

# FUSION METHODS FOR DETECTING NEURAL AND PUPIL RESPONSES TO TASK-RELEVANT VISUAL STIMULI USING COMPUTER PATTERN ANALYSIS

by

Ming Qian

Department of Electrical and Computer Engineering  
Duke University

Date: \_\_\_\_\_

Approved: \_\_\_\_\_

\_\_\_\_\_  
Dr. Loren Nolte, Supervisor

\_\_\_\_\_  
Dr. Mario Aguilar

\_\_\_\_\_  
Dr. Roger Barr

\_\_\_\_\_  
Dr. Leslie Collins

\_\_\_\_\_  
Dr. Lisa Huettel

\_\_\_\_\_  
Dr. Kishor Trivedi

Dissertation submitted in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy  
in the Department of Electrical and Computer Engineering  
in the Graduate School of  
Duke University

2008

# ABSTRACT

## FUSION METHODS FOR DETECTING NEURAL AND PUPIL RESPONSES TO TASK-RELEVANT VISUAL STIMULI USING COMPUTER PATTERN ANALYSIS

by

Ming Qian

Department of Electrical and Computer Engineering  
Duke University

Date: \_\_\_\_\_

Approved: \_\_\_\_\_

\_\_\_\_\_  
Dr. Loren Nolte, Supervisor

\_\_\_\_\_  
Dr. Mario Aguilar

\_\_\_\_\_  
Dr. Roger Barr

\_\_\_\_\_  
Dr. Leslie Collins

\_\_\_\_\_  
Dr. Lisa Huettel

\_\_\_\_\_  
Dr. Kishor Trivedi

An abstract of a dissertation submitted in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy  
in the Department of Electrical and Computer Engineering  
in the Graduate School of  
Duke University

2008

Copyright © 2008 by Ming Qian  
All rights reserved

# Abstract

A series of fusion techniques are developed and applied to EEG and pupillary recording analysis in a rapid serial visual presentation (RSVP) based image triage task, in order to improve the accuracy of capturing single-trial neural/pupillary signatures (patterns) associated with visual target detection.

The brain response to visual stimuli is not a localized pulse, instead it reflects time-evolving neurophysiological activities distributed selectively in the brain. To capture the evolving spatio-temporal pattern, we divide an extended (“global”) EEG data epoch, time-locked to each image stimulus onset, into multiple non-overlapping smaller (“local”) temporal windows. While classifiers can be applied on EEG data located in multiple local temporal windows, outputs from local classifiers can be fused to enhance the overall detection performance.

According to the concept of induced/evoked brain rhythms, the EEG response can be decomposed into different oscillatory components and the frequency characteristics for these oscillatory components can be evaluated separately from the temporal characteristics. While the temporal-based analysis achieves fairly accurate detection performance, the frequency-based analysis can improve the overall detection accuracy and robustness further if frequency-based and temporal-based results are fused at the decision level.

Pupillary response provides another modality for a single-trial image triage task. We developed a pupillary response feature construction and selection procedure to extract/select the useful features that help to achieve the best classification performance. The classification results based on both modalities (pupillary and EEG) are further fused at the decision level. Here, the goal is to support increased classification confidence through inherent modality complementarities. The fusion results

show significant improvement over classification results using any single modality.

For crucial image triage tasks, multiple image analysts could be asked to evaluate the same set of images to improve the probability of detection and reduce the probability of false positive. We observe significant performance gain by fusing the decisions drawn by multiple analysts.

To develop a practical real-time EEG-based application system, sometimes we have to work with an EEG system that has a limited number of electrodes. We present methods of ranking the channels, identifying a reduced set of EEG channels that can deliver robust classification performance.

# Contents

<b>Abstract</b>	<b>iv</b>
<b>List of Tables</b>	<b>xi</b>
<b>List of Figures</b>	<b>xii</b>
<b>Acknowledgements</b>	<b>xviii</b>
<b>1 Introduction and Background</b>	<b>1</b>
1.1 Introduction and Background . . . . .	1
1.1.1 Physiological computing and human computer interface . . . .	1
1.1.2 Cognitive categorization process and its EEG based and pupil- lography based responses . . . . .	1
1.1.3 EEG-based visual image classification system . . . . .	2
1.1.4 Pupillary-based visual image classification system . . . . .	8
1.1.5 Fusion techniques . . . . .	8
1.1.6 Single temporal window based classification scheme versus multi- level classification scheme for EEG-based classification system	9
1.1.7 Enhancement of EEG-based single trial visual recognition sig- nature detection using temporal and frequency measures at var- ious frequencies . . . . .	11
1.1.8 Decision level fusion of EEG-based and pupillary response fea- ture based single trial visual recognition detection . . . . .	12
1.1.9 Multiple-subject decision fusion . . . . .	13
1.2 Outline of the Dissertation . . . . .	13
<b>2 Experiment Setup, Signal Acquisition and Signal Preprocessing</b>	<b>14</b>
2.1 Experiment Description . . . . .	14

2.1.1	Participants . . . . .	14
2.1.2	Behavioral paradigm . . . . .	14
2.1.3	Stimuli . . . . .	15
2.1.4	Visual detection task and brain process categories . . . . .	15
2.2	Signal Acquisition . . . . .	16
2.2.1	EEG data acquisition . . . . .	16
2.2.2	Pupil data acquisition . . . . .	17
2.3	Signal Preprocessing . . . . .	17
2.3.1	EEG signal preprocessing . . . . .	17
2.3.2	Pupil dilation signal preprocessing . . . . .	21
2.3.3	Summary . . . . .	21
<b>3</b>	<b>The Role of Multi-Level Fusion in EEG-Based Classification for Image Triage Tasks</b>	<b>24</b>
3.1	Motivation . . . . .	24
3.2	Methods . . . . .	27
3.2.1	EEG signal classification using a single temporal window . . .	27
3.2.2	Feature-level fusion of multiple local classifiers' outputs . . . .	28
3.2.3	Decision-level fusion of multiple local classifiers' outputs . . . .	30
3.3	Results and discussion . . . . .	33
3.3.1	Training and test runs . . . . .	33
3.3.2	Local linear discriminant analysis (LDA) classifier performance	35
3.3.3	Feature-level fusion: feature-level linear discriminant classifier (LDA) performance . . . . .	36
3.3.4	Feature-Level fusion: feature-level RVM classification performance . . . . .	37

3.3.5	Global decision-level fusion classification performance . . . . .	40
3.3.6	Appropriate Number of Local Classifiers . . . . .	40
3.3.7	Comparison of global classifiers . . . . .	41
3.4	Summary . . . . .	42
<b>4</b>	<b>Enhancement of EEG-Based Single Trial Visual Recognition Signature Detection Using Temporal and Frequency Measures at Various Frequencies</b>	<b>44</b>
4.1	Motivation . . . . .	44
4.2	Methods . . . . .	46
4.2.1	Selective distributed oscillatory systems in the brain . . . . .	46
4.2.2	Time-frequency power map of EEG time series . . . . .	47
4.2.3	EEG temporal domain based classification . . . . .	51
4.2.4	EEG spectral band power based classification . . . . .	53
4.2.5	Decision-level fusion . . . . .	55
4.3	Results and Discussion . . . . .	58
4.3.1	Training and Test Runs . . . . .	58
4.3.2	Temporal-domain and frequency-domain classification results .	61
4.3.3	Decision fusion of temporal domain classification and frequency domain based classification results . . . . .	64
4.4	Summary . . . . .	64
<b>5</b>	<b>Decision Level Fusion of EEG-Based and Pupillary Response Based Single Trial Visual Recognition Detection</b>	<b>69</b>
5.1	Motivation . . . . .	69
5.2	Methods . . . . .	70
5.2.1	Pupil dilation feature construction . . . . .	70



5.2.2	Pupil feature selection . . . . .	73
5.2.3	EEG-based classification . . . . .	74
5.2.4	Linear discriminant analysis for pupillary response features . .	77
5.2.5	Decision fusion . . . . .	77
5.3	Results and discussion . . . . .	80
5.3.1	Training and test runs . . . . .	80
5.3.2	Fusion results . . . . .	81
5.4	Summary . . . . .	81
<b>6</b>	<b>Multi-Subject Decision Fusion</b>	<b>85</b>
6.1	Motivation . . . . .	85
6.2	Methods . . . . .	85
6.3	Results and discussion . . . . .	86
6.3.1	Training and test runs . . . . .	86
6.3.2	Decision-level fusion classification performance . . . . .	87
6.4	Summary . . . . .	88
<b>7</b>	<b>Classification Using Reduced Number of EEG Sensors</b>	<b>92</b>
7.1	Motivation . . . . .	92
7.2	Methods . . . . .	93
7.2.1	Finding the inherent dimensionality of the EEG-based detec- tion using principal component analysis . . . . .	93
7.2.2	Ranking the EEG channels using a single temporal window based LDA classifier . . . . .	93
7.2.3	Feature-level fusion of multiple local classifiers' outputs . . . .	94
7.3	Results and discussion . . . . .	95

7.3.1	Training and test runs . . . . .	95
7.3.2	Principal component analysis results . . . . .	95
7.3.3	Classification performance using reduced number of channels .	96
7.4	Summary . . . . .	97
<b>8</b>	<b>Conclusions</b>	<b>102</b>
8.1	Conclusions and Future Works . . . . .	102
8.1.1	Conclusions . . . . .	102
8.1.2	Future works . . . . .	103
	<b>Bibliography</b>	<b>109</b>
	<b>Biography</b>	<b>114</b>

# List of Tables

3.1	Number of target and non-target/distractor images in the training set.	35
3.2	Number of target and non-target/distractor trials in three test sets. .	35
4.1	Number of target and non-target/distractor images in the training set.	59
4.2	Number of target and non-target/distractor images in the test sets. .	61
4.3	Signals used for different types of classifiers. . . . .	61
5.1	Pupillary features used for classification. . . . .	82
5.2	Number of target and non-target/distractor images in the training set.	83
5.3	Number of target and non-target/distractor images, target versus non-target/distractor ratios in test sets for five subjects. . . . .	83
6.1	Number of target and non-target/distractor images in the training set.	86
6.2	Number of target and non-target/distractor images in two test sets. .	87
6.3	fusion gains for test sets by combining decisions from different subjects.	87
7.1	Number of target and non-target/distractor images in the training set.	95
7.2	Number of target and non-target/distractor images in multiple test sets.	95
A-1	Pupillary features used for classification. . . . .	106
A-2	Pupillary features used for classification. . . . .	107
A-3	Pupillary features used for classification. . . . .	108

# List of Figures

1.1	Wide-area image used in the experiment: little red blocks mark the locations of the target chips. The analysts have to zoom in to identify the targets (Imagery Credit DigitalGlobe). . . . .	5
1.2	A wide-area image can be divided into multiple 500x500-pixel image chips (Imagery Credit DigitalGlobe). . . . .	5
1.3	RSVP sequence at 100 msec/image and examples of a red framed target image and green framed non-target/distractor images (Imagery Credit DigitalGlobe). . . . .	6
1.4	Trial-average ERP and critical temporal window selection illustration; the temporal window selected is where the average target trial ERP diverges most sharply from the average non-target/distractor trial ERP. . . . .	7
1.5	Multiple local classification time windows can be extracted and the critical classification window can be identified using a training set. . . . .	10
2.1	Locations of the 64 electrodes based on the international 10-20 EEG electrode placement standard. . . . .	16
2.2	EEG signal and pupil signal recorded simultaneously while the subject views the image sequence. . . . .	17
2.3	Both the EEG and pupil data are fed into the signal processing and classification algorithms for real-time processing. . . . .	18
2.4	EEG preprocessing . . . . .	18
2.5	EEG signal before and after bandpass filtering; The DC baseline drifts are removed. . . . .	19
2.6	EEG signal on frontal channel Fp1 before and after eye blink removal; The eye blink sections of the signal are suppressed. . . . .	20
2.7	Epoch extraction: each epoch is 1-second long and time locked to stimulus onset. . . . .	22

2.8	Pupil signal preprocessing procedure. . . . .	22
2.9	Pupil area data (unit: square millimeters) before and after eye blink segment interpolation. . . . .	23
3.1	A plausible route of multiple processing stages between the retina and the muscles of the hand during a categorization task. Information from the retina is relayed by the lateral geniculate nucleus of the thalamus (LGN) before reaching the primary visual cortex (V1). Processing con- tinues in areas V2 and V4 of the ventral visual pathway before reaching the visual areas in the posterior and anterior inferior temporal cortex (PIT and AIT), which contain neurons that respond specifically to certain objects. The inferior temporal cortex projects to a variety of areas, including the prefrontal cortex (PFC), which contains the visu- ally responsive neurons that categorize objects. To reach the muscles in the hand, signals probably need to pass via the premotor cortex (PMC) and primary motor cortex (MC) before reaching the motor neurons of the spinal cord (reproduced from [41] with the permission of publisher. . . . .	25
3.2	Two-Level fusion scheme: the multiple local classifier outputs are fused either at the feature-level using a linear discriminant analysis classifier (LDA), a relevance vector machine (RVM) classifier or at the decision- level using a Bayesian decision fusion classifier. . . . .	26
3.3	Fusion of two operating points from two separate ROC curves, five fused operating points can be generated depending on where the thresh- old is set against four fused likelihood ratios. The corresponding de- cision rules are: (1) never declare “detection” regardless the decisions made by both classifiers; (2) declare “detection” if at least one classi- fier declares “detection”; (3) declare “detection” if the better classifier (in terms of joint likelihood ratio) declares “detection”; (4) declare “detection” only if both classifiers declare “detection” (consensus); (5) always declare “detection” regardless the decisions made by both clas- sifiers. . . . .	33

3.4	Illustration of fusing two local ROC curves: (a) two local ROC curves with multiple operating points; (b) one operating point on ROC curve 1 fuses with one operating point on ROC curve two generates five fused operating points; (c) a large number of fused operating points can be generated by exhausting all operating point combinations based on ROC curve one and ROC curve two; (d) the convex hull is used to represent the ideal fused operating points. . . . .	34
3.5	Local LDA classifier performance is evaluated using average area under ROC curve as a function of time for three subjects across three test sets. The variance error bars across test sets are also plotted. . . . .	36
3.6	Subject 1: average area under ROC curve as a function of time for different fusion schemes. The variance error bars across test sets are also plotted. . . . .	38
3.7	Subject 2: average area under ROC curve as a function of time for different fusion schemes. The variance error bars across test sets are also plotted. . . . .	38
3.8	Subject 3: average area under ROC curve as a function of time for different fusion schemes. The variance error bars across test sets are also plotted. . . . .	39
4.1	Topographic plots of normalized after-stimulus voltage (temporal domain) and frequency spectral power (frequency domain) amplitudes located at different frequency bands for one of the subjects; each topographic plot is based on 50-millisecond temporal window; the plot shows the evolving patterns from the stimulus onset to 500 milliseconds after onset. . . . .	50
4.2	Top subplot shows the average target and non-target/distractor responses on a single channel; the bottom subplot shows the difference between average target and non-target/distractor responses; The whole epoch can be divided into multiple non-overlapping 50-millisecond sub-windows. . . . .	52
4.3	Feature-level fusion of multiple classifiers' outputs. . . . .	53

4.4	Decision-level fusion to combine the temporal-domain classification results and the frequency-domain classification results; the first-level LDA classifiers apply to different temporal windows; the second-level LDA classifiers are feature-level classifiers; the decision-level fusion classifier fuses the decisions made by the feature-level classifiers. . . .	58
4.5	Fusion of two operating points from two separate ROC curves, five fused operating points can be generated depending on where the threshold is set against the four fused likelihood ratios. The corresponding decision rules are: (1) never declare “detection” regardless the decisions made by both classifiers; (2) declare “detection” if at least one classifier declares “detection”; (3) declare “detection” if the better classifier (in terms of joint likelihood ratio) declares “detection”; (4) declare “detection” only if both classifiers declare “detection” (consensus); (5) always declare “detection” regardless the decisions made by both classifiers. . . . .	59
4.6	Illustration of fusing two ROC curves: (a) there are multiple operating points on each ROC curve; (b) five fused operating points can be generated by fusing any two operating points from each ROC curve; (c) a large number of fused operating points can be generated by exhausting all combinations of individual ROC curves’ operating points; (d) the convex hull of all the fused operating points can be used to represent the optimum improved ROC curve for decision fused results. . . . .	60
4.7	ROC curves derived based on temporal domain voltage amplitude and the frequency spectral powers located at different frequency bands (average performance over two test sets for all four subjects). . . . .	62
4.8	Average performance of all subjects using temporal domain based method and frequency spectral band power based method (the average Azs are represented by X marks). . . . .	63
4.9	ROC curves for fused decisions (average over two test sets for all four subjects). . . . .	65
4.10	Average performance of all subjects by fusing temporal domain based method and frequency domain based method (the average Azs are represented by X marks). . . . .	66

