\hat{\sigma}^{c2}$ and $\hat{\mathbf{D}}^c$.

5.2.3 Dimension Reduction in MEM Calculations

The model parameter estimation procedure provided in the previous section involves $n^c \times n^c$ matrix inversions and determinants. These computations can be reduced to $k \times k$ where $k \ll n^c$ using the following exact rank-reduction formulas. Since these reductions apply to models of both classes, we will omit the superscript indicating class label in the following expressions throughout this section.

Simplified Formulas for Log-likelihood

Since $\mathbf{V}_i = \sigma^2\mathbf{I} + \mathbf{Z}\mathbf{D}\mathbf{Z}^T$ involves an $n \times n$ matrix inversion, we can use the following dimension-reduction formulas to exploit the relevant rank-k subspace:

$$\begin{aligned}\mathbf{V}_i^{-1} &= \sigma^{-2}\mathbf{I}_n - \sigma^{-2}\mathbf{I}_n\mathbf{Z}(\mathbf{D}^{-1} + \mathbf{Z}^T\sigma^{-2}\mathbf{I}_n\mathbf{Z})^{-1}\mathbf{Z}^T\sigma^{-2}\mathbf{I}_n \\ &= \sigma^{-2}\mathbf{I}_k - \sigma^{-4}\mathbf{Z}^T\mathbf{Z}(\mathbf{D}^{-1} + \sigma^{-2}\mathbf{Z}^T\mathbf{Z})^{-1}, \end{aligned} \tag{5.14}$$

$$|\mathbf{V}_i| = \sigma^{2(n-k)}|\sigma^2\mathbf{I}_k + \mathbf{D}\mathbf{Z}^T\mathbf{Z}|. \tag{5.15}$$

If matrix $\mathbf{D}$ is nonsingular, we can have the log of the determinant as a function of $\mathbf{D}^{-1}$.

$$\ln|\mathbf{V}_i| = \ln|\sigma^2\mathbf{D}^{-1} + \mathbf{Z}^T\mathbf{Z}| - \ln|\mathbf{D}^{-1}| + (n-k)\ln\sigma^2. \tag{5.16}$$

Simplified formulas for σ^2 and $\mathbf{D}$

To avoid inverse matrices in Equation (5.10) and (5.11), by using matrix inversion lemma, we have the following simplifications,

$$[(\mathbf{Z}\mathbf{D}\mathbf{Z}^T)^{-1} + (\sigma^2\mathbf{I}_n)^{-1}]^{-1} = \sigma^2\mathbf{I}_n - \sigma^4\mathbf{I}_n\mathbf{V}_i^{-1}, \tag{5.17}$$

$$(\frac{1}{\sigma^2}\mathbf{Z}^T\mathbf{Z} + \mathbf{D}^{-1})^{-1} = \mathbf{D} - \mathbf{D}\mathbf{Z}^T\mathbf{V}_i^{-1}\mathbf{Z}\mathbf{D}. \tag{5.18}$$

Therefore Equation (5.12) and (5.13) can be simplified as follows

$$\hat{\sigma}^2 = \frac{1}{Nn}\sum_{i=1}^{N}\hat{\boldsymbol{\varepsilon}}(\hat{\boldsymbol{\theta}})^T\hat{\boldsymbol{\varepsilon}}(\hat{\boldsymbol{\theta}}) + \sigma^2 - \frac{1}{Nn}\sigma^4\sum_{i=1}^{N}tr(\mathbf{V}_i^{-1}) \tag{5.19}$$

$$\hat{\mathbf{D}} = \frac{1}{N}\sum_{i=1}^{N}[\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})^T\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})] + \mathbf{D} - \frac{1}{N}\mathbf{D}(\sum_{i=1}^{N}\mathbf{Z}^T\mathbf{V}_i^{-1}\mathbf{Z})\mathbf{D}. \tag{5.20}$$

Using (5.14), we also obtain

$$\mathbf{Z}^T\mathbf{V}_i^{-1}\mathbf{Z} = \mathbf{Z}^T\mathbf{Z}(\sigma^2\mathbf{I}_k + \mathbf{D}\mathbf{Z}^T\mathbf{Z})^{-1}. \tag{5.21}$$

If $(\mathbf{Z}_i^T\mathbf{Z}_i)^{-1}$ exists, we can have

$$\mathbf{Z}^T\mathbf{V}_i^{-1}\mathbf{Z} = [\sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1} + \mathbf{D}]^{-1}. \tag{5.22}$$

Furthermore from (5.20),

$$\sum_{i=1}^{N}\mathbf{Z}^T\mathbf{V}_i^{-1}\mathbf{Z} = \sigma^{-2}\sum_{i=1}^{N}\mathbf{Z}^T\mathbf{Z} - \sigma^{-4}\sum_{i=1}^{N}[(\mathbf{Z}^T\mathbf{Z})(\mathbf{D}^{-1} + \sigma^{-2}\mathbf{Z}^T\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{Z})^T]. \tag{5.23}$$

Simplified formulas for $\boldsymbol{\alpha}$ and $\mathbf{b}_i$

In (5.4) we can substitute and have

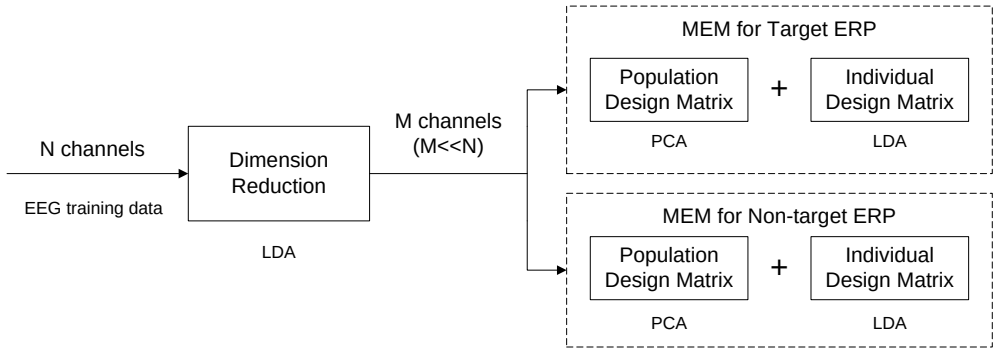
$$\sum_{i=1}^{N}\mathbf{X}^T\mathbf{V}_i^{-1}\mathbf{X} = \sigma^{-2}\sum_{i=1}^{N}\mathbf{X}^T\mathbf{X} - \sigma^{-4}\sum_{i=1}^{N}[(\mathbf{X}^T\mathbf{Z})(\mathbf{D}^{-1} + \sigma^{-2}\mathbf{Z}^T\mathbf{Z})^{-1}(\mathbf{X}^T\mathbf{Z})^T], \tag{5.24}$$

$$\sum_{i=1}^{N}\mathbf{X}^T\mathbf{V}_i^{-1}\mathbf{y}_i = \sigma^{-2}\sum_{i=1}^{N}\mathbf{X}^T\mathbf{y}_i - \sigma^{-4}\sum_{i=1}^{N}[(\mathbf{X}^T\mathbf{Z})(\mathbf{D}^{-1} + \sigma^{-2}\mathbf{Z}^T\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{y}_i)]. \tag{5.25}$$

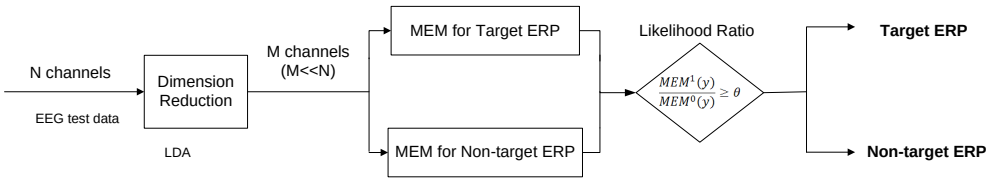
Similarly, in (5.5) we can substitute

$$\mathbf{Z}^T\mathbf{V}_i^{-1}\mathbf{y}_i = \sigma^{-2}\mathbf{Z}^T\mathbf{y}_i - \sigma^{-4}(\mathbf{Z}^T\mathbf{Z})(\mathbf{D}^{-1} + \sigma^{-2}\mathbf{Z}^T\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{y}_i), \tag{5.26}$$

$$\mathbf{Z}^T\mathbf{V}_i^{-1}\mathbf{X} = \sigma^{-2}\mathbf{X}^T\mathbf{Z}^T - \sigma^{-4}(\mathbf{Z}^T\mathbf{Z})(\mathbf{D}^{-1} + \sigma^{-2}\mathbf{Z}_i^T\mathbf{Z})^{-1}(\mathbf{X}^T\mathbf{Z})^T. \tag{5.27}$$



(a) MEM Training



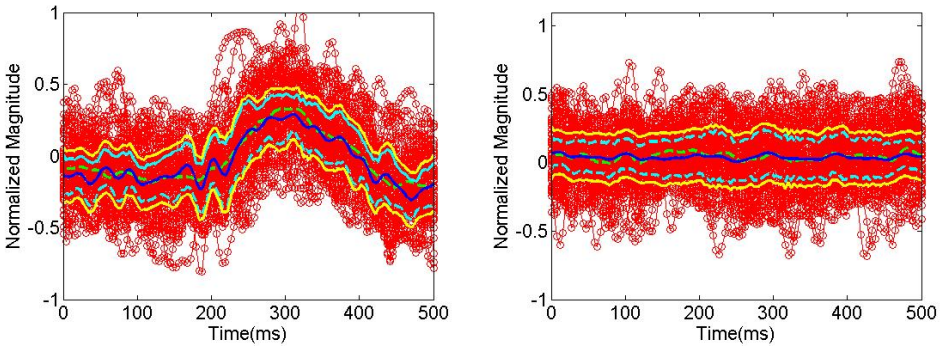
(b) MEM Test

Figure 5.1: The framework of MEM likelihood-ratio ERP detection. (a) MEM training procedures. Two MEMs are constructed, one for the ERPs associated with targets, and one for the ERPs associated with non-targets. (b) MEM likelihood ratio test.

5.2.4 MEM Likelihood-Ratio Test ERP Detector

Given one trained MEM per class, whose likelihood values for a given $\mathbf{y}$ are denoted by $MEM^c(\mathbf{y})$, one can design an ERP detector using the standard likelihood ratio test approach. Figure 5.1 shows the training and test processes of the MEMs for ERP detection. We construct two linear MEMs, one for the ERP cluster and another one for the nonERP cluster. The likelihood ratio test is simply the ratio of these two models, $\frac{MEM^1(\mathbf{y})}{MEM^0(\mathbf{y})}$, for a given test sample. Figure 5.2 shows the model fit for these two classes. The plots shows the fixed effect regression and the standard deviations from the random effects and noise.