4.11	ROC curves for LDA classifier result based on alpha band , theta band, fused alpha/theta result using LDA classifier and fused alpha/theta result using decisions fusion (average over two test sets for all four subjects). . . . .	67
5.1	Comparison of average pupil area change time course (after removing the baseline) after a target stimulus image display (left) and after a non-target/distractor stimulus image display (right) for an experiment subject. The pupil dilates more drastically after a target stimulus image is displayed. . . . .	70
5.2	Nervous system organization: the brain activities measured by the EEG signals are part of the central nervous system activity while the pupil size is controlled by the peripheral nervous system activity. . . .	71
5.3	Comparison of the average pupil area/velocity/acceleration time course after non-target/distractor image display and after target image display for all subjects. The velocity feature shows earlier differentiations between target and non-target/distractor responses compared to the area feature. The acceleration feature shows even earlier differentiations between target and non-target/distractor responses compared to the area and velocity feature. We achieve more thorough characterization of the pupil responses by deriving velocity and acceleration features. . . . .	72
5.4	Time windows defined for pupil dilation feature extraction: The time period of $-0.3 \text{ seconds} < t < 0 \text{ seconds}$ is defined as the before-stimulus period $T_{before}$ . The time period of $0 \text{ seconds} < t < 1.25 \text{ seconds}$ is defined as the after-stimulus period $T_{after}$ . A smaller time interval between 0.2 seconds and 0.8 seconds after stimulus onset ( $0.2 \text{ seconds} < t < 0.8 \text{ seconds}$ ) is regarded as the most likely time period during which the pupillary response is most salient (we call this time period “critical period” $T_{critical}$ ) in terms of the peak velocity and peak acceleration. . . . .	73
5.5	Fusion scheme: two independent classification results are derived using EEG and pupil signals; thereafter they are fused at the decision-level.	76
5.6	Five subjects’ EEG-based, pupillary response-based, and fused results on test set one and two; Blue curves are the ROC curves for the EEG-based results; Black curves are the ROC curves for the pupil-response based results; Red curves are the ROC curves for the fused results. . .	84



6.1	Decision fusion scheme to combine decisions drawn by multiple subjects.	86
6.2	Decision fusion performance using subject 1 and 2, subject 1 and 3, subject 2 and 3 and using all three subjects for test set 1. . . . .	88
6.3	Decision fusion performance using subject 1 and 2, subject 1 and 3, subject 2 and 3 and using all three subjects for test set 2. . . . .	89
6.4	Decision fusion performance using subject 4 and 5, subject 4 and 6, subject 5 and 6 and using all three subjects for test set 3 . . . . .	90
6.5	Decision fusion performance using subject 4 and 5, subject 4 and 6, subject 5 and 6 and using all three subjects for test set 4 . . . . .	91
7.1	PCA eigenvalues for seven subjects. . . . .	99
7.2	Single window classifier performance for four subjects. . . . .	100
7.3	Comparison of classification performance using different number of channels for the higher target density test set. . . . .	100
7.4	Comparison of classification performance using different number of channels for the lower target density test set. . . . .	101

# Acknowledgements

*At times our own light goes out and is rekindled by a spark from another person. Each of us has cause to think with deep gratitude of those who have lighted the flame within us.*

**Albert Schweitzer**

I would like to express my deepest gratitude to my dissertation advisor Professor Loren Nolte for his advice, help, patience, understanding and support. I thank my committee members, Professors Roger Barr, Lisa Huettel, Lesile Collins, and Kishor Trivedi for agreeing to serve on my committee.

I am also grateful to Teledyne Scientific and Imaging LLC, RTP Research Lab for the opportunity they give me to work at the lab and use the project I worked on as the basis for my dissertation research. I am especially grateful to my supervisor at Teledyne Scientific and Imaging, Dr. Mario Aguilar, for his insight and guidance on the project direction. Thanks are also due to all my co-workers in the lab: Dr. Zhou JiangYing, Dr. Mark Poet, Dr. Thomas Hawkins, Dr. Julio Castrillon, Dr. Thomas Altshuler, Dr. Venkataram Sundareswaran, Karen Zachery, Mark Anderson, Craig Anderson, Jennifer Grater, George Shepherd and Barry Ahrens.

I wish to thank our collaborators at University of California at Berkeley Visual Processing Lab for collecting and sharing experimental data for my dissertation research: Professor Stanley Klein, Dr. Thom Carney, Dr Claudia Privetira, Christian La and Ankur Garg.

I thank Mom and Dad, my wife Chang Chen, my sons Davis Qian and Norman Qian and my sister Wei Wang, for their love and encouragement all these years.

My years at Duke were not all work and no play. I especially want to thank the team members of the volleyball team I have played with for these years: Zhou Yufeng, Yang Kaiyong, Wang Dong, Zhang Jian, Li Hong, Li Hang, Ren Tong, Liu Bin, Li GuoJie, An Lei, Gao Zheng, Liu Quan, Tang Ming, Fu Xing, Chen Miao, Jiang Yong, Matthew Johannes, Kurt Wulff, Shaunna Shen, Samantha Cunningham, Jeffrey Coles, Elizabeth Hampton, Jessie Allison, Sara Cordes and Anne E Hogan.

Finally I want to thank all the people I have met at Duke University for these years: my supervisors and co-workers in different departments (Professor Miguel Nicolelis, Professor Patrick Wolf, Dr. Jose Carmena, Dr Mikhail Lebedev, Dr Roy Crist); my classmates in the Electrical and Computer Engineering Department and Computer Science Department; my lab mates in the Neurobiology Department (Center for Neuro-engineering Research), Electrical Engineering Department (Center for Advanced Computing and Communication), Mechanical Engineering Department (Lab for Intelligent Systems and Control), Biomedical Engineering Department (Center for Physiology Research) and my master thesis advisor Professor Kishor Trivedi. They have made me a better person.

# Chapter 1

## Introduction and Background

### 1.1 Introduction and Background

#### 1.1.1 Physiological computing and human computer interface

Physiological computing provides the direct interface between human physiology and computer technology ([1]). Physiological sensors can be used to acquire bioelectrical signals (e.g. heart beat signals, brain signals) and use them as input sources to interactive computer systems in order to enhance the interactive experience. For example, an electroencephalogram (EEG) signal is used to measure the brain activity and a pupillometry signal is used to measure the pupil size change in this study.

Both EEG-based and pupillometry-based methods are non-invasive methods that can be used to monitor the brain and nervous system activities during performance of a cognitive task. The successful applications of these two methods offer the potential for new brain-computer interfaces (BCI) that can detect a user's cognitive brain states or activity patterns.

This type of BCI requires the real-time assessment of brain/mental activities. The accuracy and robustness of the detection/classification methodologies and algorithms are crucial for such systems.

#### 1.1.2 Cognitive categorization process and its EEG based and pupillography based responses

Categorization is the cognitive process in which objects and ideas are recognized, differentiated and understood. Psychophysiology premises that psychological processes and states are tightly coupled with changes in physiological activity. As a result, physiological signals

can be used to monitor psychological states (e.g. categorization states). In this study, we use EEG-based and pupillography-based responses to detect/classify the cognitive categorization process states.

EEG signals provide a direct command and communication channel from the brain psychophysiology to the computer-controlled output by reading and interpreting the electrical signatures (patterns) of the brain's activity and its responses to external stimuli. EEG recording involves attaching electrodes to the scalp of a person to measure and record electrical activity on the cortex over time ([2]). Human EEG signals represent the aggregate activity of millions of neurons inside the brain. EEG-based system provides a non-invasive communication channel from the brain to the external world by reading and interpreting the electrical signatures of the brain's activity and its cognitive responses to external stimuli (e.g. visual stimuli).

Another widely emphasized cognitive response is the pupillary dilation response to a cognitive event ([3], [4]). The pupil is the expanding and contracting opening in the iris of the eye, through which light passes to the retina. It is well known that human pupils dilate in response to emotion evoking stimuli. Pupillary feature changes (e.g. diameter, diameter change velocity and acceleration) are the human body's responses to stimuli in the environment.

### **1.1.3 EEG-based visual image classification system**

#### **Motivation**

EEG is the neurophysiological measurement of the electrical activity of the brain as measured by electrodes placed on the scalp ([2], [5], [6]). Significant advances have been made in the use of EEG as a method for effective human-computer interfacing, known as Brain-Computer Interfacing (BCI) ([7], [8], [9], [10], [11], [12], [13], [14]). Additionally, more recent efforts have begun to investigate the use of high-resolution EEG signals (i.e., in contrast with the broader signals considered by BCI) to achieve fine control and responses in

a human-computer system ([15], [16], [17] and [18]). Such capabilities hinge on reliably decoding the fine-grain information carried by the spatio-temporal pattern underlying brain activities. In particular, we have been investigating the use of instantaneous EEG signal decoding to control an image analysis application where the user’s brain activity associated with the detection or recognition of targets of interest (e.g. a tumor in an MRI image) is used to manipulate image selection and processing. Our goal is to ultimately design systems that improve the efficacy and throughput of expert image analysts such as radiologists.

In visual image classification applications, human image analysts (IAs) are substantially more effective than computer visual systems at parsing a scene and recognizing target objects. Even though researchers have made significant progress in the development of automated pattern recognition systems which aim to segment image targets from background, there is still a significant gap between the performance of these computer-based systems and the capabilities of human image analysts.

A common engineering approach in implementing automated pattern recognition systems is to design a task-specific image processing algorithm which operates on the sensor data but lacks the capability to exploit task-related or context-based information. On the other hand, IAs perform extremely well on target detection tasks due to their keen sense of task constraints and their reliance on contextual information provided by the characteristics of the scene and its dynamics.

The main challenge is that we usually have a limited number of qualified human IAs and, in many applications, an enormous amount of visual images need to be classified. For a military application example, U.S. Air Force Air Intelligence Agency (AIA) reported that advances in satellite communications bandwidth generate an enormous amount of intelligence imagery and they cannot afford to put imagery analysts on every single frame ([19]). For a medical imagery example, it has been demonstrated that the accuracy of mammographic interpretation is superior when experienced and dedicated analysts interpret mammogram pictures. It was estimated that in 2004 alone, nearly 40 million mammogram

pictures were taken in United States ([20]). The number of available radiologists fall short of such overwhelming demands.

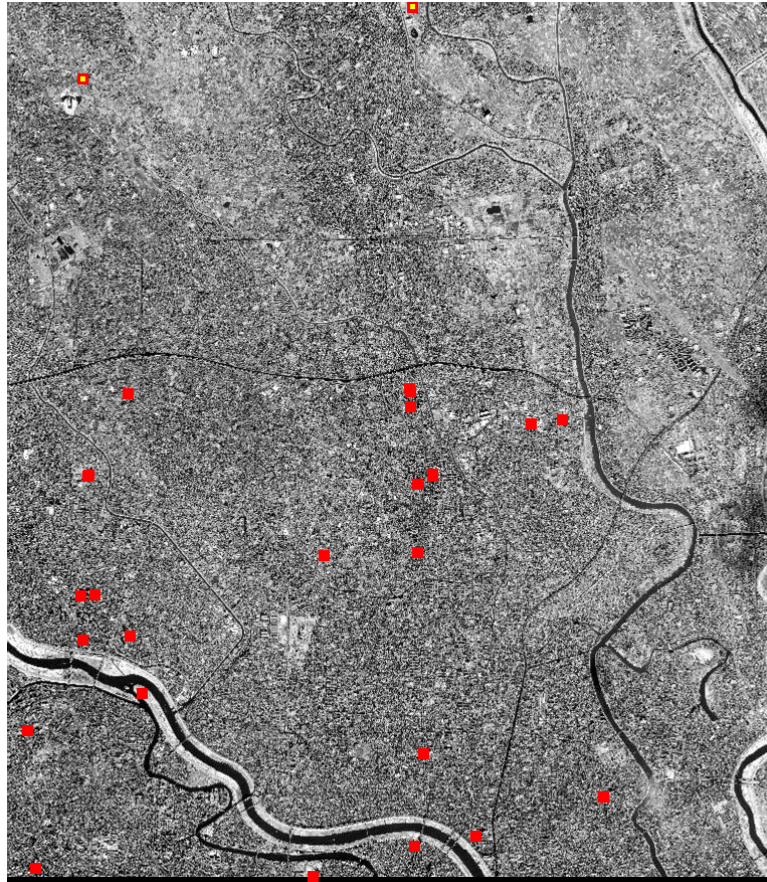
To alleviate these problems, effective triage techniques need to be developed to rapidly screen the high volume of imagery (usually comprising a large amount of non-target/distractor images and a limited number of target images) and sort out a much smaller subset of images that merits further detailed analysis.

EEG-based systems have been developed to optimally couple the superior human visual processing capability with the superior signal processing and classification capability of high-speed computers ([21], [22], [23]). This type of “cortically-coupled computer vision” system ([23]) can be used to improve the image processing throughput/speed in high volume visual triage applications ranging from satellite reconnaissance to medical image analysis. In addition to visual applications, this type of system can also be used to detect and classify auditory or information stimuli in various types of detection systems such as a lie detector or a warning system to detect potential threats.

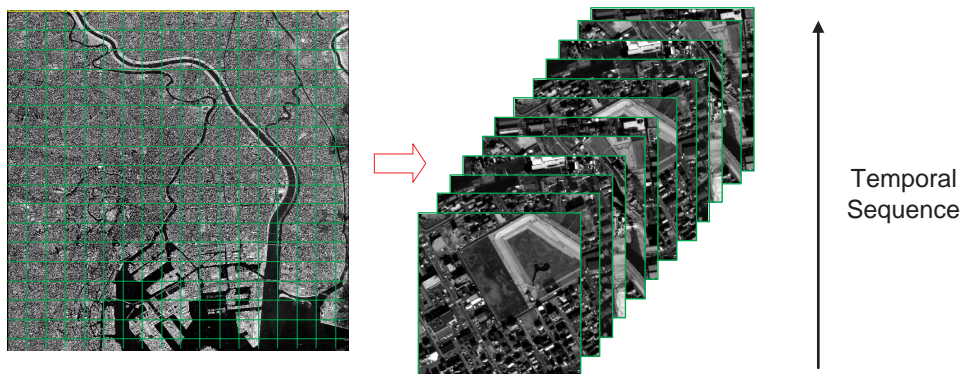
The conventional image triage usually starts with an image that covers a very big area (Fig. 1.1). Experienced analysts can identify multiple areas of interest (AOIs) and accordingly zoom in and amplify the details in those AOIs in an attempt to detect the targets. This process is tedious and inefficient. We advocate the rapid serial visual presentation (RSVP) based BCI system where the images are prepared automatically by computer software and presented rapidly as a continuous sequence of images (Fig. 1.2 and 1.3). RSVP is equivalent to riffle through a book and allows space to be traded off for time ([24]).

## **EEG and event related potential**

Human EEG signals represent the aggregate activity of millions of neurons on the cortex and have high time-resolution (capable of detecting changes in electrical activity in the brain on a millisecond-level resolution). It is well-documented both from neurophysiological evidence and EEG experiments that a variety of relevant cognitive tasks involve the activation of

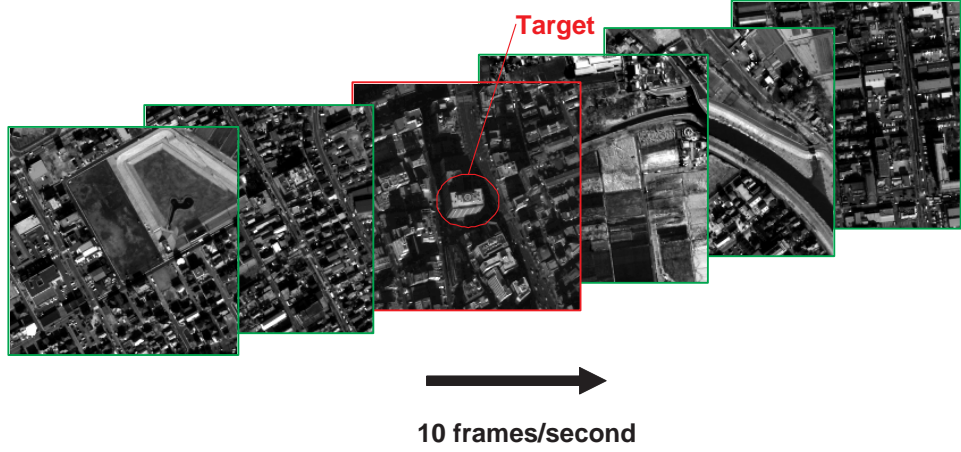


**Figure 1.1:** Wide-area image used in the experiment: little red blocks mark the locations of the target chips. The analysts have to zoom in to identify the targets (Imagery Credit DigitalGlobe).



**Figure 1.2:** A wide-area image can be divided into multiple 500x500-pixel image chips (Imagery Credit DigitalGlobe).

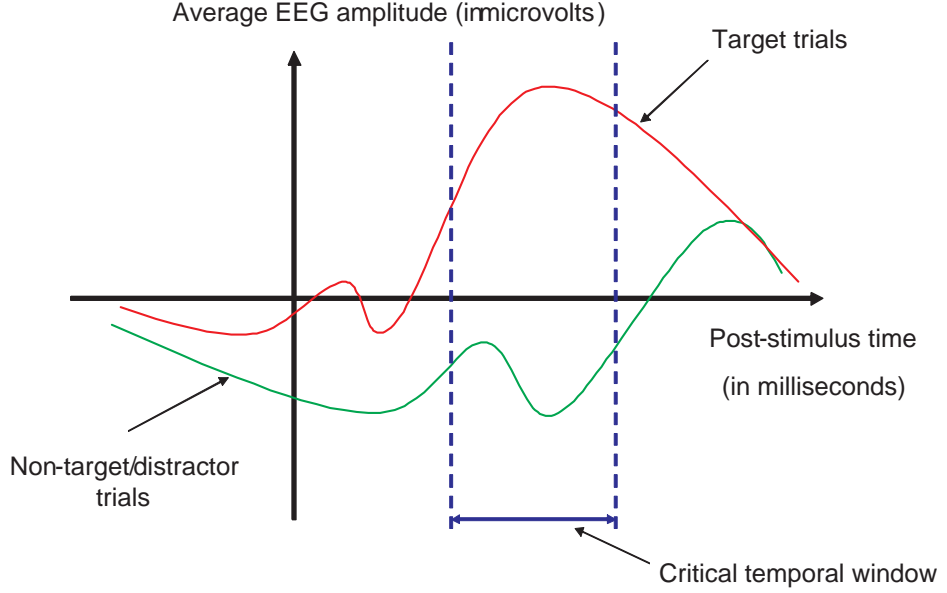




**Figure 1.3:** RSVP sequence at 100 msec/image and examples of a red framed target image and green framed non-target/distractor images (Imagery Credit DigitalGlobe).

multiple areas in the brain. Both the timing and location of these activities reveal the nature of the cognitive task. A variety of studies are underway to achieve accurate source localization and measure temporal coherence throughout the brain ([25], [26], [23], [27] and [21]).