We employ a linear channel mixture paradigm inspired by the LDA for preliminary dimensionality reduction (as described in Section 4.4) followed by the training of individual class MEM parameters based on available training data for each class. The number of channels (spatial locations on the scalp) from which EEG is acquired might be high if denser arrays are utilized. This factor directly affects the dimensionality of the measurement vector being modeled in the MEM framework and dimensionality reduction



(a) MEM fit for ERPs (b) MEM fit for non ERPs

Figure 5.2: The MEM model fit for (a) the ERPs associated with targets and (b) the ERPs associated with non-targets from the stimulus onset to 500ms latency. The red spaghetti curves are the input EEG signals. The green dash lines are the mean of the signals. The blue solid lines are the fixed effects $\mathbf{X}\boldsymbol{\alpha}$. The cyan dash lines are the population effects $\pm$ the standard deviation from the individual effects of the $\mathbf{b}_i$, i.e. $\mathbf{X}\boldsymbol{\alpha} + \sqrt{\mathbf{Z}\mathbf{D}\mathbf{Z}^T}$. The yellow solid lines are the population effects $\pm$ the standard deviation from the individual effects of the $\mathbf{b}_i$ and the noise term $\boldsymbol{\varepsilon}$, i.e. $\mathbf{X}\boldsymbol{\alpha} + \sqrt{\mathbf{Z}\mathbf{D}\mathbf{Z}^T + \sigma^2}$.

generally benefits the learning process from a parameter estimation variance perspective. We develop the population design matrices $\mathbf{X}^c$ for the ERPs associated with targets and non-targets respectively using the principle component analysis (PCA) [26] and the individual design matrix $\mathbf{Z}$ using the LDA. The intuition is that the population components in the models will attempt to capture the average large power trends in the signals of each class while the individual ERP variations will be modeled trying to exploit discriminative patterns.

Population and Individual Design Matrices

After some experimentation and cross validation, we decide to develop the population and individual design matrices for both target MEM and non-target MEM using the PCA and the the LDA, respectively [26]. This also intuitively means that the population components in the models will attempt to capture the average large power trends in the signals of each class while the individual ERP variations will be modeled trying to exploit discriminative patterns.

The population design matrices for the target ERP and non-target ERP classes are obtained as the largest few eigenvectors of the corresponding class sample covariation matrices (without subtracting the class averages as one would do in covariance calculations). Specifically, the number of eigenvectors retained for use as columns in the population design matrix is selected by a user-defined percentage of the total variation (sum of eigenvalues, or equivalently trace of the covariation matrix). The same percentage is used as the threshold for minimum retained energy for both classes/models.

The individual design matrices are developed using the LDA. Specifically the largest generalized eigenvectors of the within and between class scatter matrices are retained. Since the LDA approach uses data from both classes to select the projection directions, both models use the same individual design basis vectors.

Once the population and individual design matrices are selected, the maximum likelihood MEM parameters for each class can be obtained using the EM procedure. Figure 5.1(a) illustrates the overall block diagram of the MEM model for each class. The model order selection and parameter regularization can be obtain using cross validation [26].

MEM Operation in Testing Mode

In testing mode, for each incoming sample $\mathbf{y}_i^{Test}$, the MEM still needs to identify the best individual effect coefficient vector $\mathbf{b}_i^{c,Test}$. Specifically, for each test pattern, it is assumed that under MEM^c, the following generative model is accurate:

$$\mathbf{y}_i^{Test} = \mathbf{X}^c\boldsymbol{\alpha}^c + \mathbf{Z}\mathbf{b}_i^{c,Test} + \boldsymbol{\varepsilon}^c \tag{5.28}$$

where $\mathbf{b}_i^{c,Test} \sim N(\mathbf{0}, \mathbf{D}^c)$ and $\boldsymbol{\varepsilon}^c \sim N(\mathbf{0}, \sigma^{c2}\mathbf{I})$. Since we have

$$p(\mathbf{y}_i^{Test}|\boldsymbol{\alpha}^c, \mathbf{b}_i^{c,Test}) \sim N(\mathbf{X}^c\boldsymbol{\alpha}^c + \mathbf{Z}\mathbf{b}_i^{c,Test}, \sigma^{c2}\mathbf{I}) \tag{5.29}$$

we can maximize this posterior for each class and obtain the optimal individual random effect parameter $\mathbf{b}_i^{c,Test}$ for the test pattern. This yields:

$$\mathbf{b}_i^{c,Test^*} = \mathbf{D}^c\mathbf{Z}^T\mathbf{V}_i^{c-1}(\mathbf{y}_i^{Test} - \mathbf{X}^c\boldsymbol{\alpha}^c). \tag{5.30}$$

After we obtain $\mathbf{b}_i^{1,Test^*}$ for the target ERP model and $\mathbf{b}_i^{0,Test^*}$ for the non-target ERP model using appropriate design eigenvectors in (5.30), we can employ the likelihood ratio test using the respective model log-likelihood estimates:

$$\begin{aligned} l(\mathbf{b}_i^{Test^*}) &= \ln[N(\mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{b}_i^{Test^*}, \sigma^2\mathbf{I})] + \ln[N(\mathbf{0}, \mathbf{D})] \\ &= -\frac{1}{2\sigma^{c2}}\|\mathbf{y}_i^{Test} - (\mathbf{X}^c\boldsymbol{\alpha}^c + \mathbf{Z}\mathbf{b}_i^{c,Test^*})\|_2^2 \\ &\quad -\frac{1}{2}\mathbf{b}_i^{c,Test^*,T}\mathbf{D}^{c-1}\mathbf{b}_i^{c,Test^*} + ln(C_{\sigma^{c2}}) + ln(C_{\mathbf{D}^c}) \end{aligned}$$

where $C_{\sigma^{c2}}$ and $C_{\mathbf{D}^c}$ are normalization constants for noise and prior Gaussian densities. The discriminant value of the MEM (the estimates of target likelihood) is the difference between the log-likelihood values of the target ERP and non-target ERP models.

5.3 Fisher Kernel for SVM

The operation of SVM [126] (and any other nonparametric approach) relies heavily on the distance metric used in assessing how close or far two data points are. The distance is then monotonically related to an assumed similarity kernel (which represents an inner product in a corresponding high dimensional space determined by the eigenfunctions of the kernel selected). The Fisher kernel [48] is a particular similarity measure that is constructed using an underlying generative probabilistic model for the data. It is informed by the information geometry induced by this generative model and provides a local approximation based on the Riemannian geometry of the model. This distance metric is a natural choice for pairs of samples that are close to each other – for farther pairs, the distance is a coarse approximation, but in practice seems to provide sufficient performance.

Jaakkola et al. first introduced the Fisher kernel to create a generic mechanism for incorporating generative probability models such as hidden Markov models into discriminative models such as the SVMs to solve the variable length problem of feature vectors [48]. Jaakkola and his colleagues showed that using the Fisher kernel significantly improved over the previous methods on detecting remote protein homologies [47, 46]. Mika et al.

presented a fast training algorithm to enable the Fisher kernel algorithm viable for large datasets [75]. Tsuda et al. analyzed the statistical properties of the Fisher kernel [124]. The Fisher kernel has been successfully employed in many other applications afterwards, such as large scale web audio classification [76], document classification [111], tree structure data analysis [82], alphabet logical sequences analysis [57], standard object recognition [42], speech recognition [118], and cognitive decline detection [69].

To combine the generative model (MEM) with the discriminative model (SVM) for the purpose of improving the single-trial ERP detection performance, we propose a new ERP detector using the Fisher kernel derived from the MEM for the SVM. In Figure 5.2, it is clear that the MEM provides good fit for the target ERP and non-target ERP clusters. Instead of arbitrarily selecting the Gaussian kernel for the SVM, we use the Fisher kernel derived from the MEM for the SVM. The strength of this approach is that it combines the rich biological information encode in MEM with the discriminative power of the SVM algorithm. We use the target ERP and non-target ERP generative models offered by the MEM paradigm. Since in test mode the class label is not known, one option is to utilize a mixture of MEM models to derive the Fisher kernel. Another approach is to put the emphasis on similarities as measured by the target ERP model (or the non-target ERP model) depending on under which model we would like the similarities to be accurate. The Fisher kernel will then be utilized in the SVM formalism to achieve ERP detection. The Fisher information matrix in (5.32) is approximated by sample averaging over the training dataset.

The key idea here is to derive the Fisher kernel function from the MEM and use it to replace the Gaussian kernel in the SVM for ERP detection. First we train a MEM and use the MEM to map each new example into its Fisher score. Then we compute the Fisher kernel function on the basis of the Euclidean distance between the scores of the new sample and the training samples; Finally we can measure the discriminant value for the new example from the SVM discriminant function.

5.3.1 Fisher Kernel

The Fisher kernel operates in the parameter-gradient space of the generative model; specifically the gradient of the log-likelihood with respect to the model parameters. It utilizes information on how sensitive the parameters are to the generative model. For any data vector $\mathbf{y}_i$ and model parameters $\boldsymbol{\theta}$, the Fisher score is a row vector and which is defined as

$$\mathbf{U}_{\mathbf{y}_i} = \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_i|\boldsymbol{\theta}). \tag{5.31}$$

The Fisher Information matrix is defined as

$$\mathbb{I} = E_{p(\mathbf{y}_i|\boldsymbol{\theta})}\{\mathbf{U}_{\mathbf{y}_i}^T \mathbf{U}_{\mathbf{y}_i}\}, \tag{5.32}$$

where $E_{\mathbf{y_i}}\{\ \}$ is the expectation over $p(\mathbf{y}_i|\boldsymbol{\theta})$. The Fisher kernel is defined as

$$K_F(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{U}_{\mathbf{y}_i} \mathbb{I}^{-1} \mathbf{U}_{\mathbf{y}_j}^T, \tag{5.33}$$

where $\mathbf{y}_i$ and $\mathbf{y}_j$ are two data samples. Detailed information and properties of the Fisher kernel can be found in Jaakkola's paper [48] and Tsuda's paper [124].

5.3.2 Fisher scores derived from the MEM

Given the parametric density model of the observation from MEM (we use the ERP model MEM^1) the Fisher score is calculated from the corresponding log-likelihood as follows:

$$\begin{aligned} \mathbf{U}_{\mathbf{y}_i} &= \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_i|\boldsymbol{\theta}) \\ &= \nabla_{\boldsymbol{\alpha}} \log p(\mathbf{y}_i|\boldsymbol{\alpha}), \nabla_{vec(\mathbf{D})} \log p(\mathbf{y}_i|\mathbf{D}), \nabla_{\sigma} \log p(\mathbf{y}_i|\sigma^2), \end{aligned} \tag{5.34}$$

where the model parameters of the MEM $\boldsymbol{\theta} = (\boldsymbol{\alpha}; vec(\mathbf{D}), \sigma^2)$ and data samples obey $\mathbf{y}_i \sim N(\mathbf{X^1}\boldsymbol{\alpha}^1, {\sigma^1}^2\mathbf{I} + \mathbf{Z}^1\mathbf{D}^1\mathbf{Z}^{1T})$, for $i = 1, 2, ..., N$.

There are several options to derive the Fisher scores [46]. We use an MEM trained from ERPs associated with target examples to model a given ERP and we derive the Fisher score based on the MEM for ERPs. This intuition is that we would like to enhance

the discriminative power of the classifier. This is also based on the experimental results on our data. Alternative way is to use an MEM trained from the negative examples to derive the Fisher score as suggested in [69]. Another way of deriving the Fisher score is to construct a mixture MEM based on both the MEM for target ERPs and the MEM for non-target ERPs. We can calculate the probability density function for class c,

$$p(\mathbf{y}_i|c) = P_1 p(\mathbf{y}_i|c=1) + P_0 p(\mathbf{y}_i|c=0), \tag{5.35}$$

where $c = 1$ for ERPs and $c = 0$ for nonERPs, P_1 and P_0 are the prior probabilities for ERP class and nonERP class, respectively.