An event-related potential (ERP) is a stereotypical electrophysiological response to an internal or external stimulus ([6]). More simply, it is a measured brain response that is directly the result of a thought or perception. By the nature of recording constraints (e.g., poor sensor interface, signal cross-talk, etc.) and unrelated brain activities, EEG signals of interest have a very low signal-to-noise ratio (SNR) ([2]). For this reason, most research in the use of EEG utilizes signal average over a large number of trials in order to recover the signal and suppress the noise. For instance, evidences by Thorpe and colleagues ([28]) suggested a significant amplitude difference between EEG responses to target and non-target/distractor images can be measured if one averages over a large number of trials (illustrated in Fig. 1.4). In a similar approach, Tan and colleagues ([29]) applied the trial-averaged method in the EEG-based brain state classification task.



**Figure 1.4:** Trial-average ERP and critical temporal window selection illustration; the temporal window selected is where the average target trial ERP diverges most sharply from the average non-target/distractor trial ERP.

### Single-trial detection

In other contexts, the event or task stimulus cannot be trial-averaged. For example, the stimulus may occur only once or the application needs to classify the brain response in real-time. In the context of the application domain we consider here (instantaneous image triage in response to every single image display), we must apply signal processing techniques capable of recovering the critical EEG signal from single-trial inputs. The biggest challenge of single-trial detection is to overcome the low signal-to-noise ratio problem imposed by event-unrelated background EEG responses which usually have larger amplitude than event-related responses and could completely obscure the latter ([30]).

Recent advances in adaptive signal processing have demonstrated significant single-trial detection capability by integrating EEG signals spatially, across multiple channels of high density EEG sensors ([21], [31]). A weighted sum of all electrodes over a predefined (fixed) temporal window can be used as a new composite signal that serves as a discriminating component for the signal of interest (such as a response to a target of interest in an image

by a human observer).

#### **1.1.4 Pupillary-based visual image classification system**

A pupil monitoring system usually records data from one source only (e.g. pupil diameter size). In this study, we developed a pupillary response feature construction and selection procedure to extract the useful features that perform the best under the linear discriminant analysis (LDA) classifier. The single-trial analysis using an LDA classifier is applied using this group of pupillary response features to detect the pupillary signature of target detection.

#### **1.1.5 Fusion techniques**

To further improve the accuracy and robustness of the single-trial detection, we apply a series of fusion techniques in this study. Fusion techniques combine data from different sources together. The main objective is to produce a fused result that is more accurate and reliable.

Fusion can be executed at different levels (data-level, feature-level or decision-level). Data-level fusion is the combination of the raw data from multiple sources into a single representation. Feature-level fusion requires the extraction of different features from the source data and the features are merged afterwards. Decision-level fusion combines the results (decisions) from multiple algorithms to yield a fused decision. Under different circumstances, fusions executed at different levels have their own pros and cons. We have applied fusion techniques at all three levels in our study to improve the single trial classification performance.

Data-level fusion has advantages of extending spatial and temporal coverage and relying on the most detailed information. Feature-level fusion has advantages of reducing the problem dimensionality, simplifying the classifier design and throwing away the irrelevant information. Decision-level fusion has advantages of handling heterogeneous data sources well, reducing the problem dimensionality, simplifying the classifier design, shielding from

the noise better and being interpreted easily.

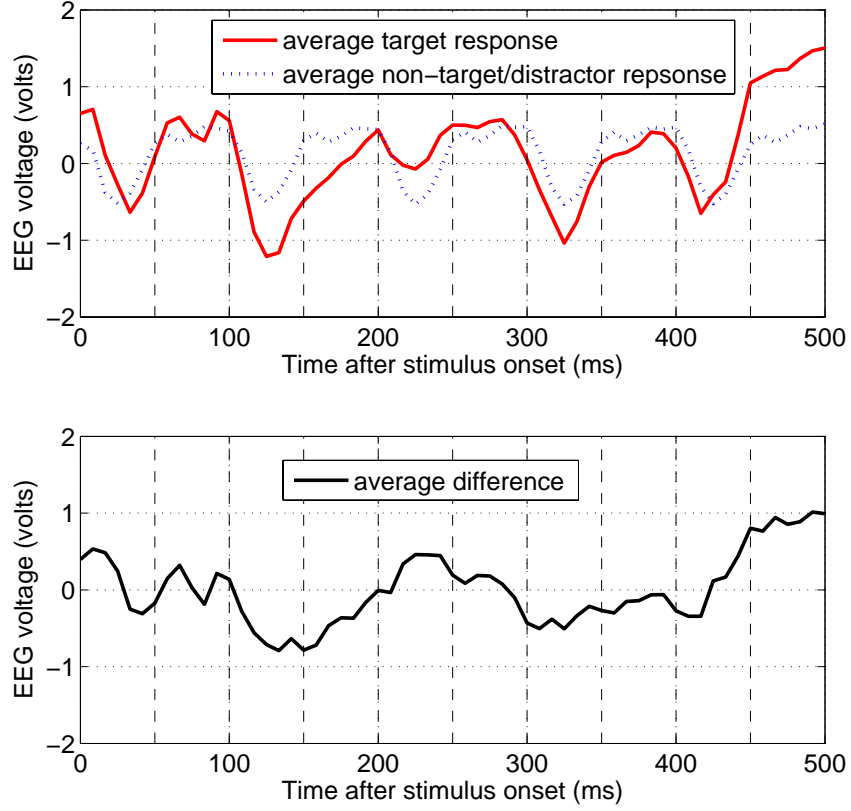
Fusion algorithms can be applied to various types of data including multi-sensor data, multi-temporal data, multi-resolution data, multi-parameter data or multi-modality data. Fusing the EEG data from multiple electrodes is a multi-sensor data fusion example (chapter three). Fusing the classifiers' outputs based on multiple temporal windows is a multi-temporal data fusion example (chapter three). Fusing the results of a temporal-domain based classifier and the result of a frequency-domain based classifier is a multi-parameter fusion example (chapter four). Fusing the result of an EEG-based algorithm and the result of a pupillary response based algorithm is a multi-modality decision fusion example (chapter five).

### **1.1.6 Single temporal window based classification scheme versus multi-level classification scheme for EEG-based classification system**

#### **Single temporal window based classifier**

For both the trial averaged approach and the spatial integrated single-trial approach, it was found that the event-related EEG response is most prominent at a certain critical time period after stimulus onset. For example, Thorpe et al. ([28]) found that the trial averaged ERP generated on target and non-target/distractor trials diverge sharply at 150-200 milliseconds after stimulus onset for a go/no-go image categorization task. Parra et al. ([32]) applied the logistic regression linear discriminant classifier to EEG data in a predefined temporal window centering around the time where the target trial averaged ERP is most sharply diverged from the non-target/distractor trial averaged ERP (illustrated in Fig. 1.4) and achieved significant classification performance.

In practice, a fixed length (e.g. 1-second wide) EEG data epoch time-locked to image stimulus can be extracted for every trial. This long ("global") epoch can be further divided into multiple smaller ("local") time windows (e.g. 50 msec wide) and the critical



**Figure 1.5:** Multiple local classification time windows can be extracted and the critical classification window can be identified using a training set.

classification window can be identified using a training set (Fig. 2.7).

To further improve the detection performance and robustness, we adopt a multi-level classification scheme to fuse the global spatio-temporal pattern across the multiple local windows.

### Multi-level classification scheme

The brain response to visual stimuli is not a localized pulse, instead it reflects time-evolving neurophysiological activities distributed selectively in the brain. To capture the evolving spatio-temporal pattern, we divide a long (“global”) EEG data epoch (e.g. 1 second wide) time-locked to each image stimulus onset into multiple non-overlapping smaller (“local”)

temporal windows (e.g. 50-msec wide). Instead of applying a linear classifier on a single local temporal window, linear classifiers can be applied on multiple local temporal windows to capture their individual signatures. To capture the evolving phenomenology of spatio-temporal changes in overall brain responses to target detection over the extended temporal window, outputs from multiple local classifiers can be fused to improve the overall detection performance. This kind of two-level fusion scheme was first proposed in [25] and [23].

In chapter three, we evaluate the performance of fusing the multiple local classifiers' outputs both at the feature-level and at the decision-level to discriminate between target and non-target/distractor trials (Fig. 5.5). Two feature-level fusion methods (a Linear Discriminant Analysis classifier known as Fisher's linear discriminant and a Relevance Vector Machine classifier known as RVM) and one Bayesian decision-level fusion method are implemented and their performance is evaluated in terms of the area under the Receiver Operating Characteristic (ROC) curves (graphical plots of correct detection rate versus false alarm rate for a binary classifier system as its discrimination threshold is varied). More importantly, we propose an effective approach to evaluate the appropriate number of local temporal windows that should be included in the global fusion. The cognitive process only lasts a limited amount of time after stimulus onset. For global fusion, we only want to include the local temporal windows related to the cognitive process. To include extra temporal windows located beyond the cognitive process completion time adds unnecessary computational load onto the system and as a result, degrades the processing throughput and speed.

### **1.1.7 Enhancement of EEG-based single trial visual recognition signature detection using temporal and frequency measures at various frequencies**

According to the concept of induced/evoked brain rhythms, the EEG response related to the visual target recognition can be decomposed into different oscillatory components and

the frequency characteristic for these oscillatory components can be evaluated separately from the amplitude characteristic. The previous studies had observed prolonged cognitive event related synchronization (power-enhancement) or de-synchronization (power blocking) responses in various frequency bands during experiments with certain cognitive tasks ([33], [34], [35] and [36]) . While the temporal-domain analysis was found to have fairly accurate detection performance by previous studies, the frequency-domain analysis can improve the overall detection accuracy and robustness further if we fuse the frequency-domain result with the temporal-domain result at the decision level.

In chapter four, using multiple human subjects across multiple test sets, we are able to show that the frequency-domain classifiers have different visual target detection signatures compared with the temporal-domain classifier. Significant performance improvement can be achieved by fusing the temporal-domain decision and the frequency-domain decision.

### **1.1.8 Decision level fusion of EEG-based and pupillary response feature based single trial visual recognition detection**

Pupil response is another psychophysiological modality we can use as a basis for single trial target image detection tasks. The pupil is the opening in the iris of the eye that expands and contracts, allowing light to pass to the retina. The expansion and contraction of the pupil are mediated by the sympathetic and parasympathetic nervous systems (both are parts of the peripheral nervous system). Task-evoked pupil dilations have been found to be a function of the cognitive workload and attention required to perform the task ([3], [4]). It also has long been known that the pupil dilates in response to emotion evoking stimuli ([37]).

In chapter five, we developed a pupillary response feature construction and selection procedure to extract the useful features that perform the best under the linear discriminant analysis (LDA) classifier. The single-trial analysis using a LDA classifier is applied

using this group of pupillary response features to detect the target detection signature. At the same time, we apply a two-level LDA classification scheme on cognitive task-related EEG response. The classification results based on the two modalities are further fused at the decision level. The fusion results show significant improvement over classification performance using any of the single modalities (EEG or pupil).

### **1.1.9 Multiple-subject decision fusion**

To further improve the classification performance of the cortically-coupled computer vision system, multiple human subjects can be used so that multiple decisions can be drawn on the same image set by multiple subjects. The fused decisions based on the multiple individual decisions are found to have improved classification performance than each individual decision.

In chapter six, we examine the performance improvements that can be achieved by fusing classification decisions drawn by multiple subjects on the same test sets.

## **1.2 Outline of the Dissertation**

The remainder of the dissertation is organized as follows: Chapter two presents the experiment setup, signal acquisition and preprocessing procedures. Chapter three examines the role of multi-level classification schemes and the appropriate number of local time windows that should be included in the multi-level classification scheme. Chapter four presents the approach of using EEG-based frequency domain classifier to enhance the EEG-based temporal domain analysis. Chapter five presents the single-trial detection method using pupillary response and the decision fusion combining the pupillary response based results and the EEG-based results. Chapter six examines the performance improvements that can be achieved by fusing classification decisions drawn by multiple subjects. Chapter seven concludes the dissertation and presents some ideas for future research.



## Chapter 2

# Experiment Setup, Signal Acquisition and Signal Preprocessing

## 2.1 Experiment Description

### 2.1.1 Participants

To evaluate the true benefit and speedup of the EEG and pupillary response based image triage systems, multiple professional image analysts (both females and males) participated and received no payment. All had normal or corrected to normal vision and reported no history of neurological problems. All participants had at least five years of experience with image analysis.

### 2.1.2 Behavioral paradigm

During the visual target detection task, participants were presented with a continuous sequence of urban scene images. Images may or may not contain the object of interest (a helipad). Each experimental run consisted of tens or hundreds of target images and thousands of non-target/distractor images (distractors are objects that could be targets in other contexts but were not part of the experimental task). For every experimental session, each subject went through several experimental runs (5-6 minutes on average) while taking a short break (1 minute) after each run to allow them to relax their eyes. The subjects were instructed to push a button as soon as they detected a target image.

Images were presented at 10Hz (ten images per second) using a rapid serial visual presentation (RSVP) paradigm where a continuous sequence of images was presented rapidly in a fixed (centered on the screen) location, thereby offering an alternative to the conventional concurrent display of images in the space domain. RSVP can help users to browse

faster than normal because the eyes do not need to move.

### **2.1.3 Stimuli**

Computer-based image processing software was used to divide wide-area urban scene images (each image covers approximately 300 square kilometers) into thousands of 500 x 500 pixel image chips (Fig. 1.2 and 1.3). Target images were selected so that they contained targets (helipads) located at the center of the images (examples are shown in Fig. 1.3 where the green framed are non-target/distractor images and the red framed image is a target image). The target image density ratio per sequence in different experiment sets ranges from 1% to 5%. These ratios reflect the realistic target density of practical image detection task (i.e., helipads in large metropolitan areas). Target versus non-target/distractor image presentation events were recorded and EEG/pupillary data were segmented based on the image presentation events.

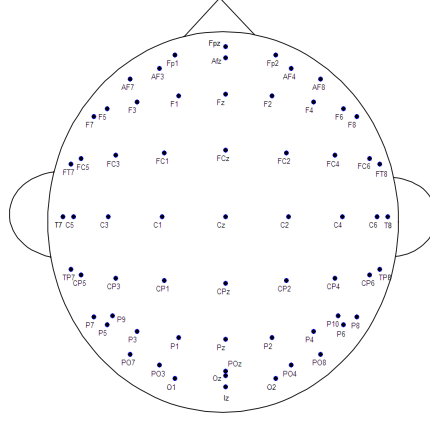
### **2.1.4 Visual detection task and brain process categories**

The chief brain cognitive functions related to the visual target detection are to distinguish targets from their background and to compare them with brain memory content. The perception must be initiated to associate the visual detected object with the existing experience stored in memory or at least find the similarity between the observed object and existing experience related to the target in brain memory. For such reasons, different parts of the brain are involved in the visual detection task.

A subject may engage in multitasking with multiple brain processes at the same time. The possible brain processes (thoughts) can be categorized as ([38]):

Task-relevant thoughts: those thoughts which are directly related to performing the visual detection task including pattern recognition, stimulus identification and response selection.

Task-related thoughts: those thoughts which are related to the task but have no direct



**Figure 2.1:** Locations of the 64 electrodes based on the international 10-20 EEG electrode placement standard.

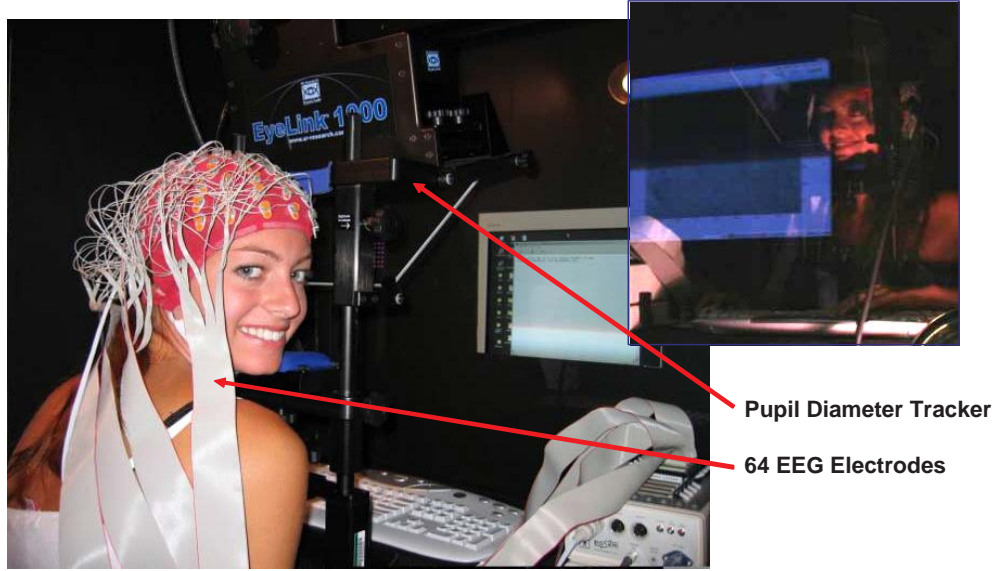
effect on performance. For example, the performance evaluation on himself/herself or other subjects (“How am I doing compared to the other guys?”) or attitudes towards the task (“The presentation speed is too fast.”).