Fisher scores of parameter $\boldsymbol{\alpha}$

Fisher scores respective to the fixed effect parameter of the MEM $\boldsymbol{\alpha}$ is a $1 \times p$ row vector

$$\frac{\partial l}{\partial \boldsymbol{\alpha}} = \left[\frac{\partial l}{\partial \boldsymbol{\alpha}_1}, ..., \frac{\partial l}{\partial \boldsymbol{\alpha}_m}, ..., \frac{\partial l}{\partial \boldsymbol{\alpha}_p} \right] \tag{5.36}$$

where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_m, ..., \boldsymbol{\alpha}_p]^T$ is a column vector. Based on the log-likelihood expression, if we let $\mathbf{a} = \mathbf{y} - \mathbf{X}\boldsymbol{\alpha}$, where $\mathbf{y}$ is a concatenated column vector of all training samples and $\mathbf{X}$ is a concatenated population design matrix, we have

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\alpha}_m} &= -\frac{1}{2}\frac{\partial(\mathbf{a}^T\mathbf{V}^{-1}\mathbf{a})}{\partial \boldsymbol{\alpha}_m} \\ &= \mathbf{a}^T\mathbf{V}^{-1}\mathbf{X}_{:m} \end{aligned} \tag{5.37}$$

where $\mathbf{X}_{:m}$ denotes the m^{th} column of basis vectors and $\mathbf{V}^{-1}$ is the symmetric blockwise covariance matrix consisting of all covariances in the the Gaussian distributions $p(\mathbf{y}_i|\boldsymbol{\theta})$, we have

$$\frac{\partial l}{\partial \boldsymbol{\alpha}^{\boldsymbol{T}}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})^T\mathbf{V}^{-1}\mathbf{X} \tag{5.38}$$

Fisher scores of parameter D

The covariance matrix $\mathbf{D}$ of the random effects of the MEM is a $k \times k$ matrix, where k is the number of basis vectors used in individual random effect modeling. The Fisher scores

with respect to the entries of $\mathbf{D}$ are given by:

$$\frac{\partial l}{\partial \mathbf{D}} = -\frac{1}{2}\left[\frac{\partial \ln|\mathbf{V}|}{\partial \mathbf{D}} + \frac{\partial(\mathbf{a}^T\mathbf{V}^{-1}\mathbf{a})}{\partial \mathbf{D}}\right]. \tag{5.39}$$

For each entry (m, l) of $\mathbf{D}$, we have

$$\frac{\partial l}{\partial \mathbf{D}_{ml}} = -\frac{1}{2}\left[\frac{\partial \ln|\mathbf{V}|}{\partial \mathbf{D}_{ml}} + \frac{\partial(\mathbf{a}^T\mathbf{V}^{-1}\mathbf{a})}{\partial \mathbf{D}_{ml}}\right]. \tag{5.40}$$

The first part of (5.40) is

$$\begin{aligned}\frac{\partial \ln|\mathbf{V}|}{\partial \mathbf{D}_{ml}} &= \sum_{ij}\frac{\partial \ln|\mathbf{V}|}{\partial \mathbf{V}_{ij}}\cdot\frac{\partial \mathbf{V}_{ij}}{\partial \mathbf{D}_{ml}} \\ &= \sum_{ij}(\mathbf{V}^{-1})_{ij}\cdot(\mathbf{Z}\cdot\mathbf{E}_{ml}\cdot\mathbf{Z}^T)_{ij},\end{aligned} \tag{5.41}$$

where $\mathbf{E}_{ij}$ is an elementary matrix with only nonzero entry of 1 occurring at location (i, j). The second part of (5.40) is

$$\frac{\partial(\mathbf{a}^T\mathbf{V}^{-1}\mathbf{a})}{\partial \mathbf{D}_{ml}} = -\mathbf{a}^T(\mathbf{V}^{-1}\cdot\mathbf{Z}\cdot\mathbf{E}_{ml}\cdot\mathbf{Z}^T\cdot\mathbf{V}^{-1})\mathbf{a}. \tag{5.42}$$

Based on (5.41) and (5.42), (5.40) can be written as

$$\begin{aligned}\frac{\partial l}{\partial \mathbf{D}_{ml}} &= -\frac{1}{2}[\sum_{ij}(\mathbf{V}^{-1})_{ij}\cdot(\mathbf{Z}\cdot\mathbf{E}_{ml}\cdot\mathbf{Z}^T)_{ij} \\ &\quad -\mathbf{a}^T(\mathbf{V}^{-1}\cdot\mathbf{Z}\cdot\mathbf{E}_{ml}\cdot\mathbf{Z}^T\cdot\mathbf{V}^{-1})\mathbf{a}].\end{aligned} \tag{5.43}$$

Therefore (5.39) can be written as

$$\frac{\partial l}{\partial \mathbf{D}} = \sum_{ml}\mathbf{E}_{ml}\frac{\partial l}{\partial \mathbf{D}_{ml}}. \tag{5.44}$$

Fisher scores of parameter σ^2

Under the white spatiotemporal noise assumption, the noise covariance matrix is determined by the scalar σ^2, which is the noise variance in any spatiotemporal sample value. The Fisher score for this parameter is

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2}\left[\frac{\partial \ln|\mathbf{V}|}{\partial \sigma^2} + \frac{\partial(\mathbf{a}^T\mathbf{V}^{-1}\mathbf{a})}{\partial \sigma^2}\right]. \tag{5.45}$$

The first term is explicitly given by

$$\frac{\partial \ln |\mathbf{V}|}{\partial \sigma^2} = tr(\mathbf{V}^{-1}). \tag{5.46}$$

The second term is

$$\frac{\partial (\mathbf{a}^T \mathbf{V}^{-1} \mathbf{a})}{\partial \sigma^2} = -\mathbf{a}^T \cdot (\mathbf{V}^{-1})^2 \cdot \mathbf{a} \tag{5.47}$$

Therefore we can write (5.45) as

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2}[tr(\mathbf{V}^{-1}) - \mathbf{a}^T \cdot (\mathbf{V}^{-1})^2 \cdot \mathbf{a}] \tag{5.48}$$

Concatenating all of these terms in 5.38, 5.44, and 5.48, we obtain the Fisher score with respective to the overall parameter vectors for each data sample as

$$\mathbf{U}_{\mathbf{y}_i} = \left[\frac{\partial l}{\partial \boldsymbol{\alpha}}, \frac{\partial l}{\partial vec(\mathbf{D})}, \frac{\partial l}{\partial \sigma^2} \right]. \tag{5.49}$$

5.3.3 Fisher Information Matrix

By definition, the Fisher information matrix $\mathbb{I}$ entry (m,n) can be written as

$$\mathbb{I}_{m,n} = \int_{\mathbf{y}_i} \frac{\partial \log p(\mathbf{y}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_m} \frac{\partial \log p(\mathbf{y}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_n} p(\mathbf{y}_i; \boldsymbol{\theta}) d\mathbf{y}_i. \tag{5.50}$$

The exact analytical calculation of the Fisher information matrix under the expectation with respect to the MEM might be infeasible or cumbersome. Assuming that the MEM is an accurate approximation of the true underlying data distribution, we employ sample averaging over the training data to obtain an approximate expression for this matrix,

$$\hat{\mathbb{I}}_{tr} = \frac{1}{N_{tr}} \sum_{k=1}^{N_{tr}} \mathbf{U}_{\mathbf{y_k^{tr}}} \mathbf{U}_{\mathbf{y_k^{tr}}}^T. \tag{5.51}$$

Other simplifications in the literature [48] suggested by Jaakkola include simply using the identity matrix in place of the Fisher information matrix. In this study we use the natural approximation for the matrix $\mathbb{I}$ by averaging over the training samples.

5.3.4 Fisher Kernel from MEM

Once the Fisher scores are available, they can be used to construct the Fisher kernel using the Mahalanobis inner product with the Fisher information matrix as the scaling matrix as in (5.52). Here we use linear Fisher kernel. The Fisher kernel between any two samples $\mathbf{y_i}$ and $\mathbf{y_j}$, is finally given by

$$K_F(\mathbf{y_i}, \mathbf{y_j}) = \frac{1}{\rho}\mathbf{U}_{\mathbf{y_i}}\hat{\mathbb{I}}_{tr}^{-1}\mathbf{U}_{\mathbf{y_j}}^T, \tag{5.52}$$

where in training both of these samples are training samples and in test mode one is a support vector sample and the other is a test sample. In 5.52, ρ is scaling constant. The Fisher kernels above can be used in the SVM formalism as a replacement for the commonly used Euclidean/Mahalanobis similarity measures. It may be advantageous to search in the feature space of quadratic decision boundaries for non linear separable examples by using the quadratic Fisher kernel

$$K_F(\mathbf{y_i}, \mathbf{y_j}) = (1 + \mathbf{U}_{\mathbf{y_i}}\hat{\mathbb{I}}_{tr}^{-1}\mathbf{U}_{\mathbf{y_j}}^T)^2, \tag{5.53}$$

or Gaussian Fisher kernel

$$K_F(\mathbf{y_i}, \mathbf{y_j}) = \exp[-\frac{1}{2\sigma^2}(\mathbf{U}_{\mathbf{y_i}} - \mathbf{U}_{\mathbf{y_j}})\hat{\mathbb{I}}_{tr}^{-1}(\mathbf{U}_{\mathbf{y_i}} - \mathbf{U}_{\mathbf{y_j}})^T], \tag{5.54}$$

where σ is a scaling parameter.

5.4 Experiments

5.4.1 Data Description

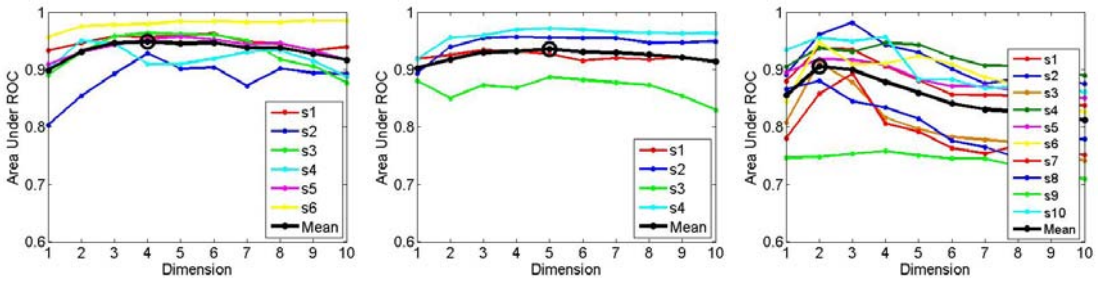
We assess dimension reduction on four datasets (Group #1 (IA Dataset #1, Group #2 (IA Dataset #2), Group #3 (naive dataset #1), Group #5 (naive dataset #3) as described in Section 2.1 and Section 2.5). We evaluate the other cross-session ERP detection performance on two IA datasets with each three subjects Group #1 (IA Dataset #1) and Group #2 (IA Dataset #2). For IA Dataset #1, there are three subjects and each has

one training session and seven test sessions. For IA Dataset #2, we exclude one subject because this subject only has one test session with single target. The other two IA subjects each has one training session. One subject has five test session and the other one has seven test sessions. The feature dimension is 129 times the reduced channel number (such as 4), selected by 10-fold cross validation. There are dozens of ERP training samples and hundreds of nonERP training samples for each subject.