Task-irrelevant thoughts: those thoughts which are unrelated to the task and have no direct effect on performance. For example, a song that keeps running through your mind, trying to decide where to have lunch after the experiment, worry or obsessive thoughts.

## 2.2 Signal Acquisition

### 2.2.1 EEG data acquisition

EEG data were acquired in a sound isolating chamber using a 64-channel Biosemi EEG cap and amplifiers (<http://www.biosemi.com>) for human subjects while they were executing visual detection tasks. The 64 EEG channels were placed according to the international 10-20 EEG placement standard shown in Fig. 2.1. Continuous EEG signals were digitized at 512 Hz and subsequently sub-sample to 120Hz to reduce the computational load.



**Figure 2.2:** EEG signal and pupil signal recorded simultaneously while the subject views the image sequence.

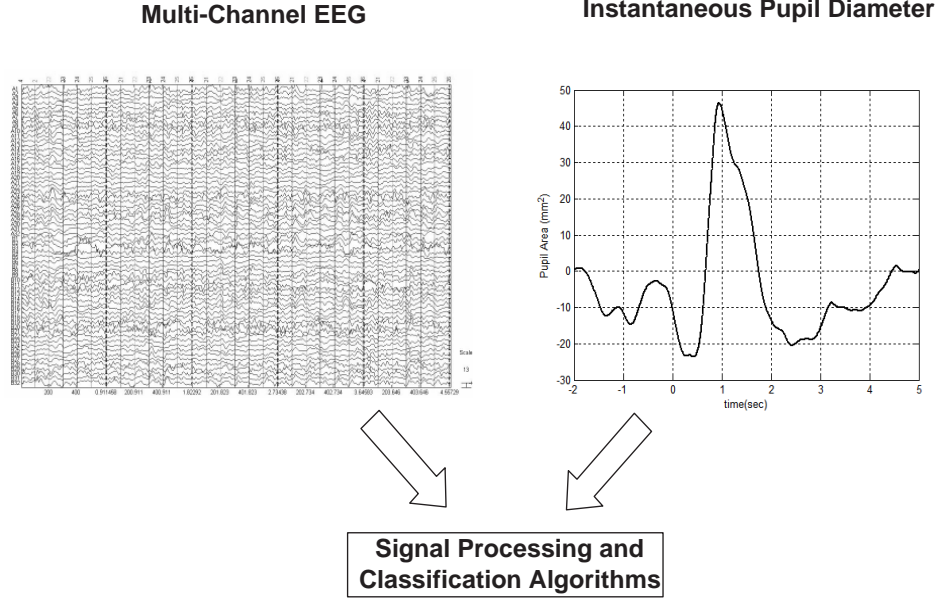
## 2.2.2 Pupil data acquisition

Pupil data were acquired using an EyeLink1000 video based eye-tracking device (<http://www.sr-research.com>) that offers a 1000Hz sampling rate using an infrared camera (Fig. 2.2 and 2.3). The area of the pupil was measured by the number of pixels. These area measures can be converted to pupil diameters in millimeter units (13.5 pixels per square millimeter). The data were further sub-sampled to 120Hz so that it can be synchronized with EEG data.

## 2.3 Signal Preprocessing

### 2.3.1 EEG signal preprocessing

The EEG signals observed on the scalp are in the microvolt range. Added to the actual signals from the brain sources are external artifacts (such as signals generated from eye blinks and facial muscle activities). The artifacts and other noises pose serious threats to the accurate analysis of EEG signals because the EEG signals usually need to be amplified tremendously before they can be used for further analysis.



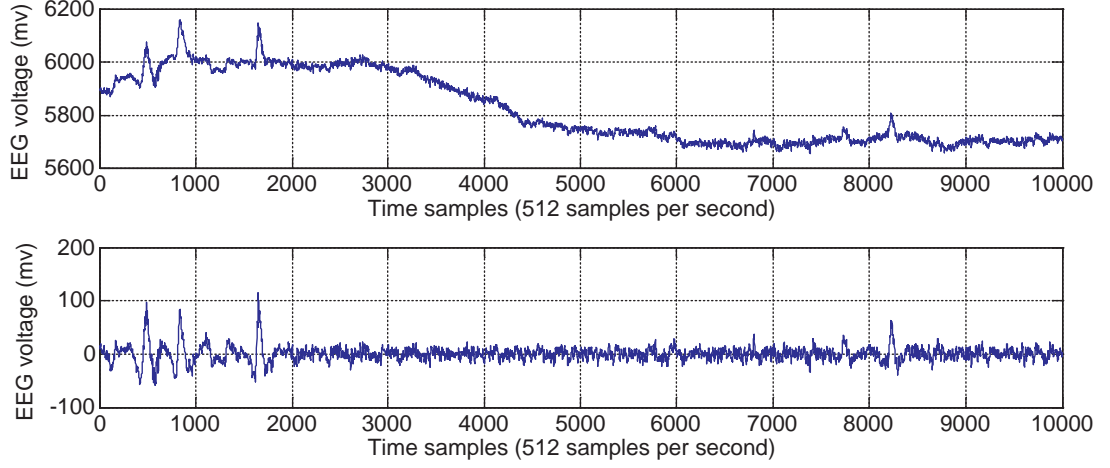
**Figure 2.3:** Both the EEG and pupil data are fed into the signal processing and classification algorithms for real-time processing.



**Figure 2.4:** EEG preprocessing

A signal preprocessing procedure was applied on the recorded continuous EEG data (Fig. 2.4). The DC baseline drifts caused by slow skin conductivity changes and abrupt voltage jumps caused by muscle movements are removed by second-order Butterworth band-pass filter (passband 2Hz-100Hz) (Fig. 2.5).

The muscle activity of eye blinks generates strong electrical signals that can have much higher amplitudes than the signal amplitudes for signals of interest. To preserve the signals of interest, the electrical signals generated by eye blinks have to be removed. The conventional method of eye blink removal is to detect eye blinks and simply discard the corresponding segments of data and interpolate the corrupted segments with neighboring valid data. In practice, this conventional method is not feasible because we would lose too many useful signals of interest embedded in the corrupted segments due to frequent eye blinks and eye motions.



**Figure 2.5:** EEG signal before and after bandpass filtering; The DC baseline drifts are removed.

A better algorithm for subtracting the artifacts using linear regression is proposed in [39] and we adopt the method in our procedure. At the beginning of each experimental session, participants were asked to have an eye blink calibration experiment during which they were instructed to blink their eyes. This enables the determination of linear components associated with eye blinks ([39]).

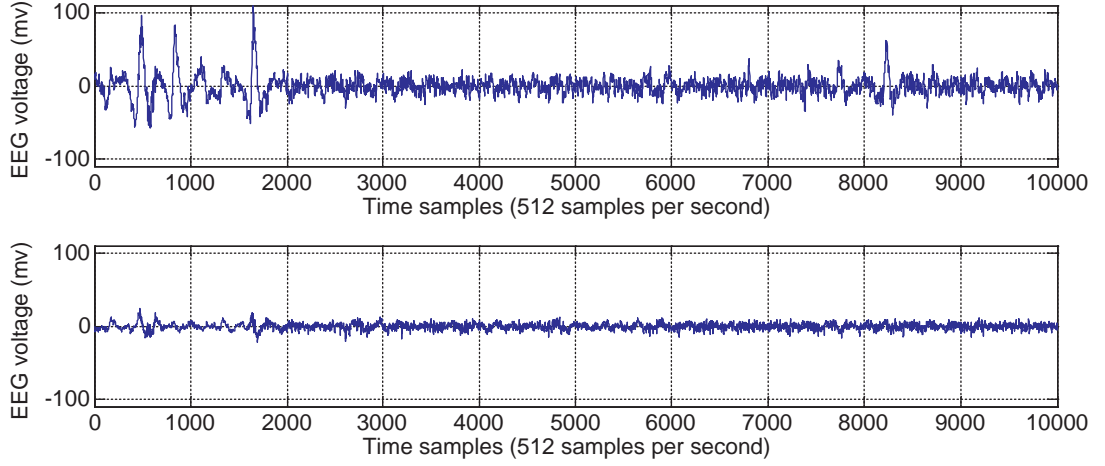
Let  $\mathbf{x}(t)$  be the observed EEG reading,  $y(t)$  be the eye blink signal, and  $\mathbf{s}(t)$  be the remaining signals of interest, then

$$\mathbf{x}(t) = \mathbf{a}y(t) + \mathbf{s}(t) \quad (2.1)$$

where  $\mathbf{a}$  is the linear coupling of the eye blink source  $y(t)$  with the EEG sensors. Given  $\mathbf{a}$ , we can generate a new signal  $\tilde{\mathbf{x}}(t)$  which only contains the signal of interest after the orientation of the eye blink is removed.

$$\tilde{\mathbf{x}}(t) = (\mathbf{I} - \mathbf{a}\mathbf{a}^T)\mathbf{x}(t) = (\mathbf{I} - \mathbf{a}\mathbf{a}^T)\mathbf{s}(t). \quad (2.2)$$

The coupling  $\mathbf{a}$  can be derived by separating EEG channels into two sets: frontal channels containing strong eye blink signals  $\mathbf{x}_1(t)$  and the remaining parietal, temporal and occipital channels containing weak or no eye blink signals  $\mathbf{x}_2(t)$ . The partitioned  $\mathbf{x}(t) = [\mathbf{x}_1(t)$



**Figure 2.6:** EEG signal on frontal channel Fp1 before and after eye blink removal; The eye blink sections of the signal are suppressed.

$\mathbf{x}_2(t)$ ] and correspondingly  $\mathbf{a}(t) = [\mathbf{a}_1(t) \ \mathbf{a}_2(t)]$ .  $\mathbf{a}_1$  can be identified as the first principal component of  $\mathbf{x}_1(t)$  estimated during eye blinks because  $y(t) \gg \mathbf{s}(t)$  in blink periods. Given  $\mathbf{a}_1$ , the estimated eye blink signal can be derived as  $\tilde{y}(t) = \mathbf{a}_1^T \mathbf{x}_1(t)$ .  $\mathbf{a}_2$  can be derived with conventional regression as the linear predictors of  $\mathbf{x}_2$  given  $\tilde{y}(t)$ . The derived signal of interest  $\tilde{\mathbf{x}}(t)$  is used for further processing so that the impact of eye blink artifact can be projected out. A comparison of EEG signal on frontal channel Fp1 (which is located right above the eye) before and after the eye blink removal is shown in Fig. 2.6.

After the eye blink artifact removal, the average voltage amplitude of 64 signal channels was used as the reference signal and this reference signal was subtracted from each of the signal channels in order to remove the unnecessary noise in the “reference free” data that was acquired by active electrodes ([40]). The overall signal-to-noise ratio can be improved by removing an averaged reference signal because 64 EEG channels have various noise levels due to their locations and how well they are placed on the scalp.

The preprocessed EEG data (after bandpass filtering, eye blink removal and reference removal) were segmented into 1-second epochs locked on visual stimulus onset events for each trial (Fig. 2.7). As a result, 1-second epochs of 64-channel data were saved for each

target and non-target/distractor trial. These segmented epochs were used as inputs for single-trial analysis.

### 2.3.2 Pupil dilation signal preprocessing

Signal preprocessing procedures were applied on the recorded continuous pupillary data (Fig. 2.8). The pupil size data can be corrupted by eye blinks. The pupil size monitoring device has an eye blink detection mechanism implemented. First we removed all corrupted pupil size data associated with eye blink regions. Then we interpolated data to fill in missing data segments created by eye blinks. In Fig. 2.9 we illustrate an example of the continuous pupil size data before and after the eye blink data segment interpolation.

Averaging filter was applied on the pupil area data to improve signal to noise ratio. Pupil diameter data (in millimeter unit) were generated using pupil area data (13.5 pixels per square millimeter).

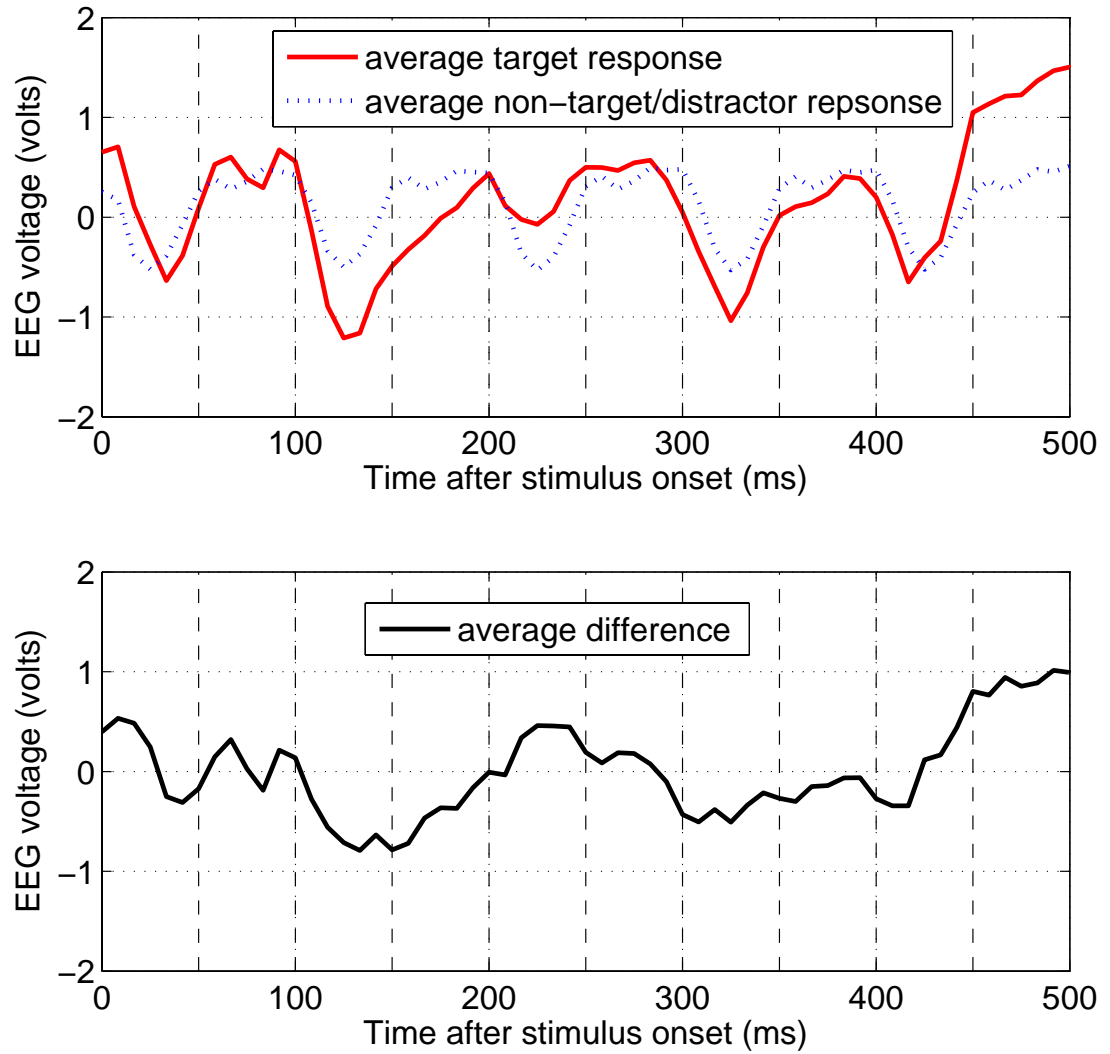
Event markers for target/distractor stimulus onset were set upon the preprocessed pupil diameter data (after eye blink segment interpolation, averaging filtering). As a result, the data were segmented for each target and non-target/distractor event. The segmented data were used for further classification analysis.

### 2.3.3 Summary

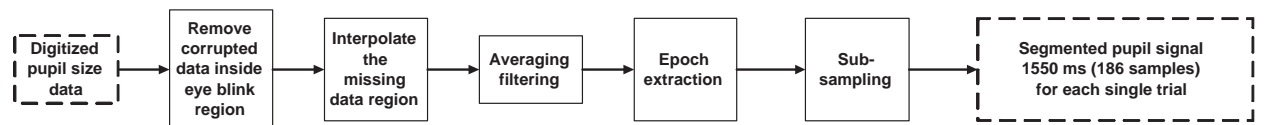
EEG data are high dimensional and noisy due to unpredictable neurophysiological processes in the brain. As a result, a well-organized experiment setup and a proper sequence of signal preprocessing procedures to clean up the noisy data are crucial to the success of classification algorithms that will be applied as the next step.

Pupillary diameter is a single-channel signal and we generate multiple features based on that single-channel signal. A proper sequence of signal preprocessing procedures to clean up the noise is crucial to the success of classification algorithms that will be applied as the next step.

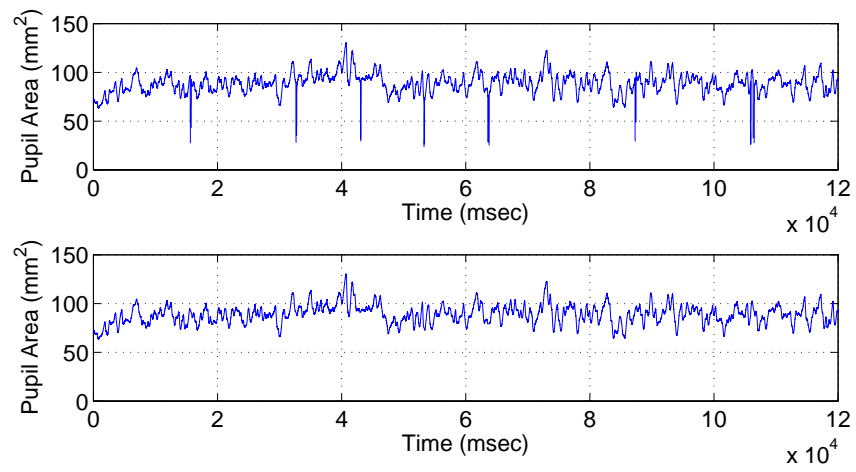




**Figure 2.7:** Epoch extraction: each epoch is 1-second long and time locked to stimulus onset.



**Figure 2.8:** Pupil signal preprocessing procedure.



**Figure 2.9:** Pupil area data (unit: square millimeters) before and after eye blink segment interpolation.

## Chapter 3

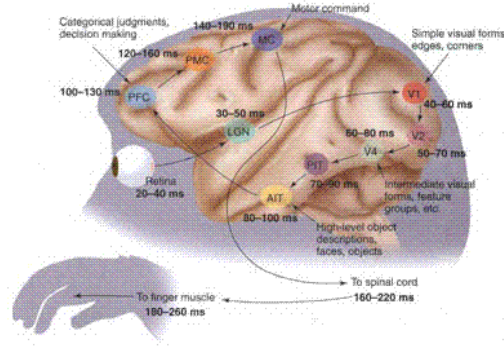
# The Role of Multi-Level Fusion in EEG-Based Classification for Image Triage Tasks

### 3.1 Motivation

Recent advances in adaptive signal processing have demonstrated significant single-trial detection capability by integrating EEG data spatially, across multiple channels of high density EEG sensors ([21], [31]). A weighted sum of all electrodes over a predefined temporal window can be used as a new composite signal that serves as a discriminating component between responses to target versus non-target/distractor stimuli.

For both the trial-averaged approach and the spatially integrated single-trial approach (see section 1.1.3), it was found that the event-related EEG response triggered by target detection is most prominent at certain critical time periods after stimulus onset. For example, Thorpe et al. ([28]) found that the trial-averaged ERP generated on target versus non-target/distractor trials diverged very sharply at 150-200 milliseconds after stimulus onset for a go/no-go image categorization task. Parra et al. ([32], [21], [39], [31]) applied the logistic regression linear classifier on EEG data in a predefined temporal window centering around the time where the target trial averaged ERP is most sharply diverged from the non-target/distractor trial averaged ERP and achieved significant single-trial classification performance.

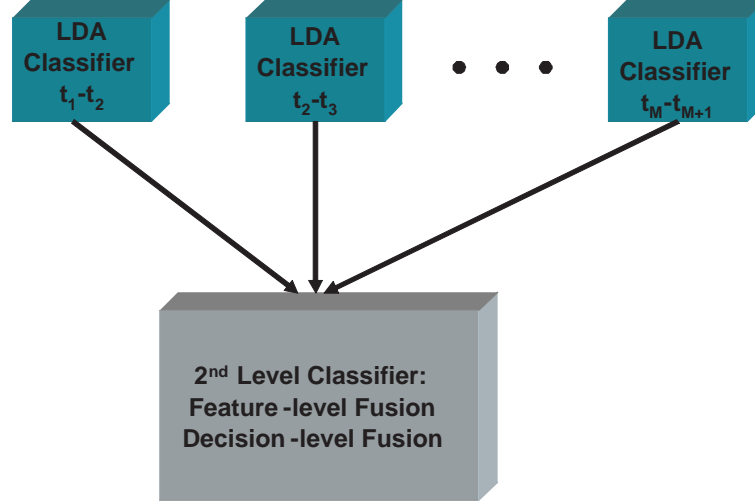
However, the brain response to visual stimuli is not a localized pulse, instead it reflects neurophysiological activities located in selectively distributed sites of the brain evolving



**Figure 3.1:** A plausible route of multiple processing stages between the retina and the muscles of the hand during a categorization task. Information from the retina is relayed by the lateral geniculate nucleus of the thalamus (LGN) before reaching the primary visual cortex (V1). Processing continues in areas V2 and V4 of the ventral visual pathway before reaching the visual areas in the posterior and anterior inferior temporal cortex (PIT and AIT), which contain neurons that respond specifically to certain objects. The inferior temporal cortex projects to a variety of areas, including the prefrontal cortex (PFC), which contains the visually responsive neurons that categorize objects. To reach the muscles in the hand, signals probably need to pass via the premotor cortex (PMC) and primary motor cortex (MC) before reaching the motor neurons of the spinal cord (reproduced from [41] with the permission of publisher).

with a continuous time course (Fig. 3.1). To capture the evolving spatio-temporal pattern, we propose to divide an extended (“global”) EEG data epoch (e.g. 900 milliseconds), time-locked to each image stimulus onset, into multiple non-overlapping smaller (“local”) temporal windows (e.g. 50 msec wide). Instead of applying a linear classifier on a single local temporal window, linear classifiers can be applied on multiple local temporal windows to capture their individual signatures (patterns). To capture the evolving phenomenology of spatio-temporal changes in overall brain responses to target detection over the extended temporal window, outputs from multiple local classifiers can be fused to improve the overall detection performance. This kind of two-level fusion scheme was proposed in [25] and [23].

In this chapter, we evaluate the performance of fusing multiple local classifiers’ outputs, both at the feature-level and at the decision-level, to discriminate between target and non-target/distractor images (Fig. 3.2). Two feature-level fusion methods (a Linear Dis-



**Figure 3.2:** Two-Level fusion scheme: the multiple local classifier outputs are fused either at the feature-level using a linear discriminant analysis classifier (LDA), a relevance vector machine (RVM) classifier or at the decision-level using a Bayesian decision fusion classifier.

criminant Analysis classifier known as Fisher’s linear discriminant and a Relevance Vector Machine classifier known as RVM) and one decision-level fusion method are implemented and their performance is evaluated in terms of the area under ROC curves (graphical plots of correct detection rate versus false alarm rate for a binary classifier system as its discrimination threshold is varied).

More importantly, we develop an effective approach to evaluate the appropriate number of selected local temporal windows that should be included in the global fusion. The brain response phenomenology to visual stimuli lasts a limited amount of time after stimulus on-set. For global fusion, we only need to include a selected number of local temporal windows that maximize our capability to observe/interpret the brain phenomenology. Extra temporal windows add unnecessary computational load onto the system, degrade the processing throughput/speed and yet do not improve (sometimes even degrade) overall classification performance. Due to inter-subject variability in brain response phenomenology, a “one size fits all” solution should not be embraced. Instead, each subject should be evaluated individually to identify the “threshold” local time window so that we can keep the number

of local classifiers incorporated in the two-level fusion scheme at the minimum.

## 3.2 Methods

### 3.2.1 EEG signal classification using a single temporal window

The high density EEG system simultaneously records surface potentials in multiple electrodes encompassing most of the head surface. The spatial diversity can be explored by integrating information over space to improve resolution and signal-to-noise ratio (SNR).

By linearly combining EEG channels, an aggregate representation of the data can be obtained ([31]). Let  $\mathbf{x}(t)$  be the observed vector of multidimensional EEG amplitude reading (after signal preprocessing) at time  $t$ , an optimal projection weighting vector  $\mathbf{w}_{\text{amp}}$  can be derived based on a training set and so that a one-dimensional projection  $y_{\text{amp}}(t)$  can be generated:

$$y_{\text{amp}}(t) = \mathbf{w}_{\text{amp}}^T \mathbf{x}(t) = \sum_{i=1}^N w_{\text{amp}_i} x_i(t) \quad (3.1)$$

where  $N$  is the number of EEG electrodes (channels). For the visual detection task, we can find an optimal projection weight matrix  $\mathbf{w}_{\text{amp}}$ , which can generate a  $y_{\text{amp}}(t)$  that maximally discriminates at time  $t$  corresponding to target and non-target/distractor trial conditions. It was found that the linear discriminators trained on certain crucial short temporal windows have robust single-trial discrimination performance when these discriminators were applied to analyze the spatio-temporal changes in neural processing ([31]) during visual cognitive tasks.

A more robust output can be derived by averaging over the  $T$  samples in the short temporal window  $\bar{y}_{\text{amp}} = T^{-1} \sum_{t=1}^T y_{\text{amp}}(t)$ . The hypotheses are  $\mathbf{H}_0$  (non-target/distractor in the image) and  $\mathbf{H}_1$  (target in the image). The posterior likelihood that a trial belongs to the target class is assumed to follow a logistic function distribution  $p(\mathbf{H}_1 | y_{\text{amp}}) = \frac{1}{1 + e^{-y_{\text{amp}}}}$  ([39]).

ROC curves can be obtained using  $p(\mathbf{H}_1|y_{amp})$  and comparing it with a threshold  $\theta$ .  $\theta$  can take on values within the range  $[0, 1]$ . The decision rule can be 
$$\begin{cases} p(\mathbf{H}_1|y_{amp}) \geq \theta, & u_{amp} = 1 \\ p(\mathbf{H}_1|y_{amp}) < \theta, & u_{amp} = 0 \end{cases}$$
 or vice versa where  $u_{amp} = 1$  represents a classifier's decision to declare a target and  $u_{amp} = 0$  represents a classifier's decision to declare a non-target/distractor.

In this chapter, local LDA classifiers are applied on non-overlapping 50-millisecond temporal windows with onset time ranging from the stimulus onset to 850 milliseconds following the stimulus onset. Since the length of the trial epoch is 900 milliseconds, 18 local classifiers are trained and used to classify the trials in test sets.

### 3.2.2 Feature-level fusion of multiple local classifiers' outputs

While LDA classifiers can be applied on EEG data located in multiple local temporal windows with different window onset times after stimulus events, the likelihood output values of these local LDA classifiers can be used as inputs to a global classifier as illustrated in Fig. 3.2. The objective of using the global classifier is to enhance the triage performance by exploring the global spatio-temporal pattern.

In this chapter we applied two feature-level fusion methods: a linear discriminant analysis classifier (LDA) and a relevance vector machine (RVM) classifier while previous works ([25],[23]) only considered the LDA feature-level fusion. The feature-level classifiers are applied on multiple local LDA classifiers' outputs to capture the linear spatio-temporal pattern across multiple local windows.

#### Feature-level linear discriminant analysis classifier (LDA)

By linearly combining local classifier outputs, an aggregate representation of global pattern can be obtained. Let  $\mathbf{y}$  be the observed vector of local classifier outputs, a weighting vector  $\mathbf{w}_{\text{window}}$  can be derived based on training data to generate a one-dimensional projection

$z$ :

$$z = \mathbf{w}_{\text{window}}^T \mathbf{y} = \sum_{i=1}^M w_{\text{window}i} y_i \quad (3.2)$$

where  $M$  is the number of local classifiers. The projection  $z$  serves as an estimate of global pattern. The likelihood of a trial belonging to the target class is assumed to follow a logistic function distribution  $p(\mathbf{H}_1|\mathbf{y}) = \frac{1}{1+e^{-z}}$ .

ROC curves can be obtained using  $p(\mathbf{H}_1|\mathbf{y})$  and comparing it with a threshold  $\eta$ .  $\eta$  can take on values ranged in  $[0, 1]$ . The decision rule can be  $\begin{cases} p(\mathbf{H}_1|\mathbf{y}) \geq \eta, & u = 1 \\ p(\mathbf{H}_1|\mathbf{y}) < \eta, & u = 0 \end{cases}$  or vice versa where  $u = 1$  represents a classifier's decision to declare a target and  $u = 0$  represents a classifier's decision to declare a non-target/distractor.

### Feature-level Relevance Vector Machine (RVM) classifier

A RVM classifier ([42], [43]) models the likelihood of a trial belonging to the target class as a sigmoid logistic function distribution:

$$p(\mathbf{H}_1|\mathbf{y}) = \frac{1}{1 + e^{-f_{RVM}(\mathbf{y})}} \quad (3.3)$$

where  $f_{RVM}$  is given by:

$$f_{RVM}(\mathbf{y}) = \sum_{i=1}^M \alpha_i K(\mathbf{y}, \mathbf{y}_i) + b \quad (3.4)$$

where  $K(\mathbf{y}, \mathbf{y}_i)$  is the kernel function (the first order polynomial kernel is used in this study where  $K(\mathbf{y}, \mathbf{y}_i) = \mathbf{y}^T \mathbf{y}_i + 1$ ),  $\alpha_i$  is the weight parameter for each local classifier output and  $b$  is a threshold.

To determine the  $\alpha_i$  using a Bayesian approach, they are encoded to have a sparse prior: statistically independent from each other and following a zero-mean, Gaussian distribution with variance  $\lambda_i^{-1}$ . In addition, a gamma distribution is assumed on the hyper-parameter



$\lambda_i$ . Therefore, prior  $\alpha_i$  are highly concentrated around 0 and generate very few nonzero terms in  $f_{RVM}(z)$ .

Let  $\lambda_i^*$  be the maximum a posterior (MAP) estimate of the hyperparameter  $\lambda_i$ . The MAP estimate for the parameters  $\alpha_i$  can be obtained by maximizing the posterior distribution of the class labels given the training set. It is equivalent to maximize the following objective function:

$$J(\alpha_1, \alpha_2, \dots, \alpha_M) = \sum_{i=1}^M \log(p(\mathbf{H}_i|\alpha_i)) + \sum_{i=1}^M \log(p(\alpha_i|\lambda_i^*)) \quad (3.5)$$

where the first term represents the likelihood of the class labels and the second term represents the prior on the parameters  $\alpha_i$ . Only the samples associated with nonzero coefficients  $\alpha_i$  (called relevance vectors) can contribute to the decision function.

ROC curves can be obtained using  $p(\mathbf{H}_1|\mathbf{y})$  and comparing it with a threshold  $\phi$ .  $\phi$  can take on values ranged in  $[0, 1]$ . The decision rule can be  $\begin{cases} p(\mathbf{H}_1|z) \geq \phi, & u = 1 \\ p(\mathbf{H}_1|z) < \phi, & u = 0 \end{cases}$  or vice versa where  $u = 1$  represents a classifier's decision to declare a target and  $u = 0$  represents a classifier's decision to declare a non-target/distractor.

### 3.2.3 Decision-level fusion of multiple local classifiers' outputs

Decision-level fusion achieves fusion gain by combining binary decisions (given by individual local classifiers) optimally to determine the presence or absence of a target in a stimulus image. While the feature-level fusion relies on detecting a global pattern across multiple local classifier outputs, decision-level fusion explores complementarities of local classifiers' decisions ([44], [45]).

An effective approach is to use Bayesian inference where local classifiers' binary decisions (target or non-target/distractor) are treated as multiple hypotheses or evidences that need

to be combined optimally. The hypotheses are  $\mathbf{H}_0$  (non-target/distractor in the image) and  $\mathbf{H}_1$  (target in the image). The local classifier output vector has joint probability density function  $P(y_1, \dots, y_k | \mathbf{H}_j)$  under hypothesis  $\mathbf{H}_j$ , for  $j = 0, 1$  and  $k = 1, 2, \dots, M$ . For individual local amplitude-based classifiers, they receive as inputs the N-dimension observation vector  $\mathbf{x}$  (amplitude) and make the decisions based on LDA classifier outputs (given a fixed value of decision threshold). The decisions drawn from  $M$  individual local classifiers are denoted as  $u_k$ , where  $k = 1, 2, \dots, M$  and  $u_k = 0$  if the  $k_{th}$  local classifier decides  $\mathbf{H}_0$  and  $u_k = 1$  if the  $k_{th}$  local classifier decides  $\mathbf{H}_1$ . Individual classifier's decision  $u_k$  depends only on the local classifiers' output vectors  $\mathbf{y}$ .

$$u_k = \alpha(\mathbf{x}_k) = \begin{cases} 0, & k_{th} \text{ amplitude-based local classifier decides } \mathbf{H}_0, \\ 1, & k_{th} \text{ amplitude-based local classifier decides } \mathbf{H}_1. \end{cases}$$

The performance characteristic of individual classifier  $k$  can be specified by  $P(u_k | \mathbf{H}_j)$ , where  $P(u_k=1 | \mathbf{H}_0) = P_{f_k}$  = probability of false alarm and  $P(u_k=1 | \mathbf{H}_1) = P_{d_k}$  = probability of detection.