5.4.2 Performance Evaluation

For performance comparison, we examine the four classifiers : (i) the MEM likelihood ratio test (simplified as MEM) between the MEM for ERPs and the MEM for nonERPs, (ii) the SVM using linear Fisher kernel (simplified as FKSVM), derived from the MEM for ERPs, (iii) the linear kernel SVM (simplified as LinearSVM), (iv) the nonlinear SVM using Gaussian kernel (simplified as GKSVM). We use the overall areas under ROC curves (AUCs) to depict the ERP detection performance. The overall ROC for each subject is constructed by congregating multiple test session outputs from the classifiers. We use a non-parametric approach to evaluate the correlated AUCs by calculating the statistical significance as describe in Chapter 4 Section 5.4.

We employ 10-fold cross-validation [26] on each subject for parameter regularization. Considering small samples of training data and high feature dimensionality, we select the model order of MEM equal one in this study for reliable model fitting. The optimal number of channel – LDA generalized eigenvectors – in initial dimension reduction, the percentage of energy retained in the population design matrices, and the scaling constant and cost parameter in the FKSVM are selected using exhaustive search within discrete sets of values. Cross-validation performance measure utilized for these assessments is the average of the the AUC estimates within the 10-fold validation framework. We also employ the same approach for parameter regularization in GKSVM training when obtaining baseline performance results for comparisons. These parameters include the kernel width for the isotropic Gaussian kernel and the overlap penalty parameter in its training.



(a) Dataset 1&2 (b) Dataset 3 (c) Dataset 4

Figure 5.3: Dimension order selection. We apply the LDA for dimension reduction for the MEM classifier on four datasets: (a) Group #1 (IA Dataset #1) and Group #2 (IA Dataset #2), (b) Group #3 (naive dataset #1), (c) Group #5 (naive dataset #3). We evaluate 10-fold cross validation performance in term of area under ROC curve on the training sessions for each subject. The color thin curves represent the detection performance for each subject. The thick lines represent the averaged performance across all subjects. The black circle represents the optimal order at the highest performance point. It is clear that the optimal number of channel projection dimension is much lower than the the original number of channel dimensions (32 for (a) and (b), 64 for (c)) across all subjects for four datasets.

5.4.3 Results

Dimension Reduction

To select dimension order of dimension reduction for the MEM, we apply 10-fold cross validation on the training session for each subject. We use the LDA approach for reduce channel dimensionality on four datasets (Group #1 (IA Dataset #1, Group #2 (IA Dataset #2),Group #3 (naive dataset #1),Group #5 (naive dataset #3)as described in Section 2.1 and Section 2.5). The dimension reduction performance is tested on the MEM classifier. The original channel number is 32 for two IA datasets and naive dataset #1. The original channel number is 64 for naive dataset #3. We first run similar dimension reduction procedure as Section 4.4 to select a few discrete numbers in the range from $1-32$ or $1-64$ and find that the best range is $1-10$ for all datasets. Therefore we conduct the dimension order selection in this range. Figure 5.3 shows the 10-fold cross validation performance in term of area under ROC for four datasets. The results show that the optimal dimension order averaged across subjects for these groups of subjects are $4, 5, 2$ respectively. It is clear that the optimal number of channel projection dimension is much

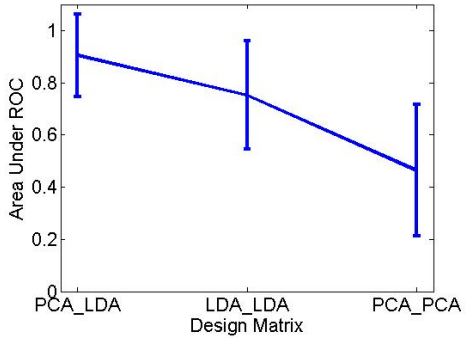


Figure 5.4: Detection performance in term of area under ROC curve for the MEM using three design matrix selections. The design matrix selections are: (1) the PCA for population design matrix and the LDA for individual design matrix, (2) the LDA for population design matrix and the LDA for individual design matrix, and (3) the PCA for population design matrix and the PCA for individual design matrix. The MEMs are applied to two IA datasets, Group #1 (IA Dataset #1) and Group #2 (IA Dataset #2), which consist of 6 subjects and 33 test sessions. The performance are averaged across 33 test sessions. It is clear that the first design matrix selection (the PCA for population design matrix and the LDA for individual design matrix) achieves the best performance in average.

lower than the the original number of channel dimensions (32 or 64) across all subjects for four datasets. These results further confirm the dimension reduction performance in Section 4.10 by using a difference classifier. It demonstrates that the dimension reduction using projection features produces better performance with much less computational cost.

Design Matrix Selection

To select the population design matrix and the individual design matrix for the MEMs, we compare the ERP performance using three design matrix selections. The performance in terms of area under ROC curves is evaluated on 33 test sessions from six subjects of two IA datasets , Group #1 (IA Dataset #1) and Group #2 (IA Dataset #2). Figure 5.4 shows the averaged performance across all test sessions for six subjects. One can see that the MEM with the PCA for population design matrix and the LDA for individual design matrix achieves the best performance in average. Therefore we use this design matrix selection for both the ERP MEM and the nonERP MEM.

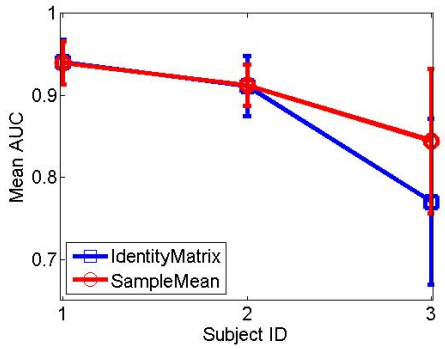


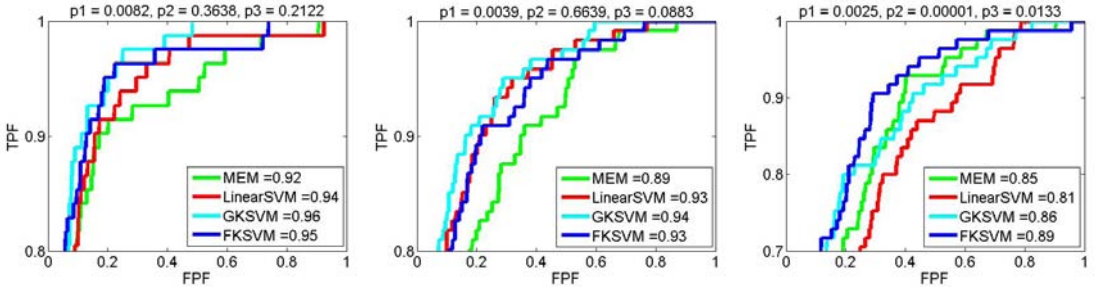
Figure 5.5: Comparison of Fisher Information Matrices based on the approximation of sample averaging over the training data and the identity matrix on three IA subjects, Group #1 (IA Dataset #1). The detection performance is in term of averaged AUC across seven test sessions for each subject.

Comparison of Fisher Information Matrices

We compare the Fisher information matrices based on the approximation of sample averaging over the training data and the identity matrix. Figure 5.5 shows the detection performance using the linear FKSVM based on the sample mean and the identity matrix. The detection performance is in term of averaged AUC across seven test sessions for each subject. The result shows the linear FKSVM using the sample mean has less variance and better performance in terms of the mean AUC than the model using the identity matrix on this dataset. Therefore in this study we use the natural approximation for the Fisher Information matrix $\mathbb{I}$ by averaging over the training samples.

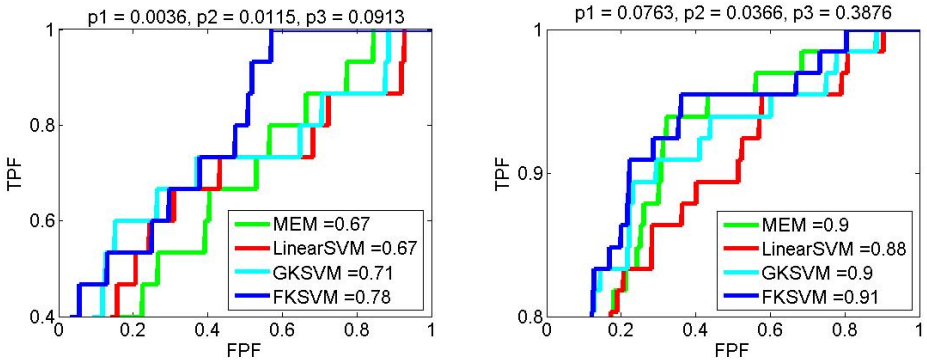
Comparison of Classifiers

We compare the hybrid model (the FKSVM) with the generative model (the MEM) and the discriminative models (the LinearSVM and the GKSVM). For each subject, using the reduced dimension data, four classifiers are trained using one training session. The ROC of each classifier as well as the corresponding overall AUC values over the corresponding test sessions for each subject (32 test sessions in total) for IA Dataset #1 and IA Dataset #2 are shown in Fig. 5.6 and Fig. 5.7 respectively. The significance levels of the hypothesis comparing the AUCs of the FKSVM to others are shown on the title



(a) Subject 1 (b) Subject 2 (c) Subject 3

Figure 5.6: Comparison of the ERP detection performance in terms of the overall area under ROC (AUC) between four classifiers on three subjects in Group #1 (IA Dataset #1). The p-values in the titles are based on pairwise correlated AUCs comparison between the FKSVM and the following classifiers, respectively: (p1) the MEM, (p2) the LinearSVM, (p3) the GKSVM.



(a) Subject 2 (b) Subject 3

Figure 5.7: Comparison of the ERP detection performance in terms of area under ROC (AUC) between four classifiers on two subjects in Group #2 (IA Dataset #2). The p-values in the titles are based on pairwise correlated AUCs comparison between the FKSVM and the following classifiers, respectively: (p1) the MEM, (p2) the LinearSVM, (p3) the GKSVM.

of each subgraph. This figures show that the FKSVM achieves substantial improvement over the original generative model (the MEM) in single-trial ERP detection. The FKSVM using the linear Fisher kernel has better performance than the linear SVM and has comparable performance as the GKSVM. On average across five subjects, the FKSVM outperforms the other classifiers with the mean values of the AUCs for the MEM, the LinearSVM, the GKSVM and the FKSVM as $0.846, 0.846, 0.874$, and 0.892 respectively. These results indicate that the FKSVM significantly outperforms the generative model (the MEM), provides better performance than the discriminative models (the LinearSVM and the GKSVM) in average.

5.5 Discussions

Discriminative classifiers have been more successful than their generative counterparts in many tasks since the task of the former is to model the lower dimensional classification boundary while the generative model needs to distribute its accuracy effort across the whole data space. In BCI literature, discriminative classifiers are quite popular for this reason, however, generative models of EEG signals also offer additional opportunities for future BCI algorithm development, separation of muscle and other artifacts from components of interest while detecting relevant brain signals within one generative model framework being the most important one.