The global decision fusion classifier receives decisions of the individual local classifiers as its inputs. The decision at the fused level,

$$u = \varphi(u_1, u_2, \dots, u_k) = \begin{cases} 0, & \text{global decision } \mathbf{H}_0, \\ 1, & \text{global decision } \mathbf{H}_1. \end{cases} \text{ depends only on local decisions, their}$$

probability of detection  $P_{d_k}$ , probability of false alarm  $P_{f_k}$  and how complementary they are to each other. Since multiple local LDA classifiers base their decisions on EEG raw signals at non-overlapping temporal windows, the simplest assumption is that these decisions are statistically independent. Assuming the local decision variables are statistically independent, the likelihood ratio of the fused classifier  $\lambda$  is:

$$\lambda_{fusion}(u_1, \dots, u_M) = \frac{P(u_1, \dots, u_M | \mathbf{H}_1)}{P(u_1, \dots, u_M | \mathbf{H}_0)} = \prod_{k=1}^M \frac{P(u_k | \mathbf{H}_1)}{P(u_k | \mathbf{H}_0)} = \prod_{k=1}^M \left( \frac{P_{d_k}}{P_{f_k}} \right)^{u_k} \prod_{k=1}^M \left( \frac{1 - P_{d_k}}{1 - P_{f_k}} \right)^{1 - u_k} \quad (3.6)$$

The optimal fusion decision rule uses the fusion likelihood ratio as a classification decision variable and compares it to a threshold  $\beta$  for decision  $u$  ([44]):

$$u = F(u_1, u_2, \dots, u_M) = \begin{cases} 1, & \text{if } \lambda_{fusion}(u_1, \dots, u_M) \geq \beta \\ 0, & \text{if } \lambda_{fusion}(u_1, \dots, u_M) < \beta \end{cases} \quad (3.7)$$

The chosen threshold  $\beta$  determines the operating point on the ROC curve. By varying the threshold  $\beta$  value, the entire decision fusion ROC curve can be obtained by deriving a series of  $Pd_{fusion}(\beta)$  and  $Pf_{fusion}(\beta)$ . The decision fusion classifier's probability of detection and false alarm can be calculated based on equation (3.8) and (3.9):

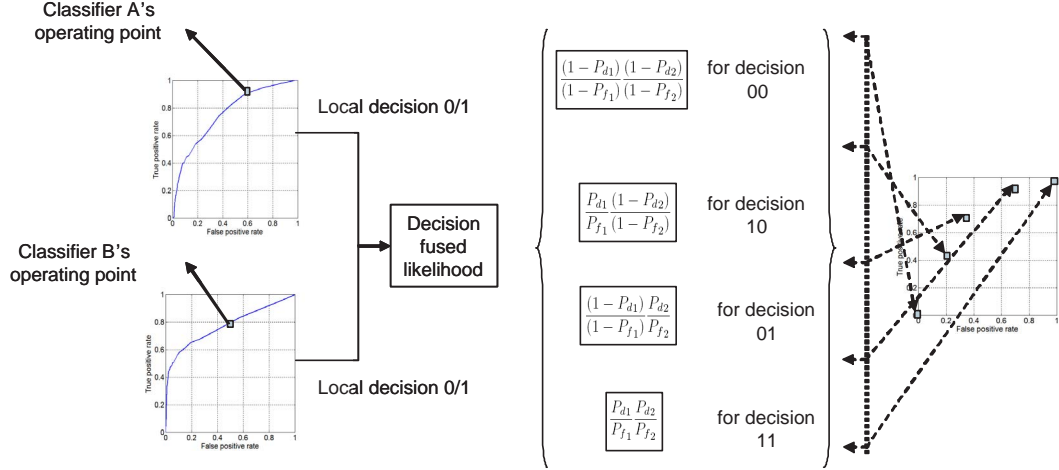
$$Pd_{fusion}(\beta) = \sum_{\lambda_{fusion} \geq \beta} P(\lambda = \lambda_{fusion} | H_1) \quad (3.8)$$

$$Pf_{fusion}(\beta) = \sum_{\lambda_{fusion} \geq \beta} P(\lambda = \lambda_{fusion} | H_0) \quad (3.9)$$

In practice, to derive the realistic ROC curve, decisions made by individual local classifiers should be compared with the ground truth so that  $P(u_1, u_2, \dots, u_M | \mathbf{H}_0)$ ,  $P(u_1, u_2, \dots, u_M | \mathbf{H}_1)$  and  $\lambda_{fusion}(u_1, \dots, u_M)$  can be evaluated. Using a training set, we can obtain ROC curves and choose the operating points associated with the specific local classifier thresholds  $\theta_k$ , and the specific global decision threshold  $\beta$ .

Fig. 3.3 and 3.4 illustrate the process of generating a global decision fused ROC curve by fusing two local classifier ROC curves: originally there are two local ROC curves; each of them has multiple operating points with different  $P_d$  and  $P_f$  (shown in Fig. 4.6 subplot (a)). Every time, one operating point from one ROC curve can be fused with one operating point from another ROC curve as illustrated in Fig. 4.6 subplot (b) and Fig. 4.5. Five fused operating points can be generated based on where we set the threshold to compare with the fused likelihood ratio in equation (3.7). Since each ROC curve can have multiple (e.g. 100) operating points, a large number of fused operating points (as illustrated in Fig. 4.6 subplot (c)) can be generated to exhaust all operating point combinations. To evaluate the optimum improvement achievable by the decision-level fusion, the convex hull of these

points (as illustrated in Fig. 4.6 subplot (d)) is used as the fused ROC curve so that this ROC curve stands for the best possible Area under ROC curve ( $A_z$ ) for the fused result. This convex hull based ROC curve is used in this chapter to evaluate the best possible decision fused classification performance.

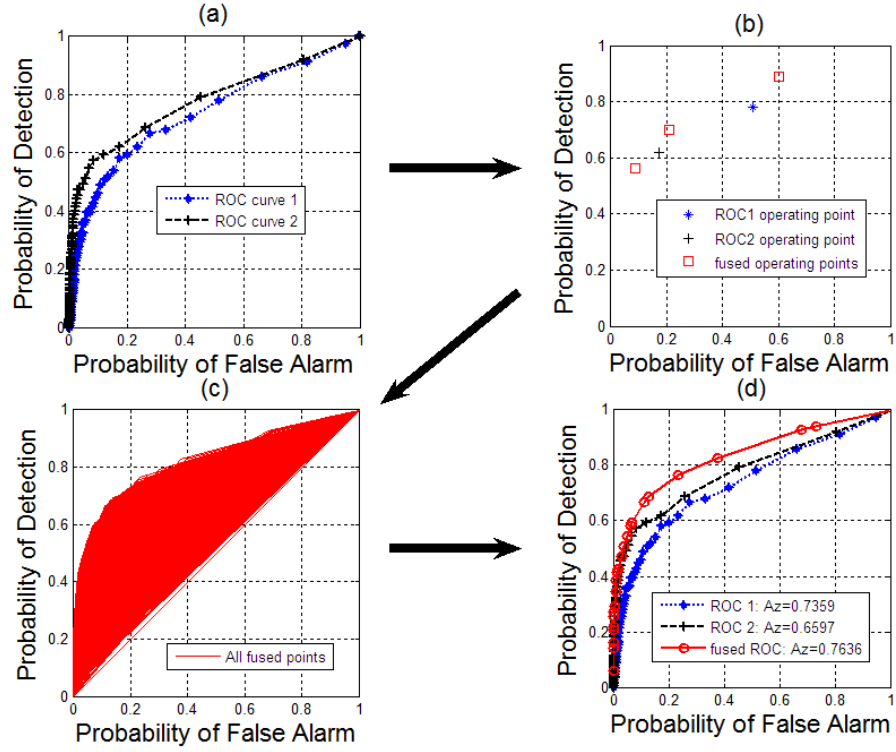


**Figure 3.3:** Fusion of two operating points from two separate ROC curves, five fused operating points can be generated depending on where the threshold is set against four fused likelihood ratios. The corresponding decision rules are: (1) never declare “detection” regardless the decisions made by both classifiers; (2) declare “detection” if at least one classifier declares “detection”; (3) declare “detection” if the better classifier (in terms of joint likelihood ratio) declares “detection”; (4) declare “detection” only if both classifiers declare “detection” (consensus); (5) always declare “detection” regardless the decisions made by both classifiers.

### 3.3 Results and discussion

#### 3.3.1 Training and test runs

The same training set of stimulus images was used for three subjects. The number of target and non-target/distractor images in the training set are listed in Table 3.1. The target image density was about 10% (while this ratio can be smaller than 1% in real applications, we have to have enough target trials to train the classifiers). The training set was used to derive weights for local LDA classifiers at different temporal windows, to derive weights for the global LDA classifier, to derive MAP weights estimate for the global RVM classifier



**Figure 3.4:** Illustration of fusing two local ROC curves: (a) two local ROC curves with multiple operating points; (b) one operating point on ROC curve 1 fuses with one operating point on ROC curve two generates five fused operating points; (c) a large number of fused operating points can be generated by exhausting all operating point combinations based on ROC curve one and ROC curve two; (d) the convex hull is used to represent the ideal fused operating points.

**Table 3.1:** Number of target and non-target/distractor images in the training set.

Non-target/distractor images	Target images
2415	256

**Table 3.2:** Number of target and non-target/distractor trials in three test sets.

<i>Testset</i>	Non-target/distractor images	Target images
1	3341	156
2	3464	27
3	2146	83

and to derive the operating points for the global decision fusion classifier.

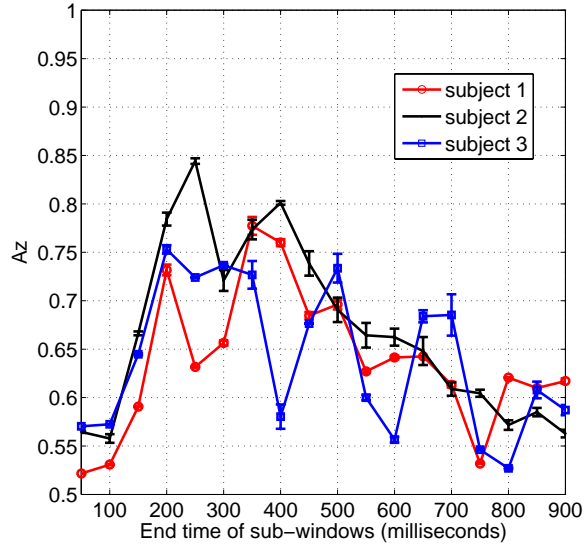
The classification performance is evaluated based on multiple test sets. There are three different test sets that were used in the experiments. The number of target and non-target/distractor images of three test sets are listed in Table 3.2. The ratios of target versus non-target/distractor images differ in different test sets so that we can evaluate the classification performance for different target density profiles.

### 3.3.2 Local linear discriminant analysis (LDA) classifier performance

The classification performance can be evaluated using the area under ROC curve  $A_z$ : chance performance corresponds to  $A_z = 0.5$  and perfect classification performance corresponds to  $A_z = 1$ . Fig. 3.5 shows the average  $A_z$  time course for running LDA classifiers on 18 50-millisecond local time windows for each subject over three test sets (and variance error bar). The area under ROC  $A_z$  value are plotted as a function of the end of 50-millisecond discrimination windows.

There is variability among different subject's  $A_z$  time course as a function of 50-millisecond discrimination windows. For example, subjects 2 and 3 reach their  $A_z$  peaks earlier than subject 1 reaches hers. It shows that each subject has its own consistent visual target detection signatures across consecutive local time windows after the visual stimulus onset.

We can observe that the classification performance based on 50-millisecond windows is poor at the first few windows after the visual stimulus onset due to the lag between stimulus onset and response phenomenology. We can also observe that the  $A_z$  values significantly increase at later time windows: for subject 1,  $A_z$  peaks at the 300-350 millisecond window; for subject 2,  $A_z$  peaks at the 200-250 millisecond window; for subject 3,  $A_z$  peaks at the 150-200 millisecond window. A significant  $A_z$  value in a specific local temporal window can serve as the indication that the signature of the visual target detection is very strong for the corresponding temporal window.



**Figure 3.5:** Local LDA classifier performance is evaluated using average area under ROC curve as a function of time for three subjects across three test sets. The variance error bars across test sets are also plotted.

### 3.3.3 Feature-level fusion: feature-level linear discriminant classifier (LDA) performance

A two-level classification structure described in Fig. 3.2 (with a global LDA classifier) is used for detecting the global visual target detection signature. The outputs of multiple local LDA classifiers are fed into the global LDA classifier as inputs. The global LDA classifier projection weight matrix is trained based on the training set and the performance

is evaluated on all test sets.

At the end of  $k_{th}$  time window (where  $k$  can be 1, 2, ... 17 or 18), the outputs of the local LDA classifiers from time window 1 to time window  $k$  are fed into the global classifier. By feeding the local classifier outputs from all existing windows into the global classifier, we can take full advantage of all existing information up to  $k_{th}$  time window. The average  $A_z$  (and the variance errorbar) time course as a function of the end of 50-millisecond discrimination windows for three subjects are plotted on Fig. 3.6, 3.7 and 3.8. The classification performance ( $A_z$ ) of the two-level linear classifier structure improves as outputs from more time windows join in and enhance the global classification performance.

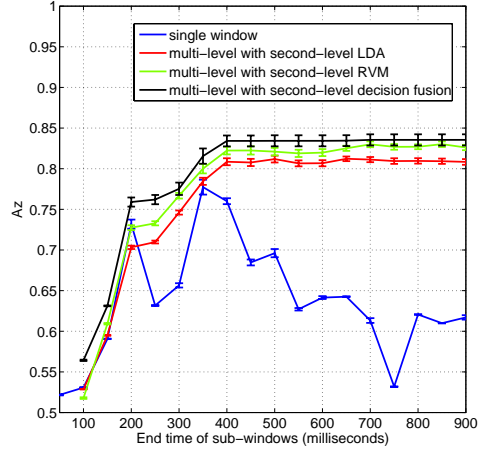
Among all test sets for all subjects, as the  $k$  increases to a certain threshold, the classification performance reaches an asymptote. After that threshold time window, classifier outputs from later time windows enhance the overall performance, insignificantly or even degrade the overall performance. For subject 1, the 350-400 millisecond window is the threshold window. For subject 2, the 200-250 millisecond window is the threshold window. For subject 3, the 250-300 millisecond window is the threshold window. The performance threshold can be interpreted as the point at which the classifier maximizes its capture of the cognitive response phenomenology at this period of time and captured signatures from later time windows do not contribute significant information.

### 3.3.4 Feature-Level fusion: feature-level RVM classification performance

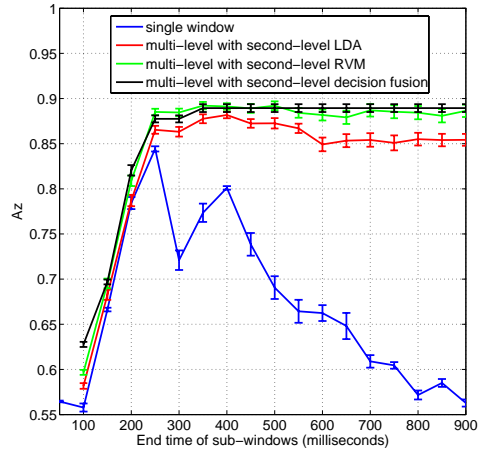
A two-level classification structure described in Fig. 3.2 (with a second-level RVM classifier) is used for detecting the global visual target detection signature. The outputs of multiple local LDA classifiers are fed into the second level RVM classifier as inputs. The second-level RVM classifier is trained based on the training set and the performance is evaluated based on the test set one, two and three.

At the end of time window  $k$  (where  $k$  can be 1, 2, ... 17 or 18), the outputs of the

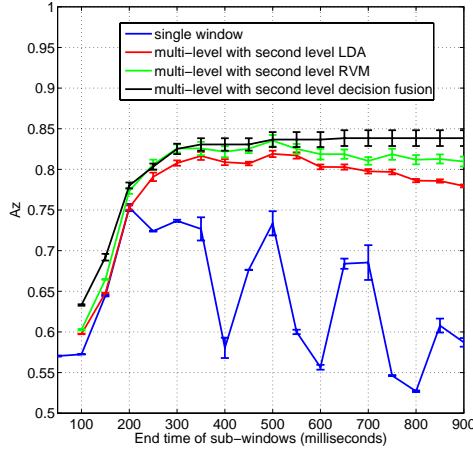




**Figure 3.6:** Subject 1: average area under ROC curve as a function of time for different fusion schemes. The variance error bars across test sets are also plotted.



**Figure 3.7:** Subject 2: average area under ROC curve as a function of time for different fusion schemes. The variance error bars across test sets are also plotted.



**Figure 3.8:** Subject 3: average area under ROC curve as a function of time for different fusion schemes. The variance error bars across test sets are also plotted.

local LDA classifiers from time window 1 to time window  $k$  are fed into the global classifier. By feeding the local classifier outputs from all existing windows into the global classifier, we can take full advantage of all existing information up to time window  $k$ . The average  $A_z$  time course as a function of the end of 50-millisecond discrimination windows for three subjects (and the variance errorbar) are plotted on Fig. 3.6, 3.7 and 3.8. The classification performance ( $A_z$ ) of the two-level RVM classifier structure improves as outputs from more time windows join in and enhance the global classification performance.