We have developed a generative model for single-trial ERP responses using a hierarchical Bayesian model, referred to as an MEM. The MEM approaches the single-trial ERP detection problem by characterizing between trial variation in signal statistics. Our dimension(rank)-reduced version of the MEM significantly lowers the computational complexity from $O(n^2)$ to $O(k^2)$ $(k << n)$. We employ channel dimension reduction using the LDA to benefit the parameter estimation for the models. To capture the average large power trends in the signals of each class and to exploit discriminative patterns, we

develop the population design matrices for each MEM using the PCA and the individual design matrix using the LDA. Our experiments using cross-validation showed that in most datasets, using the largest generalized eigenvector give optimal generalization capability, while adding more basis vectors did not improve performance significantly. To estimate the gain from distinguishing random effect from observation noise in the MEMs, we compare the MEM to a a least square fitting, which is a simpler generative model that assumes no random effect. The least square model can be viewed as a special case of the MEM with constraint on random effect covariance $\mathbf{D} = 0$. Our experimental results show no significant different between the MEM and the the least square model. We suspect that some form of regularization effort needed for the MEM to improve the discriminant power of the random effect on the high dimensional data.

Fisher kernel formalism provides a way to incorporate information obtained from a generative model into the design of a kernel that can be utilized by an SVM classifier. We begin with a trained MEM and use this MEM to map each new data vector we wish to detect into its Fisher score. We then compute the Fisher kernel function on the basis of the Euclidean distance between the score vector for the new ERP example and the score vectors for training examples. Next we measure the discriminant value for this new ERP example from the discriminant function of the SVM. Jaakkola suggests [48] using the Fisher kernel found by setting the Fisher information matrix to an identity matrix for computational simplicity. However our data analysis show that the sample mean approximation in (5.51) for the Fisher information matrix produces better performance than an identity matrix.

The performance of the FKSVM was compared with that produced by the generative model (the MEM) and a competing discriminative models (the LinearSVM and the GKSVM). The likelihood ratio test based on the MEM, as expected did not outperform a well designed Gaussian kernel SVM. One of the reason of lower performance of the MEM might be the high dimension of EEG data ($4 * 129$) and small training samples (around 50

ERP samples). It indicates that the performance could be improved by some form of regularization. However, upon introducing Fisher kernels obtained from this model, we have obtained significant improvements on single-trial ERP detection accuracy over the original generative models (the MEM). The hybrid model also outperforms the LinearSVM and the GKSVM (the improvement is more significant in the fair comparison between the two linear kernels, the linear FKSVM and the LinearSVM), especially in subjects where the overall performance is lower. Clearly, for subjects where the performance is on the high end, improvements are also more difficult to obtain. The results show that the proposed method, which combines the MEM and SVM via the Fisher kernel, outperforms both the MEM and the SVM used in isolation for single-trial ERP detection in our data.

This work indicates the promising use of the mixed modeling techniques in the BCI design and the potential of the Fisher kernel formalism in SVM design. While the concept of our underlying MEM has been relatively simple, we believe that future work in this direction where the generative model will be developed using more rigorous signal propagation models such as those used in source localization. In that case, it could yield Fisher kernels that dramatically outperform generic kernels such as Gaussian in BCI design. This general approach of combining a generative model with a discriminative method may have applications in other areas of biosignal classification and analysis as well.

Chapter 6

Conclusions

6.1 Summary

The two main purposes of this thesis were to investigate ERP characteristics and to improve single-trial ERP detection performance. Specifically, the work enhanced our understanding on the neural signatures in human visual information perceptual system and we developed an effective ERP-based BCI that increased throughput in image analysis. In contrast to the previous studies [94, 40, 52, 96, 53, 39], in this thesis we used the ERP as a physiological marker to examine human visual perception and cognition in a dynamic recognition task. As opposed to the traditional ERP trial-averaging method, we built on Sajda's single-trial ERP detection approach, which used spatial integration to improve signal to noise ratio. We proposed an online BCI algorithm that used cross-session training and a novel, hybrid generative/discriminative classifier in attempt to address problems associated with high dimensionality, noise, non-stationarity and a dearth of training samples.

We provided an overview of BCI and the state-of-the-art developments in this emerging field. We also reviewed recent works related to neural signatures of visual information processing and the EEG-based BCI applications for visual object recognition. We described ERPs and their relationship with cognitive events, and discussed problems and solutions related to ERP detection on a single-trial basis. We described experimental procedures for acquiring high-quality EEG data and the pre-processing techniques that were

used to for EEG data analysis.

We presented insights on attention allocation in visual information processing by listing and describing several neural correlates of human cognitive processing that occurred in a dynamic task. We examined allocation processes used in visual detection as a function of both task difficulty (image duration) and target difficulty on real target detection tasks. Our results suggested that the cognitive visual processing system dynamically allocated attentional resources as a function of both target and task difficulty.

We reported the design and performance of a BCI system for target image search using single-trial ERP. The feasibility of cross-session training and the efficacy of incremental learning on cross-session data were demonstrated. We found that the two SVMs significantly outperformed the linear logistic classifier on high dimensional EEG data. In addition, we demonstrated that the LDA successfully extracted a subset of salient electrodes, which reduced the computational cost of the classifier. The MEM was a generative classifier that incorporated domain knowledge by modeling the underlying distributions of the variables and constructing a probability density from which decision functions for ERP detection were computed. We applied the MEM as a modeling tool and a classifier as well in ERP detection. The support vector machine (SVM) is a discriminative classifier, which is widely considered to be one of the best classifiers. In Chapter 5 we showed how to combine these two classifiers, the MEM and the SVM (the latter of which used the Fisher kernel), to produce a classifier that significantly outperformed the MEM alone and performed as well as or better than the SVM alone for single-trial ERP detection.

We found that there was high level of correlation between the underlying ERP characteristics and visual detection processes. Our single-trial ERP detection system, which used incremental learning, cross-session training and channel dimension reduction, was amenable for real-time implementations. The improved performance based on the hybrid generative / discriminative classifier enhanced the functionality of ERP-based BCIs.

6.2 Major Contributions

The major contributions of this work are:

- We developed a BCI that used cross-session training and incremental learning and showed that, for our data, the system performed well for single-trial ERP detection and it had low computational complexity. More specifically, our cross-session method outperformed the standard single-session method, in spite of the problems associated with intra-trial variability, and incremental learning had a computational load that was $1/3$ that of the standard SVM for the same level of detection performance.
- We developed a hybrid generative/discriminative classifier for single-trial ERP detection that had better detection performance then either system used in isolation. Morover, we used dimension-reduction within the MEM, which lowered the computational complexity from $O(n^2)$ to $O(k^2)$ where $(k << n)$.
- The channel dimension reduction using the LDA successfully extracts representative features and improved the ERP detection performance with less computational cost. This dimension reduction technique was validated on different classifiers including the linear and nonlinear SVMs, and the MEM.
- We quantified stimulus difficulty for complex real stimuli and subject performance. The stimuli was estimated as the minimum exposure duration necessary for subjects to detect a specified target correctly at a fixed detection rate. The subject performance was estimated as the image duration thresholds at a fixed detection rate of behavior performance averaged across all targets.
- We made several important discoveries concerning the dynamics of attention allocation that occured in a rapidly changing environment. First, EEG components occurring at 110ms and 250ms were related to both target and task difficulty. The target and task difficulty effects were more sensitive when behavioral performance

is poor. The earlier component was related to feature extraction and the later component was highly predictive of behavioral performance. The earlier component was more sensitive than the earlier component to the degree of attention. Based on our findings, we hypothesized that subjects needed to allocate additional neural resources, such as attention, to detect increasingly difficult targets or to detect targets in increasingly difficult tasks.

6.3 Future Work

We could extend this work in the direction of improving the single-trial ERP detection performance. In this work, we developed a method for single-trial ERP detection that used a learning method that approximated online learning. It would be beneficial for real-time applications to develop a BCI that uses (exact) online learning. Cauwenberghs' online learning method [14]), constructs the solution recursively, one point at-a-time. This work indicates the potential of MEM on ERP detection. Another avenue is to investigate the optimal configuration for the random effects model in the MEM classifier. For example, it would be worthwhile to explore the performance of MEM classifiers when multilevel random effects on trials, sessions or subjects are incorporated. We could develop a good model for the ERP signals. We envision using wavelet-based methods, such as Saito's local discriminant basis [107], to reduce the dimensionality in hopes of further enhancing the performance. To improve the overall ERP-based BCI performance, we could explore the effects of minimizing the error rate by detecting error related negativity (ERN) [138, 90], which is an EEG signal that is correlated with perceived error.

Here we only use the ERP as the input to the BCI for rapid image search. We have also investigated a fusion approach, which combines (behavioral) button presses and (neurological) ERPs [43]. Another set of potential inputs is gaze directions [110, 11, 101]. Eye gaze interaction can provide a convenient and natural addition to BCI dialogues [49, 98] The previous works demonstrated that the gaze information can be a useful source for

target image search. By combining the gaze features with the current source information, we can develop a fusion scheme to improve the target detection performance.

The experimental design we used for investigating the neural correlates to human visual information system is limited in that could not, because we used real images, carefully control the characteristics of the stimuli. We could use synthetic stimuli, which we could have multi-dimension controlling of the detectabilty, to confirm some of our findings. Synthetic stimuli may allow us to control the experiment with more precision and provides us more flexibility on the design of the perceptual task. In addition to analyzing neural responses associated with hits and true negatives, we could also assess the neural responses associated with false positives and misses. The analysis of these responses will help us understand the perceptual mechanism even better.

Bibliography

[1] AMAR, S. *Differential-Geometrical Methods in Statistics*. Kluwer Academic Publishers Group, Dordrecht, Netherlands, 2004.

[2] B. CAMPBELL, K., COURCHESNE, E., PICTON, T. W., AND SQUIRES, K. Evoked potential correlations of human information processing.

[3] BANDOS, A., ROCKETTE, H. E., AND GUR, D. A permutation test sensitie to differences in areas for comparing roc curves from a paired design. *Statistics in Medicine 24* (2005), 2873–2893.

[4] BERGER, H. Uber das electrenkephalogramm des menchen. *Archiv fur Psychiatrie Nervenkrankheiten 87* (1929), 527–570.

[5] BIGDELY-SHAMLO, N., VANKOV, A., RAMIREZ, R. R., AND MAKEIG, S. Brain activity-based image classification from rapid serial visual presentation. *IEEE Trans. on Neural Systems and Rehabilitation Engineering 16*, 5 (2008), 432–441.

[6] BIRBAUMER, N. Braincomputer interfaces research : coming of age. *Clinical Neurophysiology 117* (2006), 479–483.

[7] BIRBAUMER, N., KUBLER, A., GHANAYIM, N., HINTERBERGER, T., PERELMOUTER, J., KAISER, J., IVERSEN, I., AND KOTCHOUBEY, B. The thought translation device (ttd) for completely paralyzed patients. *IEEE Transactions on rehabilitation Engineering 8*, 2 (2000), 190–193.

[8] BISHOP, C. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, UK, 1995.

[9] BLANKERTZ, B., G.CURIO, AND MULLER, K. Classifying single trial eeg: Towards brain computer interfacing. *Advances in Neural Inf. Proc. Systems (NIPS 01) 14* (2002), 157–164.

[10] BROADBENT, D., AND BROADBENT, M. From detection to identification: response to multiple targets in rapid serial visual presentation. *Perception and Psychphysics 42*, 2 (1987), 105–113.

[11] BRUIJN, O. D., AND SPENCE, R. Patterns of eye gaze during rapid serial visual presentation. In *In Proceedings of Advanced Visual Interfaces (AVI 2002)* (Trento, Italy, 2002), pp. 209–217.

[12] BURGES, C. J. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2* (1998), 121–167.

[13] CATON, R. The electrical currents of the brain. *British Medical Journal 2* (1875), 278.

[14] CAUWENBERGHS, G., AND POGGIO, T. Incremental and decremental support vector machine learning. In *Proc. of the Conf. on Advances in neural information processing systems* (Vancouver, Canada, 2001), pp. 409–415.

[15] CHAPIN, J., MOXON, K., MARKOWITZ, R., AND NICOLELIS, M. Real-time control of a robot arm using simultaneous recorded neurons in the motor cortex. *Nature Neurosciences 2* (1999), 664–670.

[16] CHUN, M., AND POTTER, M. A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Huamn Perception and Performance 21*, 1 (1995), 109–127.

[17] CNAAN, A., LAIRD, N., AND SLASOR, P. Tutorial in biostatistics : Using the general linear mixed models to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine 16* (1997), 2349–2380.

[18] COLES, M., AND RUGG, M. *Event-related brain potentials: an introduction*. Oxford University press, Oxford, England, 1995, ch. 1 and 2.

[19] CORTES, C., AND VAPNIK, V. Support-vector networks. *Machine Learning 20* (1995), 273–297.

[20] CRISTIANINI, N., AND SHAWE-TAYLOR, J. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, UK, 2000.

[21] DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics 44* (1988), 837–845.

[22] Delorme, A., and Makeig, S. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics. *Journal of Neuroscience methods 134* (2004), 9–21.

[23] Demidenko, E. *Mixed Models Theory and Application*. John Wiley&Sons, Inc., New Jersey, NY, 2004.

[24] Dempster, A., Laird, N., and Rubin, D. Maximum likelihood with incomplete data via the e-m algorithm. *Journal of the Royal Statistical Society B39* (1977), 1–38.

[25] Donchin, E. The mental prosthesis: assessing the speed of a p300-based brain computer interface. *IEEE Transaction of Rehabilitation Engineering 8*, 2 (June 2000), 174–179.

[26] Duda, R., Hart, P., and Stork, D. *Pattern Classification*. Wiley, Inc., New Jersey, NY, 2001.

[27] Epstein, C. M., and Andriola, M. R. *An Introduction to EEG and Evoked Potentials*. J.B. Lippincott Company, Philadelphia, PA, 1983.

[28] Eriksen, C., and Spencer, T. Rate of information processing in visual percepton: some results and methodological considerations. *Journal of Experimental Psychology Monograph Supplement 79*, 2 (1969), Part 2.

[29] Fabiani, M., Gratton, G., Karis, D., and Donchin, E. Definition, identification, and reliability of measurements of the p300 component of the event-related brain potential. *Advances in Psychophysiology 2* (1988).

[30] Farwell, L., and Donchin, E. Talking off the top of your head: toward a mental prosthesis utilizing event-ralated brain potential. *Electroencephalograph & Clinical Neurophysiology 70*, 6 (1988), 510–523.

[31] Fawcett, T. Roc graphs: notes and practical considerations for data mining reserachers. Tech. Rep. HPL-2003-4, HP Laboratories, 2003.

[32] FORSYTH, D. A., AND PONCE, J. *Computer Vision: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2003.

[33] GALAMBOS, R., AND SHEATZ, G. An electroencephalographic study of classical conditioning. *American Journal of Physiology 203*, 0 (1962), 173–184.

[34] GERSON, A. D., PARRA, L. C., AND SAJDA, P. Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage 28*, 2 (2005), 342–353.

[35] GERSON, A. D., PARRA, L. C., AND SAJDA, P. Cortically-coupled computer vision for rapid image search. *IEEE Transactions on Neural Systems & Rehabilitation Engineering 14*, 2 (2006), 174–179.

[36] GUGER, C., DABAN, S., SELLERS, E., HOLZNER, C., KRAUSZ, G., CARABALONA, R., GRAMATICA, F., AND EDLINGER, G. How many people are able to control a p300-based brain-computer interface (bci)? *Neuroscience Letters 462*, 1 (2009), 94–98.

[37] HANLEY, J. A., AND MCNEIL, B. J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology 148* (1983), 839–843.

[38] HE, B. *Neural Engineering*. Kluwer Academic Publishers, New York, NY, 2005.

[39] HILLYARD, S., AND ANLLO-VENTO, L. Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences 95*, 3 (1998), 781–787.

[40] HILLYARD, S., AND KUTAS, M. Electrophysiology of cognitive processing. *Annual Review of Psychology 34* (1983), 33–61.

[41] HO, S., AND WECHSLER, H. Learning from data streams via online transduction. In *Proc. of the 4th IEEE Int. Conf. on Data Mining* (Brighton, UK, 2004).

[42] HOLUB, A., WELLING, M., AND PERONA, P. Combining generative models and fisher kernels for object recognition. In *Proc. of the 10th IEEE EMBS Int. Conf. on Computer Vision* (Washington DC, 2005), vol. 1, pp. 136–143.

[43] HUANG, Y., ERDOGMUS, D., MATHAN, S., AND PAVEL, M. A fusion approach for image triage using single trial erp detection. In *Proc. of the 3rd Int. IEEE EMBS Conf. on Neural Engineering* (Kohala Coast, HI, 2007), pp. 473–476.

[44] HUANG, Y., ERDOGMUS, D., PAVEL, M., HILD, K. E., AND MATHAN, S. Target detection using incremental learning on single-trial evoked response. In *Proc. of IEEE Intl. Conf. on ICASSP* (Taipei, Taiwan, 2009), pp. 481–484.

[45] HUK, A. C., AND N.SHADLEN, M. Neural activity in the macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *The Journal of Neuroscience 25*, 45 (2005), 10420–10436.

[46] JAAKKOLA, T., DIEKHANS, M., AND HAUSSLER, D. Using the fisher kernel method to detect remote protein homologies. In *Proc. of the 7th IEEE EMBS Int. Conf. on Intelligent Systems for Molecular Biology* (Menlo Park, CA, 1999), pp. 149–158.

[47] JAAKKOLA, T., DIEKHANS, M., AND HAUSSLER, D. A discriminative framework for detecting remote protein homologies. *Journal of Computational Biology 7*, 1-2 (2000), 95–114.

[48] JAAKKOLA, T., AND HAUSSLER, D. Exploiting generative models in discriminative classifiers. In *Proc. of the Conf. on Advances in neural information processing systems* (Denver,CO, 1998), pp. 487–493.

[49] JACOB, R., AND KARN, K. *Eye tracking in humancomputer interaction and usability research: Ready to deliver the promises*. Elsevier, London, UK, 2003, ch. The mind's eye: Cognitive and applied aspects of eye movements, pp. 573–05.

[50] JEBARA, T. *Machine Learning Distriminative and Generative*. Kluwer Academic Publishers Group, Dordrecht, Netherlands, 2004.

[51] JOACHIMS, T. Text categorization with support vector machines. In *Proc. of European Conference on Machine Learning* (1998), pp. 137–142.

[52] JOHNSON, J., AND B.A.OLSHAUSEN. Timecouse of neural signatures of object recognition. *Journal of Vision 3* (Sep 2003), 499–512.

[53] JOHNSON, J., AND OLSHAUSEN, B. The earliest eeg signatures of object recognition in a cued-target task are postsensory. *Journal of Vision 5*, 4 (2005), 299–312.

[54] JONES, R. Brain waves in phase. *Neuroscience 3*, 2 (2002), 167–167.

[55] J.R.WOLPAW. Braincomputer interfaces for communication and control. In *Proc. of the 9th Intl. ACM SIGACCESS Conf. on Computers and Accessibility* (Tempe,Arizona, 2007), pp. 1–2.

[56] KAPOOR, A., SHENOY, P., AND TAN, D. Combining brain computer interfaces with vision for object categorization. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* (Anchorage,AL, 2008).

[57] KERSTING, K., AND GARTNER, T. Fisher kernels for logical sequences. In *Proc. of the 15th European Conf. on Machine Learning* (Pisa, Italy, 2004), pp. 205–216.

[58] KEYSERS, C., XIAO, D.-K., FOLDIAK, P., AND PERRETT, D. The speed of sight. *Journal of Cognitive Neuroscience 13*, 1 (2001), 90–101.

[59] KIVINEN, J., SMOLA, A., AND WILLIAMSON, R. Online learning with kernels. In *Proc. of the Conf. on Advances in neural information processing systems* (Vancouver, Canada, 2001), pp. 785–792.

[60] KOSTOV, A., AND POLAK, M. Parallel man machine training in development of eeg based cursor control. *IEEE Transacktions on Rehabilitation Engineering 8*, 2 (2000), 203–205.

[61] KUBLER, A., KOTCHOUBEY, B., AND WOLPAW, J. Brain computer communication: Unlocking the locked in. *Psychlogical Bulletin 127*, 3 (2001), 358–375.

[62] LAIRD, N., AND WARE, J. Random-effect models for longitudinal data. *Biometrics 38* (1982), 963–974.

[63] LASKO, T. A., BHAGWAT, J., ZOU, K. H., AND OHNO-MACHADO, L. The use of receiver operating chanracteristic curves in biomedical inofrmatics. *Journal of Biomedial Informatics 38* (2005), 404–415.

[64] LASKOV, P., AND MULLER, K. Incremental support vector learning: analysis, implementation and applications. *Journal of Machine Learning Research 7* (2006), 1909–1936.

[65] LASKOV, P., SCHAFER, C., AND KOTENKO, I. Intrusion detection in unlabeled data with quarter-sphere support-vector machines. In *Proc. of the Conf. on Detection of Intrusions and Malware & Vulnerability Assessment* (Dortmund, Germany, 2004), pp. 71–82.

[66] LAWRENCE, D. Two studies of visual search for word targets with controlled rates of presentation. *Perception and Psychphysics 10* (1971), 85–89.

[67] LEBEDEV, M. A., AND NICOLELIS, M. A. L. Brainmachine interfaces: past, present and future. *Trends in Neurosciences 29*, 9 (2006), 536–546.

[68] LOTTE, F., CONGEDO, M., LECUYER, A., LAMARCHE, F., AND ARNALDI, B. A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of Neural Engineering 4*, 0 (2007), R1–R13.

[69] LU, Z., AND LEEN, T. Hierarchical fisher kernels for longitudinal data. *Advances in Neural Inf. Proc. Systems 21* (2008), 1961–1968.

[70] LUCK, S. *An Introduction to the Event-Related Potential Technique*. The MIT Press, Cambridge, MA, 2005.

[71] MAKEIG, S., BELL, A., JUNG, T., AND SEJNOWSKI, T. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems 8* (1996), 145–151.

[72] MAKEIG, S., DELORME, A., WESTERFIELD, M., J.TOWNSEND, COURCHENSE, E., AND SEJNOWSKI, T. Electroencephalographic brain dynamics following visual targets requiring manual responses. *Public Library of Science Biology 2*, 6 (2004), 747–762.