Again, among all test sets for all three subjects, the classification performance stops improving significantly after a “threshold time window”. After that time window, classifier outputs from later windows only enhance the overall performance insignificantly or even degrade performance. For subject 1, the 350-400 millisecond window is the threshold window. For subject 2, the 200-250 millisecond window is the threshold window. For subject 3, the 250-300 millisecond window is the threshold window. For each subject, the classification performance for the global RVM classifier reaches its threshold at approximately the same time window as the classification performance for the global LDA classifier reaches its threshold.

### 3.3.5 Global decision-level fusion classification performance

A two-level classification structure described in Fig. 3.2 (with a global decision fusion classifier) is used for exploring the complementarities among decisions made by local LDA classifiers. The decisions (instead of classifier outputs) given by local classifiers are fused to gain enhanced global classification performance. Even though theoretically we can fuse decisions from as many local classifiers as we want, the computational load and the number of training cases required to train a global decision fusion classifier increase substantially as the number of local classifiers increases. We only attempted fusion of two local LDA classifier decisions in this chapter. At the end of time window  $k$  (where  $k$  can be 1, 2, ... 17 or 18), decision fusions are attempted on all possible combinations of two local classifiers located at existing 50-millisecond time windows  $i$  and  $j$  ( $i \leq k, j \leq k, i \neq j$ ) and the optimal combination that gives the best area under ROC curve ( $A_z$ ) is used as the performance measure of decision fusion for  $k_{th}$  time window.

As we can observe from the performance plots (Fig. 3.6, 3.7 and 3.8), the classification performance improves when additional useful information are added by the new local classifier's decision and the performance stops improving after a "threshold time window" is hit. That implies that decisions given by the local classifiers after the "threshold time window" do not add additional performance. For each subject, the classification performance for the decision-level fusion classifier reaches its threshold at approximately the same time window as the classification performance for the global LDA classifier and the global RVM classifier reach their threshold.

In practice, we can identify the two best local time windows during training and use the same two local times windows for all test sets.

### 3.3.6 Appropriate Number of Local Classifiers

Regardless of the global classifier we choose to use, the location of the "threshold time window" is consistent for each individual subject. To include the local classifier outputs

from time windows beyond the “threshold time window” into the global classifier does not contribute significant performance. We are better off to exclude these additional local classifiers to reduce the computational load, to increase the detection speed and to improve the processing throughput.

For subjects 1 and 3, the “threshold time window” is located at 350-400 milliseconds. For subject 2, the “threshold time window” is located at 200-250 milliseconds. Instead of using all 18 local classifiers that covers from stimulus onset to 900 milliseconds following stimulus onset, we should use eight local classifiers that covers from stimulus onset to 400 milliseconds for subjects 1 and 3, and we should use five local classifiers for subject 2 that covers from stimulus onset to 250 milliseconds. The processing speed can be improved accordingly. For example, the processing speed for subject 2 can be improved 3.6 times by using five local classifiers instead of 18 local classifiers.

### 3.3.7 Comparison of global classifiers

On average, the global RVM classifier and decision-level fusion classifier (using two local classifier decisions) show better performance than the global LDA classifier. RVM classifier is less susceptible to the noise and channel failure because it only uses the relevance vectors for its decision function and decision-level fusion classifier is less susceptible to the noise because decision level is more robust and resistant to noise and channel failures.

Compared to the LDA classifier, the RVM classifier and the decision-level fusion classifier take longer time for training: the RVM classifier has a time-consuming hyperparameter estimation routine while the decision level fusion classifier has to exhaust all combinations of operating points from two local classifiers’ ROC curves in order to find an optimized fused operating point.

On the other hand, the decision-level fusion classifier requires less computational time for testing. The decision-level fusion classifier using the data from two local classifiers can achieve equivalent or better performance than the LDA classifier (that requires data from

five local time windows for subject 2; data from seven to eight local times windows for subjects 1 and 3 to achieve their maximum performance). In addition, the decision-level classifier only needs to know two local classifier decisions and it would make the fused decision accordingly and there is no calculation needed. The RVM classifier takes a similar amount of computational time as the LDA classifier for test trials but achieves a little better accuracy than the LDA classifier.

In real applications, the training set is usually much smaller than the testing set and yet the classification speed for the test set is more crucial. Therefore, the global decision-level fusion classifier is a favorable solution.

### 3.4 Summary

This chapter applies a series of fusion techniques to EEG signals during RSVP presentation of visual imagery, in order to improve the accuracy of single-trial neural signature detection of object recognition events. Using a two-level fusion scheme, multiple linear classifiers are applied on EEG amplitude signals from multiple non-overlapping temporal windows and the local classifier outputs are fed into a global feature-level classifier (either an LDA classifier or a RVM classifier) to establish the weighting of each local classifier’s output. In the alternative Bayesian decision fusion scheme, individual decisions obtained by different local classifiers are combined at the decision-level using Bayesian fusion. The experiment results show that the decision-level global fusion and the global fusion scheme with a global feature-level RVM classifier give better classification performance in terms of the area under ROC curve than the two-level fusion scheme with a global LDA classifier scheme. We also find that, for each of the individual subjects, the performance of different fusion schemes peaks at approximately the same time period following the visual stimulus onset. The location of the “threshold” time window is unique for each individual subject due to his/her unique brain response phenomenology. In real application systems that require fast processing speed and high processing throughput, the “threshold” time window for a subject can be

identified using a training set so that we can keep the number of local classifiers incorporated in the two-level fusion scheme at a minimum. The global decision-level fusion classifier using decisions from only two local classifiers can achieve better classification accuracy than the global feature-level LDA classifier. Even though the decision-level fusion classifier needs longer training time, the classification computational load for test trials can be reduced because it only needs decisions from two local classifiers and no computation is needed for the fused decision.

## Chapter 4

# Enhancement of EEG-Based Single Trial Visual Recognition Signature Detection Using Temporal and Frequency Measures at Various Frequencies

### 4.1 Motivation

While raw EEG signals measured on the scalp are in the form of voltage amplitude, spectral band power measures are frequency-domain signal features that can be extracted from the raw voltage data. According to the oscillatory brain theory developed by Basar et al. ([46], [26] and [47]), EEG consists of the activities of an ensemble of generators producing rhythmic oscillations in different frequency bands. The synchronized evoked (event-related) rhythms (oscillations) can be generated when the generators are coupled in a coherent way by means of a sensory stimulus or a cognitive event. After stimulation, responses within certain oscillatory frequency bands reflect different cognitive processes and mental states. These responses are distributed in the brain in a selective way (they exist in the brain stem, in the cerebellum, in the thalamus, in the limbic system, and in sensory/association/motor areas of the cortex). The event-related spectral band power measures can be used to quantify the event-related synchronization/de-synchronization that results in spectral power enhancement ([35]) or blocking within certain frequency bands ([48]). Thus, it makes sense to relate the oscillating spectral band power to the dynamics of cognitive processes.

Multiple previous studies provide evidences that the EEG power patterns in the delta, theta, alpha and gamma frequency ranges are related to the cognitive performance ([47], [34], [49], [33]). For example:

1. The delta responses were found to be related to target matching recognition and

decision making tasks ([50]).

2. The event-related theta oscillations were found to be involved in several tasks (focused attention, signal detection, anticipation and expectation) ([50]).

3. The alpha activity was found to be strongly correlated with phasic alertness (e.g. alertness change after presentation of a target stimulus) and expectancy ([49]).

4. The gamma oscillatory neural activity was related to the comparison of memory contents with stimulus-related information and the utilization of signals derived from this comparison. The early gamma-band response reflects the match between bottom-up and top-down information and the late gamma-band activity reflects the readout and utilization of the information resulting from this match ([51]).

5. The desynchronized neural firing (power decrease) within the alpha band and the synchronized neural firing (power increase) within the theta band can be found in response to a variety of cognitive tasks ([52], [49]).

6. It was found that the frequency components of the event-related potential vary independently of each other in terms of their relation to the event and their topographic distribution ([46]).

In this chapter, single-trial analysis using a two-level linear discriminant analysis (LDA) scheme is applied to EEG spectral power amplitude within delta, theta, alpha and gamma bands to detect spatio-temporal patterns located at various frequencies. We found that the spectral power spatio-temporal patterns at various frequencies are different from the temporal-domain (voltage amplitude) spatio-temporal pattern, and therefore can be used to enhance the overall detection performance.

In addition, this chapter presents a decision-fusion based enhancement method for detecting brain activity pattern during visual target recognition process by combining the temporal and the frequency based classification results.

A popular and simple way to think about the relationship between brain (neural ensembles) dynamics and cognitive states is to describe the sequence of brain areas that “light up”



during the various stages in the performance of a cognitive task ([53]). Consequently, the evolving spatial pattern in the temporal domain is the most important basis for cognitive state detection.

A more complex view treats the cognitive dynamics as reverberations of reentrant feed-forward and feedback processes in a complex neural network ([54]). The interactions of these parallel, reentrant feed-forward and feed-back processes are better represented by the coordinated associated oscillations of neuron cell assemblies in a hebbian sense ([55], [52]). From this perspective, we should focus on analyzing the frequency characteristics to emphasize the role of oscillation synchronization (power enhancement) and oscillation de-synchronization (power blocking). In this chapter, a multi-level LDA method is applied to analyze the EEG delta/theta/alpha/gamma band power amplitude over multiple EEG channels after visual stimulus onset and draw detect/no-detect decisions.

Since the voltage amplitude based (referred as temporal-domain response/method in this chapter from now on) and the power spectral amplitude (referred as frequency-domain response/method in this chapter from now on) based decisions are founded on alternative perspectives on brain cognitive dynamics, the fused classification decision derived using the combination of two decisions can significantly improve target detection accuracy and reduce false alarm rate.

## **4.2 Methods**

### **4.2.1 Selective distributed oscillatory systems in the brain**

According to the oscillatory brain theory, the brain can generate evoked rhythms (oscillations) when it is brought into an excited state by means of a sensory stimulus or a cognitive event. These responses are distributed selectively. The event related spectral power measure can be used to quantify the event-related synchronization that results in spectral power enhancement ([35]) or event-related de-synchronization that results in spectral power blocking within certain frequency bands ([48]). Thus, it makes sense to relate

the oscillating spectral power to the dynamics of cognitive processes.

We use the EEG recordings to reveal the activities of an ensemble of neuronal oscillatory generators producing synchronized oscillatory activities at various frequency bands in response to the cognitive events. The EEG frequencies most commonly under examinations by cognitive neuroscientists range from 0.5Hz to 100Hz. EEG frequencies are traditionally subdivided into frequency bands such as delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (13-30 Hz) and gamma bands (30-60 Hz) ([2]). For example, the amount of EEG power enhancement/blocking in the theta and alpha frequency range were found to be related to cognitive events ([34] [33] [26]).

### **4.2.2 Time-frequency power map of EEG time series**

EEG records a time series of electrical voltages across multiple sites on the scalp. EEG signals are temporal-domain voltage signals in their raw format. Demiralp et al. ([56]) pointed out that the event-related EEG potential analyses were traditionally executed in the time-domain, where only the EEG amplitude and latencies of some prominent peaks and valleys (e.g. P300 response) are considered; therefore, the information contributed by frequency domain is missing. On the other hand, in frequency-domain representation (e.g. conventional frequency power spectra), the temporal relations are lost. Demiralp went on to claim that it is more beneficial to analyze the EEG signals in time-frequency planes so that patterns at both domains can be captured.

Time-frequency analysis reveals the temporal characteristics of the EEG signal by its spectral components in the frequency domain so that one can determine both the dominant frequency of variability and their development over time. Short-Time Fourier Transform (STFT), developed by Gabor ([57]), provides the time information by computing different Fourier transforms for consecutive time intervals (windows), putting them together and maps a 1-D time domain signal to a 2-D time-frequency signal. The drawback is that once a particular size is chosen for the time window, that window is the same for all frequencies

(the resolution is set for both time and frequency). A wide analysis window offers poor time resolution but good frequency resolution. A narrow analysis window, on the other hand, offers good time resolution but poor frequency resolution.

The wavelet transform ([58]) overcomes the preset resolution problem by generating windows from dilation/compression of translating (shifting) and/or scaling a single prototype basis function (so-called parent wavelet function). Narrower windows are used for high frequency analysis and wider windows are used for low frequency analysis. While discrete wavelet transform (DWT) produces scales and positions based on powers of 2 (dyadic scales and positions), the continuous wavelet transform (CWT) produces every scale in range. The redundancy in the continuous wavelet transform offers a more accurate time-frequency spectrum.

To accurately observe the spectral power of the EEG signal at certain frequency bands after stimulus onset, the continuous wavelet transform (CWT) based time-frequency decomposition ([59]) is performed on multi-channel EEG time series. For each single trial, we have a time series of  $Q$  samples (for a 500-millisecond epoch,  $Q = 60$  samples with 120 Hz sampling rate) of  $\mathbf{x}_i$  for EEG channel  $i$ . The continuous wavelet transform coefficient is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the parent wavelet function  $\psi(t)$ :

$$\mathbf{C}_i(t, f) = CWT(\mathbf{x}_i(t), \psi(t), \tau, s) = \frac{1}{\sqrt{(|s|)}} \int_0^{0.5s} \mathbf{x}_i(t) \psi^*\left(\frac{t - \tau}{s}\right) dt \quad (4.1)$$

where  $\tau$  denotes the translation parameter and  $s$  denotes the scale parameter ( $s = 1/f$  where  $f$  is the frequency) that defines the dilation/compression of the parent wavelet. By choosing different types of wavelets and dilating/compressing them, we can access different time and frequency resolutions in the time-frequency plane. In our analysis, the wavelet function selected is the Morlet wavelet  $\psi(t) = e^{-t^2/2} \cos(5t)$  ([60]) because it strikes a good balance between time and frequency resolutions.

While the inputs are EEG signal  $\mathbf{x}_i$ , a time window  $(t_1, t_2)$  and a frequency range

$(f_1, f_2)$ , a matrix of the time-frequency power decomposition can be obtained using:

$$\mathbf{P}_i(t, f) = |\mathbf{C}_i(t, f)|^2 \quad (4.2)$$

where  $t$  is any time between  $t_1$  and  $t_2$  and  $f$  is any frequency between  $f_1$  and  $f_2$ . Each element of the  $\mathbf{P}_i(t, f)$  matrix gives us the instantaneous power at a specific time and frequency combinational cell on the time-frequency plane for EEG channel  $i$ .

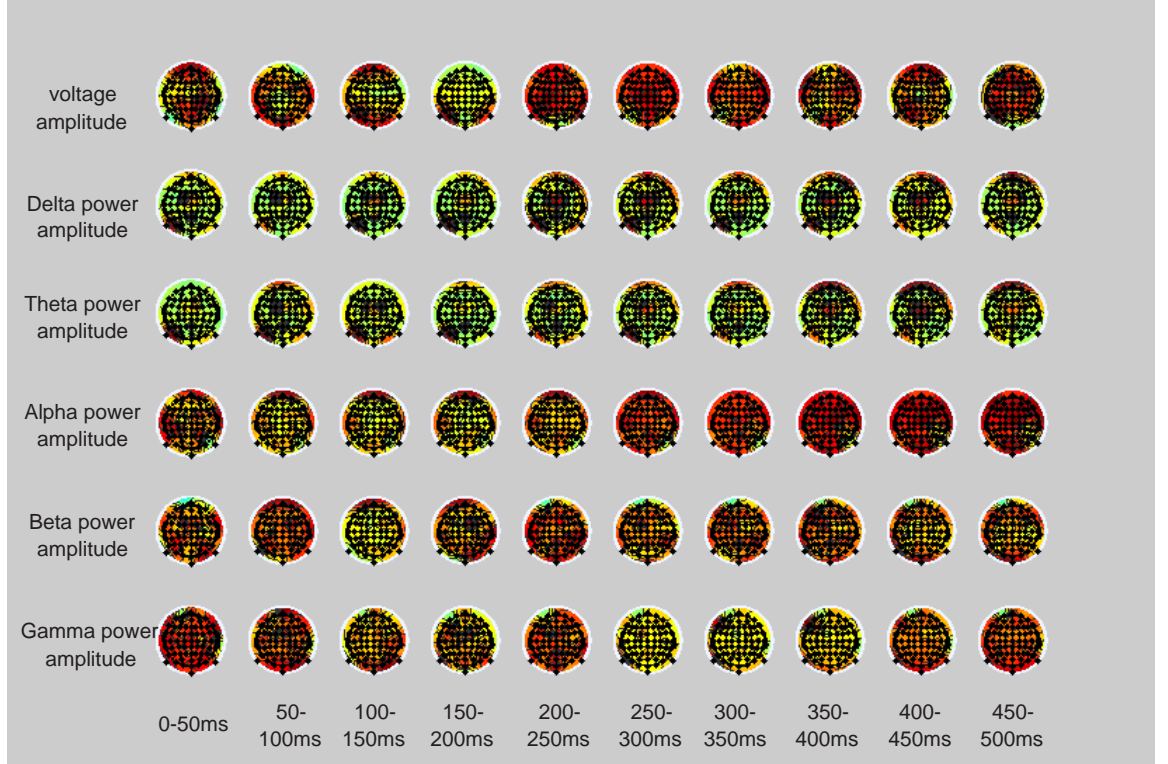
Examples of average target response and non-target/distractor response for raw voltage amplitude signals and spectral power amplitude signals at different frequency bands are shown in Fig. 4.1 based on subject 1's data. Frequency bands illustrated are the delta band, theta band, alpha band, beta band and gamma band. The average responses are based on 2415 non-target/distractor trials and 256 target trials. The plots on the top show the average response for frontal channel 38 (Fz). The plots on the bottom show the average response for rear channel 31 (Pz). The plots on the left show the average response for left parietal channel 14 (C5). The plots on the right show the average response for right parietal channel 51 (C6). Significant differences can be found:

(1) Between average temporal domain response and average frequency domain responses at different frequency bands for the same channel. For example, the frontal channel 38 (Fz) has a negative peak around 200 milliseconds and a positive peak around 300 milliseconds after stimulus onset for the average temporal domain response; the same channel has peaks and valleys at different time intervals after stimulus onset for the average frequency domain responses at different frequency bands.