[73] MAKEIG, S., WESTERFIELD, M., JUNG, T., ENGHOFF, S., AND J.TOWNSEND. Dynamic brain sources of visual evoked responses. *Science 295* (2002), 690–694.

[74] MATHAN, S., VERVERS, P., , ERDOGMUS, D., CARCIOFINI, J., HUANG, Y., DORNEICH, M., PAVEL, M., AND WHITLOW, S. Rapid image analysis using neural signal. In *Proc. of the 26th Int. Conf. for Human-Computer Interaction* (Florence, Italy, 2008).

[75] MIKA, S., SMOLA, A., AND SCHOLAKPF, B. An improved training algorithm for kernel fisher discriminants. In *Proc. of the 8th Int. Workshop on Artificial Intelligence and Statistics* (Key West, Florida, 2001), pp. 98–104.

[76] MORENO, P., AND RIFKIN, R. Using the fisher kernel method for web audio classification. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Istanbul, Turkey, 2000), pp. 2417–2420.

[77] MUCULLAGH, P., AND NELDER, J. A. *Generalized Linear Models*. Chapman & Hall, 1990.

[78] MULLER, K., ANDERSON, C., AND BIRCH, G. Linear and non-linear methods for brain-computer interfaces. *IEEE Trans. on Neural Sys. Rehab. Eng. 11*, 0 (2003), 165–169.

[79] MULLER, K., TANGERMANN, M., DORNHEGE, G., KRAULEDAT, M., G.CURIO, AND BLANKERTZ, B. Machine learning for real-time single-trial eeg-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods 167*, 1 (2008), 82–90.

[80] MURATA, N., KAWANABE, M., ZIEHE, A., MULLER, K., AND AMARI, S. On-line learning in changing environment with applications in supervised and unsupervised learning. *Journal of Neural Networks 15*, 4-6 (2002), 743–760.

[81] NELDER, J. A., AND WEDDERBURN, R. Generalized linear models. *Journal of the royal Statistical Society 135* (1974), 370–384.

[82] NICOTRA, L., MICHELI, A., AND STARITA, A. Fisher kernel for tree structured data. In *Proc. of the IEEE Int. Joint Conf. on on Neural Networks* (Budapest, Hungary, 2004), pp. 1917–1922.

[83] NOCEDAL, J., AND WRIGHT, S. *Numerical Optimization*. Springer, New York, NY, 1999.

[84] OKEN, B. S. *Evoked Potentials in Clinical Medicine*. LippincottRaven, Philadelphia, PA, 1997, ch. 15.

[85] OSUNA, E., FREUND, R., AND GIROSI, F. Support vector machines: Training and applications. Tech. Rep. AIM-1602, 1997.

[86] OSUNA, E., FREUND, R., AND GIROSI, F. Training support vector machines: an application to face detection. In *Proc. of IEEE Computer Society Int. Conf. on Computer Vision and Pattern Recognition* (San. Juan, Puerto Rico, 1997), p. 1306.

[87] PARRA, L., ALVINO, C., TANG, A., PEARLMUTTER, B., AND YEUNG, N. Linear spatial integration for single-trial detection in encephalography. *NeuroImage 17* (2002), 223–230.

[88] PARRA, L., ALVINO, C., TANG, A., PEARLMUTTER, B., YEUNG, N., OSMAN, A., AND SAJDA, P. Single-trial detection in eeg and meg: Keeping it linear. *Neurocomputing 52-54* (2003), 177–183.

[89] PARRA, L., CHRISTOFOROU, C., GERSON, A., DYRHOLM, M., LUO, A., WAGNER, M., PHILIASTIDES, M., AND SAJDA, P. Spatio-temporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks. *IEEE Signal Processing Magazine Jan* (2008), 107–115.

[90] PARRA, L., SPENCE, C., GERSON, A., AND SAJDA, P. Response error correction a demonstration of improved human-machine performance using real-time eeg monitoring. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 11*, 2 (2003), 173–177.

[91] PARRA, L. C., SPENCE, C. D., GERSON, A. D., AND SAJDA, P. Recipes for the linear analysis of eeg. *Neuroimage 28*, 2 (2005), 326–341.

[92] PEPE, M. S. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, New York, NY, 2003.

[93] PFURTSCHELLER, G., NEUPER, C., MULLER, G., OBERMAIER, B., KRAUSZ, G., SCHLOGL, A., R. SCHERER, GRAIMANN, B., LEINRATH, C., SKLIRIS, D., WORTZ, M., SUPP, G., AND SCHRANK, C. Graz-bci: state of the art and clinical applications. *Psychlogical Bulletin 11*, 2 (2003), 117–180.

[94] PHILIASTIDES, M. G., RATCLIFF, R., AND SAJDA, P. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram.

[95] PHILIASTIDES, M. G., AND SAJDA, P. Temporal characterization of the neural correlates of perceptual decision making in the human brain.

[96] PICTON, T. The p300 wave of the human event-related potential. *Journal of Clinical Neurophysiology 9*, 4 (1992), 456–479.

[97] PICTON, T., BENTIN, S., BERG, P., DONCHIN, E., HILLYARD, S., JR., R. J., MILLER, G., RITTER, W., RUCHKIN, D., RUGG, M., AND TAYLOR, M. Guidelines for using human event-related potentials to study cognition: Recording starndards and publication criteria. *Psychophysiology 37* (2000), 127–152.

[98] POOLE, A., AND BALL, L. *Eye tracking in human-computer interaction and usability research: current status and future prospects*. Idea Group, Pennsylvania, 2005, ch. Encyclopedia of Human Computer Interaction.

[99] POTTER, M. Short-term conceptual memory for pictures. *Journal of Experimental Psychology 2*, 5 (1976), 509–522.

[100] POTTER, M., AND LEVY, E. Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology 81*, 1 (1969), 10–15.

[101] PRIVITERA1, C., RENNINGER, L., CARNEY, T., KLEIN, S., AND AGUILAR, M. The pupil dilation response to visual detection. *Human Vision and Electronic Imaging XIII 6806* (2008), 6806–6827.

[102] P.SAJDA, GERSON, A., PHILIASTIDES, M., AND PARRA, L. *IEEE Neural-Engineering Handbook*. Wiley/IEEE, 2006, ch. Single-trial analysis of EEG for Enabling Cognitive User Interfaces.

[103] P.SAJDA, GERSON, A., PHILIASTIDES, M., AND PARRA, L. *Towards Brain-Computer Interfacing*. MIT Press, 2007, ch. Single-trial analysis of EEG during rapid visual discrimination: Enabling cortically-coupled computer vision.

[104] RITTER, W., AND VAUGHAN, H. Average evoked responses in vigilance and discrimination. *Science 164* (1969), 326–328.

[105] ROBERTS, S., AND PENNY, W. Real-time brain computer interfacing: a preliminary study using bayesian learning. *Medical and Biological Engineering and Computing 38* (2000), 56–61.

[106] RUPING, S. Incremental learning with support vector machines. In *Proc. of the IEEE Int. Conf. on Data Mining* (San Jose, CA, 2001), pp. 641–642.

[107] SAITO, N., AND COIFMAN, R. Local discriminant bases and their applications. *Mathematical Imaging and Vision 5*, 4 (1995), 337–358.

[108] SAJDA, P., GERSON, A., AND PARRA, L. High-throughput image search via single-trial event detection in a rapid serial visual presentation task. In *Proc. of the 1st Intl. IEEE EMBS Conf. on Neural Engineering* (capri island, Italy, 2003), pp. 7–10.

[109] SAJDA, P., GERSON, A., AND PARRA, L. Spatial signature of visual object recognition events learned from single-trial analysis of eeg. In *Proc. of the 25th Intl. Conf. of the IEEE EMBS* (Cancun, Mexico, 2003), pp. 2087–2090.

[110] SALOJARVI, J., PUOLAMAKI, K., SIMOLA, J., KOVANEN, L., KOJO, I., AND KASKI, S. Inferring relevance from eye movements: Feature extraction. Tech. Rep. Technical Report A82, Computer and Information Science, Helsinki University of Technology, 2005.

[111] SAUNDERS, C., TAYLOR, J., AND VINOKOUROV, A. Document classification employing the fisher kernel derived from probabilistic hierarchic corpus representations. In *Proc. of the 23rd European Colloquium on Information Retrieval Research* (Darmstadt, 2001), pp. 24–40.

[112] SCHALL, J., AND HANES, D. Neural basis of saccade target selection in frontal eye field during visual search. *Nature 366* (1993), 467–469.

[113] SCHOLKOPF, B., AND SMOLA, A. *Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, NY, 2001.

[114] SCHOROGER, E., AND WIDMANN, A. Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology 35*, 0 (1998), 755–759.

[115] SCHWAIGHOFER, A. Svm toolbox for matlab (version 2.51 of january 2002). Tech. rep., 1997.

[116] S.DUMAIS, PLATT, J., HECKERMAN, D., AND SAHAMI, M. Inductive learning algortithms and respresentations for text categorization. In *Proc. of the 7th Intl. Conf. on Information and Knowledge Management* (Bethesda, Maryland, 1998), pp. 148–155.

[117] SHENOY, P., AND TAN, D. Human-aided computing: utilizing implicit human processing to classify images. In *Proc. of the 1st Intl. IEEE EMBS Conf. on Neural Engineering* (capri island, Italy, 2003), pp. 7–10.

[118] SMITH, N., AND GALES, M. Speech recognition using svms. In *Proc. of the Conf. on Advances in neural information processing systems* (Vancouver, Canada, 2002), pp. 1197–1204.

[119] SOLTANI, M., AND KNIGHT, R. Neural origins of the p300. *Critical Reviews in Neurobiology 14*, 3-4 (2001), 199–224.

[120] SUTTON, S., BRAREN, M., ZUBIN, J., AND JOHN, E. Evoked potential correlated of stimulus uncertainty. *Science 150* (1965), 1187–1188.

[121] SYED, N., LIU, H., AND SUNG, K. Incremental learning with support vector machines. In *Proc. of the Int. Joint Conf. on Artificial Intelligence* (Stockholm, Sweden, 1999).

[122] THORPE, S., FIZE, D., AND MARLOT, C. Speed of processing in the human visual system. *Nature 381* (1996), 520–522.

[123] TOMIOKA, R., AIHARA, K., AND MULLER, K. Logistic regression for single trial eeg classification. *Advances in Neural Inf. Proc. Systems (NIPS) 19* (2006), 1377–1384.

[124] TSUDA, K., AKAHO, S., KAWANABE, M., AND MULLER, K. Asymptotic properties of the fisher kernel. *Journal of Neural Computation 16*, 1 (2004), 115137.

[125] VANRULLEN, R., AND THORPE, S. The time course of visual processing: From early perception to decition-making. *Journal of Cognitive Neuroscience 13*, 4 (2001), 454–461.

[126] VAPNIK, V. *Statistical learning theory*. John Wiley & Sons, Inc., New York, NY, 1998.

[127] VAUGHAN, T., HEETDERKS, W., TREJO, L., RYMER, W., WEINRICH, M., MOORE, M., KBLER, A., DOBKIN, B., BIRBAUMER, N., E, E. D., WOLPAW, E., AND WOLPAW, J. Brain computer interface technology: a reviw of the second international meeting. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 11*, 2 (2003), 94–109.

[128] VIDAL, J. Towards direct braincomputer communication. *Annual Review of Biophysics and Bioengineering 2* (1973), 157–180.

[129] VIDAL, J. Real time detection of brain events in eeg. *Proceedings of the IEEE, Special issue on Biological Signal Processing and Analysis 65*, 5 (1977), 633–664.

[130] WALTER, W., COOPER, R., ALDRIDGE, V., MCCALLUM, W., AND WINTER, A. Contingent negative variation:a n electric sign of sensorimotor association and expectancy in the human brain. *Nature 203*, 0 (1964), 380–384.

[131] WOLPAW, J. Brainmachine interfaces for communication and control: a minireview. *Clinical Neurophysiology 57* (2004), 607–613.

[132] WOLPAW, J., AND BIRBAUMER, N. *neural repair and rehabilitation; neural repair and plasticity*. ambridge University Press, 2006, ch. Brain-computer interfaces for communication and control, pp. 602–614.

[133] WOLPAW, J., BIRBAUMER, N., HEETDERKS, W., MCFARLAND, D., PECKHAM, P., SCHALK, G., DONCHIN, E., QUATRANO, L., ROBINSON, C., AND VAUGHAN, T. Brain computer interface technology: a reviw of the first international meeting. *IEEE Trans on Rehabilitation Engineering 8*, 2 (2000), 164–172.

[134] WOLPAW, J., BIRBAUMER, N., MCFARLAND, D. J., PFURTSCHELLER, G., AND VAUGHAN, T. M. Braincomputer interfaces for communication and control. *Clinical Neurophysiology 113* (2002), 767–791.

[135] WOLPAW, J., AND MCFARLAND, D. Two dimensional movement control by scalp recorded sensorimoto rhythms in humans. *Proceedings of the National Academy of Sciences of the United States of America 101*, 51 (2004), 17849–17854.

[136] WOLPAW, J., MCFARLAND, D., VAUGHAN, T., AND SCHALK, G. The wadsworth center brain-computer interface research and development program. *IEEE Trans on Neural Systems & Rehabilitation Engineering 11*, 2 (2003), 204–7.

[137] WOODMAN, G., AND LUCK, S. Electrophysiological measurement of rapid shifts of attention during visual search. *Nature 400* (1999), 867–869.

[138] YEUNG, N., AND COHEN, J. The impact of cognitive deficits on conflict monitoring. predictable dissociations between the error-related negativity and n2. *Psychological Science 17* (2006), 164–171.

Biographical Note

Yonghong Huang was born on 6 December 1968 in Nanning China. She was joined by Zhong Lei in marriage in 1993. Their first child, Jordan was born in 1998 in China. Her family immigrated to Canada in 2000 and they became Canadian citizens in 2004. They moved to Portland in summer 2004 and her young son, Matthew was born in 2005.

She earned her Bachelor degree in Electrical Engineering at the South China University of Technology in 1990. She joined China Construction Bank in 1990 and was promoted as a division head in 1992, a senior software engineer in 1996 and associate director in 1998. She won numerous awards for her projects and leadership. After she immigrated to Canada, she started her master graduate study in the department of Computer Science and Electrical Engineering (CSEE) at the University of New Brunswick (UNB) in 2002. She served as a teaching assistant and a substitute lecturer for several courses during her master program. In 2003 she joined the institute of Biomedical Engineering, which was world-renowned for developing the systems that control artificial limbs. She started to expose herself in the research field of biomedical engineering and conducted her thesis, entitled as a Classification Scheme for Myoelectric Control of Upper Limb Prostheses. She was granted to PhD program in March 2004. She got straight A on her master course study. She was awarded the top Canadian national scholarship, 3-year Postgraduate Scholarships, by Natural Sciences and Engineering Research Council of Canada for three years (2005 to 2008) and the Governors Tuition Award in 2005. She earned her master degree in Electrical Engineering from UNB in September 2005. She had one journal paper published and several conference papers based on her research.

Yonghong transferred her PhD program to the CSEE department at OHSU in October 2005. She has since spent her time studying the event-related potential (ERP) in Electroencephalography (EEG) and the brain computer interface technology using signal processing and machine learning techniques. She transferred from the CSEE department to the BME department in June 2008. During her Ph.D. program she has acquired specialized training in the computational modeling of psychological phenomena on cognitive neuroscience in addition to a broad foundation in the core areas of the CSEE. The interdisciplinary nature of the program has profoundly impacted her thesis work. She has been one of the major participants at OHSU in Neurotechnology for Intelligence Analysts (NIA) project to develop an ERP-based BCI for rapid object search. This work was featured in "A Brainy approach to Image Sorting," IEEE Spectrum Online, 2008 (http://www.spectrum.ieee.org/biomedical/imaging/a-brainy-approach-to-image-sorting). The project has been won phase 1 to phase 3 from DARPA. She also participated in the Behavioral Assessment & Intervention Consortium project. During her PhD tenure, Yonghong has been awarded the Student Travel Grant from IEEE ICASSP 2008, the Excellence in Neural Engineering Travel Award from IEEE EMBS 2007 and the Best Topic Paper Award from Augmented Cognition International Society 2006.

Her research interests are broad, spanning from machine learning to computational neuroscience, from signal processing to the psychological concept formation. Continuing with the interdisciplinary nature of her thesis research, she would also like to interact with neuroscientists, experimental psychologists, and other cognitive scientists. Her career goal is to apply technology to clinical medicine and health research.

Yonghong has been served as a reviewer for several journals and conferences. She has presented her research at several national and international conferences and has summarized her work in a handful of peer-reviewed publications. A list of publications and abstracts to date follows:

- Peer Reviewed Journal Papers and Book Chapter

1. Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, and K. E. Hild. A framework for visual image search using single-trial brain responses. In review with the Journal of Neurocomputing.

2. Y. Huang, T. Leen, Z. Lu, D. Erdognus and T. Hayes. Medication adherence predicts cognitive function. In review with IEEE Trans on Info Tech in BioMed.

3. Y. Huang, D. Erdogmus, K. E. Hild , M. Pavel, and S. Mathan. Mixed effects models for single-trial ERP detection in noninvasive brain computer interface design. To be appeared in Recent Advances in Biomedical Signal Processing. J.M. Grriz, Elmar W. Lang, Javier Ramrez (Eds), Bentham Science Publishers, 2010.

4. Y. Huang, K. Englehart, B. Hudgins, and A.D.C Chan. A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. IEEE Trans on Biomed Engi, vol. 52, no. 11, pp. 1801-1811, 2005.

- Peer Reviewed Conference Papers

1. Y. Huang, D. Erdogmus, M. Pavel, K. E. Hild and S. Mathan. A hybrid generative/discriminative method for EEG evoked potential detection. The 4th Intl. IEEE EMBS Conf. on Neural Engineering (CNE), Antalya, Turkey, 2009.

2. Y. Huang, D. Erdogmus, M. Pavel, K. E. Hild and S. Mathan. Target detection using incremental learning on single-trial evoked response. IEEE Intl. Conf. on ICASSP, Taipei, Taiwan, pp. 481-484, 2009.

3. Tian Lan, Y. Huang and D. Erdogmus. A comparison of temporal windowing schemes for single-trial ERP detection. The 4th Intl. IEEE EMBS CNE, 2009.

4. Y. Huang, D. Erdogmus, M. Pavel, S. Mathan. Mixed effects models for EEG evoked response detection. IEEE Intl. Workshop on MLSP, pp. 91-96, 2008.

5. Y. Huang, D. Erdogmus, S. Mathan, M. Pavel. Detecting EEG evoked responses for target image search with mixed effect models. The 30th Intl. IEEE EMBC, pp. 4988-4991, Vancouver, Canada, 2008.

6. Y. Huang, D. Erdogmus, Z. Lu, T.K. Leen. Detecting mild cognitive loss with continuous monitoring of medication adherence. IEEE Intl. Conf. on ICASSP, Las Vegas, NV, pp. 609-612, 2008.

7. Y. Huang, D. Erdogmus, S. Mathan, M. Pavel. Large-scale image database triage via EEG evoked responses. IEEE Intl. Conf. ICASSP, pp. 429-432, 2008.

8. Z. Lu, T.K. Leen, Y. Huang, D. Erdogmus. A reproducing kernel Hilbert space framework for pairwise time series distances. The 25th Intl. Conf. on Machine Learning (ICML), Helsinki, Finland, pp. 624-631, 2008.

9. S. Mathan, D. Erdogmus, Y. Huang, M. Pavel, P. Ververs, J. Carciofini, M. Dorneich, S. Whitlow. Rapid image analysis using neural signals. The 26th Conf. on Human Factors in Computing System, pp. 3309-3314, Italy, 2008.

10. Y. Huang, D. Erdogmus, S. Mathan, M. Pavel. A fusion approach for image triage using single-trial ERP detection. The 3rd IEEE EMBS CNE, pp. 473-476, 2007.

11. Y. Huang, D. Erdogmus, S. Mathan, M. Pavel. Comparison of linear and nonlinear approaches in single-trial ERP detection in rapid serial visual presentation tasks. IEEE World Congress on Computational Intelligence, pp1136-1142, 2006.

12. Y. Huang, D. Erdogmus, S. Mathan, M. Pavel. Boosting linear logistic regression for single-trial ERP detection in rapid serial visual presentation tasks. The 28th Intl. IEEE EMBC, pp3369-3372, New York, NY, 2006.

13. T. Lan, Y. Huang, D. Erdogmus. A comparison of linear ICA and local linear ICA for mutual information based feature ranking. The 6th Intl. Conf. Independent Component Analysis and Blind Signal Separation, pp. 823-830, 2006.

14. T. Lan, D. Erdogmus, U. Ozertem and Y. Huang. Estimating mutual information using Gaussian mixture model for feature ranking and selection. The IEEE World Congress on Computational Intelligence, Vancouver, Canada, pp5034-5039, 2006.

15. S. Mathan, P. Ververs, M. Dorneich, S. Whitlow, J. Carciofini, D. Erdogmus, M. Pavel, C. Huang, T. Lan, A. Adami. Neurotechnology for image analysis: searching for needles in haystacks efficiently. Augmented Cognition Intl. Conf., 2006.

16. Y. Huang, K. Englehart, B. Hudgins, and A.D.C Chan. Optimized Gaussian mixture models for upper limb motion classification. The 26th Intl. IEEE EMBC, San Francisco, CA, pp72-75, 2004.

17. Y. Huang, K. Englehart, B. Hudgins, and A.D.C Chan. Robust upper limb motion classification using Gaussian mixture models. The 28th Conf. of the Canadian Medical and Biological Engineering Society), Canada, pp. 149-152, 2004.

- Abstracts

1. M. Pavel, Y. Huang, K. E. Hild, S. Mathan, and D. Erdogmus. The dynamics of visual detection processes in RSVP paradigms. The Society of Neuroscience (SFN), Chicago, IL, 2009.

2. Y. Huang, K. Englehart, B. Hudgins, and A.D.C Chan. A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. XVth Congress of Intl Electrophysiology & Kinesiology Society, pp67, 2004.

3. Y. Huang, K. Englehart, B. Hudgins, and A.D.C Chan. Classification of myoelectric signals using Gaussian mixture models. The Mathematics of Information Technology and Complex Systems 5th Annual Conf, Canada, p31, 2004.