(2) Among average frequency domain responses at different frequency bands for the same channel. For example, the frontal channel 38 (Fz) has a positive peak between 500-600 milliseconds after stimulus onset for the average delta band response, a positive peak around 500 milliseconds after stimulus onset for the average theta band response, a positive peak around 600 milliseconds after stimulus onset for the average theta band response, a positive peak around 100 milliseconds and a negative peak around 500 milliseconds after stimulus onset for the average beta response, a negative peak around 500 milliseconds after

stimulus onset for the average gamma band response.

(3) Among channels for the same average temporal domain response or for the same average frequency domain response. For example, the average alpha band power response has early peaks for channel 14 (C5) and 51 (C6), but late peaks for channel 31 (Pz) and 38 (Fz).



**Figure 4.1:** Topographic plots of normalized after-stimulus voltage (temporal domain) and frequency spectral power (frequency domain) amplitudes located at different frequency bands for one of the subjects; each topographic plot is based on 50-millisecond temporal window; the plot shows the evolving patterns from the stimulus onset to 500 milliseconds after onset.

Based on the above observations, we postulate that EEG temporal response and frequency responses at different frequency bands have different spatio-temporal signatures (patterns). We could draw independent classification decisions by detecting these unique spatio-temporal signatures located at different frequency bands.

### 4.2.3 EEG temporal domain based classification

The spatial diversity of a high density EEG system can be explored by integrating information over space to improve resolution and signal-to-noise ratio (SNR). By linearly combining EEG voltage amplitude data from multiple channels, an aggregate representation of the data can be obtained ([31]). Let  $\mathbf{x}(t)$  be the observed vector of multidimensional EEG voltage amplitude reading (after signal preprocessing) at time  $t$ , an optimal projection weighting vector  $\mathbf{w}_{\text{voltage}}$  can be derived based on a training set and so that a one-dimensional projection  $y_{\text{voltage}}(t)$  can be generated:

$$y_{\text{voltage}}(t) = \mathbf{w}_{\text{voltage}}^T \mathbf{x}(t) = \sum_{i=1}^N w_{\text{voltage}_i} x_i(t) \quad (4.3)$$

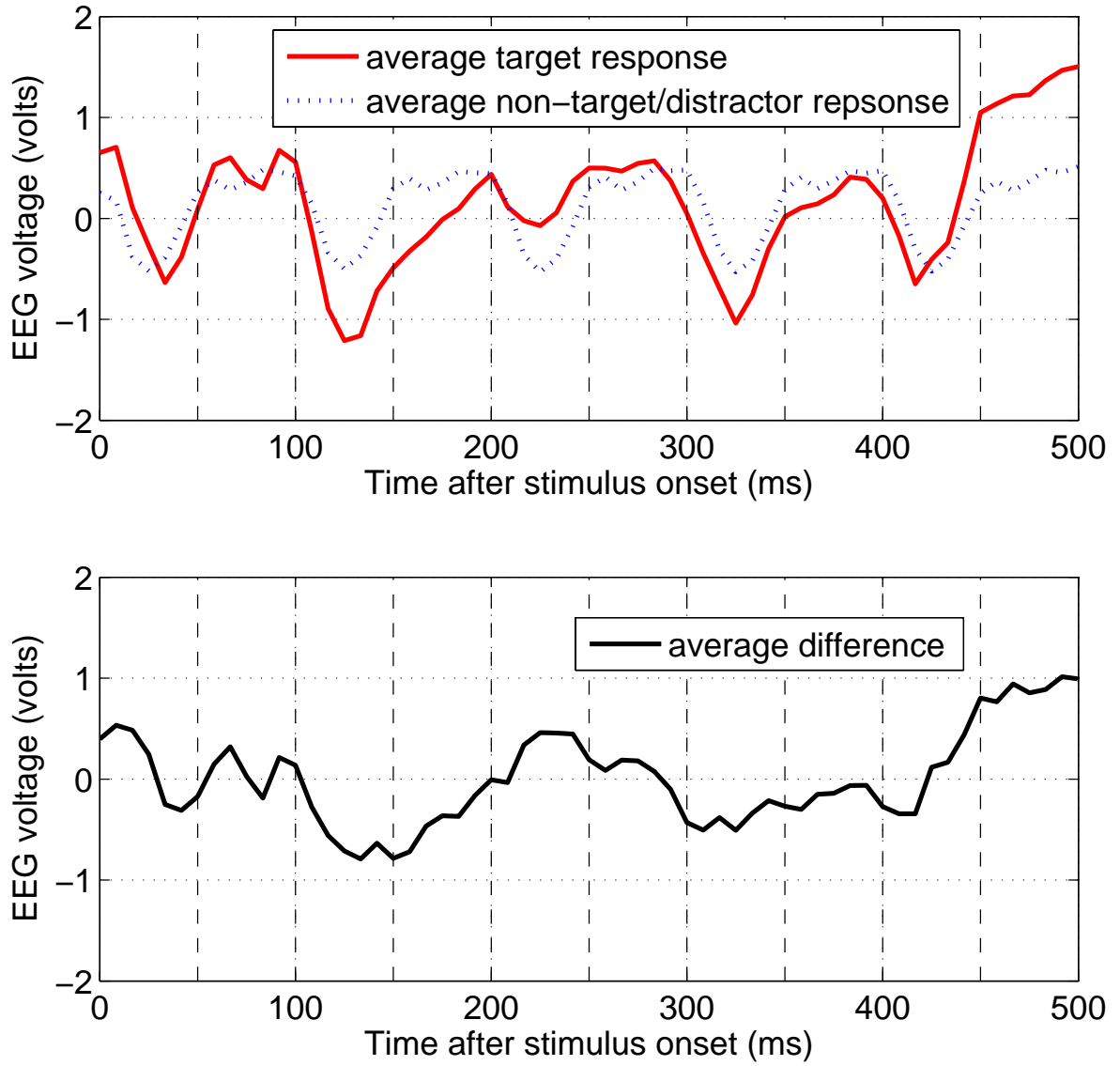
where  $N = 64$  is the number of EEG electrodes (channels). For the visual detection task, we can find an optimal projection weight matrix  $\mathbf{w}_{\text{voltage}}$ , which can generate a  $y_{\text{voltage}}(t)$  that maximally discriminates at time  $t$  corresponding to target and non-target/distractor trial condition.

By capturing neurophysiological activities located in selectively distributed sites of the brain evolving with a continuous time course, a two-level fusion scheme ([61]) can achieve further classification performance by fusing the output values of multiple first-level LDA classifiers at the feature-level (illustrated in Fig. 4.3). Let  $\mathbf{y}_{\text{voltage}}$  be the observed vector of classifier outputs, a weighting vector  $\mathbf{w}_{\text{voltagewin}}$  can be derived based on training data to generate a one-dimensional projection  $z_{\text{voltage}}$ :

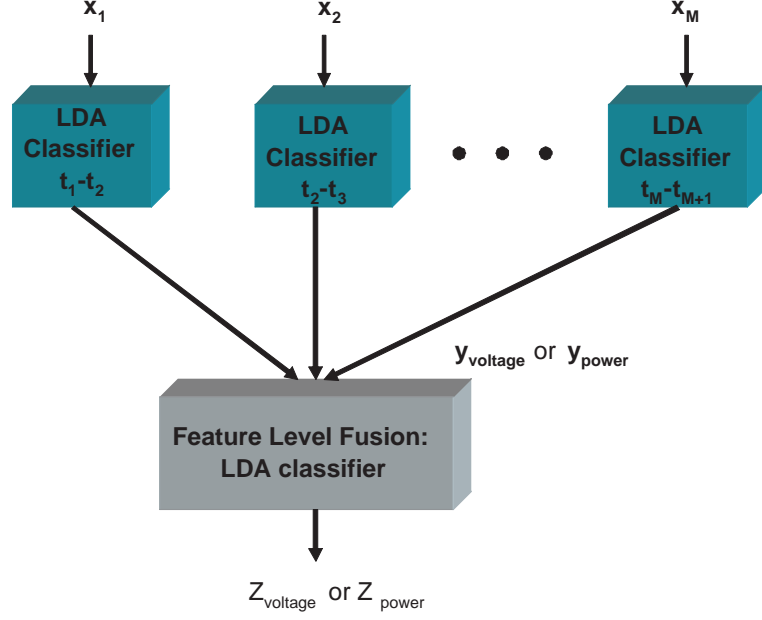
$$z_{\text{voltage}} = \mathbf{w}_{\text{voltagewin}}^T \mathbf{y}_{\text{voltage}} = \sum_{i=1}^M w_{\text{voltagewin}_i} y_{\text{voltage}_i} \quad (4.4)$$

where  $M$  is the number of first level classifiers. The projection  $z_{\text{voltage}}$  serves as an estimate of the global pattern across multiple first-level classifier temporal windows.

In this chapter, LDA classifiers are applied to non-overlapping 50-millisecond temporal sub-windows with onset time ranging from 0 milliseconds to 450 milliseconds following the



**Figure 4.2:** Top subplot shows the average target and non-target/distractor responses on a single channel; the bottom subplot shows the difference between average target and non-target/distractor responses; The whole epoch can be divided into multiple non-overlapping 50-millisecond sub-windows.



**Figure 4.3:** Feature-level fusion of multiple classifiers' outputs.

stimulus onset (Fig. 4.2). Since the trial epoch length is 500 milliseconds,  $M = 10$  classifiers based on first-level 50-millisecond windows and one feature-level classifier are trained and used to test the trials.

#### 4.2.4 EEG spectral band power based classification

In addition to applying the two-level LDA analysis on EEG voltage amplitude, we apply the two-level linear discrimination analysis on EEG spectral powers inside the delta, theta, alpha, beta and gamma frequency bands over the full set of 64 EEG channels:

$$y_{\text{power}}(t) = \mathbf{w}_{\text{power}}^T \mathbf{p}(t) = \sum_{i=1}^N w_{\text{power}_i} p_i(t) \quad (4.5)$$

where  $N = \text{number of electrodes} = 64$  and  $p(t)$  is the vector of multidimensional power in a specific band. The hypothesis here is that each frequency band has its unique spatio-temporal signature.

To explore the possible coupled power change patterns in the theta and alpha bands (power decrease in the alpha band is closely coupled with the power increase in the theta



band) ([49]), we also applied the two-level linear discrimination analysis on EEG power inside theta and alpha bands over the full set of 64 EEG channels:

$$y_{power}(t) = \mathbf{w}_{\mathbf{power}}^T \mathbf{p}_{\alpha\theta}(t) = \sum_{i=1}^{NP} w_{power_i} p_{\alpha\theta}(t) \quad (4.6)$$

where  $p_{\alpha\theta}$  is the vector of multidimensional power in both theta and alpha bands,  $NP = \text{number of electrodes} \times 2 = 128$  because both the theta and the alpha band powers are calculated for each EEG channel. The corresponding  $y_{power}(t)$  is the maximal discriminating projection of the EEG power at time  $t$  corresponding to target and non-target/distractor trial condition.

The same feature-level fusion scheme as we used in temporal domain based feature-level fusion (illustrated in Fig. 4.3) can be used to fuse the LDA classifiers' outputs and to boost the classification performance by exploring the global power based spatio-temporal pattern. By linearly combining LDA classifiers' outputs, an aggregate representation of the global pattern can be obtained. Let  $\mathbf{y}_{\mathbf{power}}$  be the observed vector of local classifiers' outputs, a weighting vector  $\mathbf{w}_{\mathbf{powerwin}}$  can be derived based on the training data to generate a one-dimensional projection  $z_{power}$ :

$$z_{power} = \mathbf{w}_{\mathbf{powerwin}}^T \mathbf{y}_{\mathbf{power}} = \sum_{i=1}^M w_{powerwin_i} y_{power_i}. \quad (4.7)$$

where  $M = 10$  is the number of local temporal windows. The projection  $z_{power}$  is the estimate of the global power pattern.

Similar to what we did for the temporal-domain classification, LDA classifiers are applied to non-overlapping 50-millisecond temporal windows with onset time ranging from 0 milliseconds to 450 milliseconds following the stimulus onset (Fig. 4.2). Since the length of the trial epoch is 500 milliseconds, 10 classifiers based on first level 50-millisecond windows and a feature-level classifier are trained and used to test the trials in test sets.

### 4.2.5 Decision-level fusion

In the two-level classification scheme introduced in the previous sections, separate classification decisions, one based on voltage (temporal-domain) amplitude and the others based on spectral power (frequency-domain) amplitude at specific frequency bands, can be derived. While temporal-based classifier captures the evolving activity levels of neuron at different spatial locations, frequency-based classifiers capture the underlying oscillation patterns associated with detection, decision making, attention and memory associative tasks. An effective fusion approach is to use Bayesian inference ([62], [44], [63], [64] and [45]) where classifiers' binary decisions (target or non-target/distractor) are treated as multiple hypotheses or evidences that need to be combined optimally. This decision-level fusion method can improve the overall classification performance further by combining the temporal-domain decision and the frequency-domain decision (illustrated in Fig. 4.5 and 4.4).

The decision-level fusion optimally fuses all local decisions according to the operating points on their ROC curves at which the local decisions were made with certain probability of detection ( $P_d$ ) and probability of false alarm ( $P_f$ ). The EEG voltage amplitude observation vector has joint probability density function  $P(\mathbf{x}|\mathbf{H}_j) = P(x_1, \dots, x_N|\mathbf{H}_j)$  under hypothesis  $\mathbf{H}_j$ , for  $j = 0$  (non-target/distractor),  $1$  (target) and  $N = 64$ . The EEG delta/theta/alpha/gamma power observation vector has joint probability density function  $P(\mathbf{p}|\mathbf{H}_j) = P(p_1, \dots, p_N|\mathbf{H}_j)$  under hypothesis  $\mathbf{H}_j$ , for  $j = 0, 1$  and  $N = 64$ . The EEG theta/alpha power observation vector has joint probability density function  $P(\mathbf{p}|\mathbf{H}_j) = P(p_1, \dots, p_{NP}|\mathbf{H}_j)$  under hypothesis  $\mathbf{H}_j$ , for  $j = 0, 1$  and  $NP = 128$ . Feature-level LDA classifiers receive as inputs the first-level LDA classifiers' outputs. The decisions drawn by feature-level classifiers are denoted as  $u_k$ , for  $k = 1$  (temporal-domain) and  $k = 2$  (frequency-domain), where  $u_k = 0$  if  $k_{th}$  classifier's decision is  $\mathbf{H}_0$  and  $u_k = 1$  if  $k_{th}$  classifier's decision is  $\mathbf{H}_1$ . Individual classifier's decision  $u_k$  depends only on the observation vectors  $\mathbf{x}$  and  $\mathbf{p}$ .

$$u_1 = \alpha(\mathbf{x}) = \begin{cases} 0, & \mathbf{H}_0 \text{ (classified as non-target/distractor) ,} \\ 1, & \mathbf{H}_1 \text{ (classified as target).} \end{cases} \quad (4.8)$$

$$u_2 = \beta(\mathbf{p}) = \begin{cases} 0, & \mathbf{H}_0 \text{ (classified as non-target/distractor) ,} \\ 1, & \mathbf{H}_1 \text{ (classified as target).} \end{cases} \quad (4.9)$$

The performance characteristics of  $k_{th}$  classifier can be specified by  $P(u_k|\mathbf{H}_j)$ , where  $P(u_k=1|\mathbf{H}_0) = P_{f_k}$  = the probability of false alarm (false positives) and  $P(u_k=1|\mathbf{H}_1) = P_{dk}$  = the probability of detection (true positives). Using these probabilities, the likelihood ratio value of a binary decision variable has a simple form as

$$\lambda_{decision}(u_k) = \frac{Pr(u_k|\mathbf{H}_1)}{Pr(u_k|\mathbf{H}_0)} = \begin{cases} \frac{P_{dk}}{P_{f_k}} & \text{if } u_k = 1 \\ \frac{(1-P_{dk})}{(1-P_{f_k})} & \text{if } u_k = 0 \end{cases} \quad (4.10)$$

The decision-level fusion classifier receives the decisions of the individual classifiers as its inputs. The decision at the fused level,

$$u = \varphi(u_1, u_2) = \begin{cases} 0, & \text{global decision } \mathbf{H}_0, \\ 1, & \text{global decision } \mathbf{H}_1. \end{cases} \quad (4.11)$$

depends only on local decisions and their probability of detection  $P_{dk}$  and probability of false alarm  $P_{f_k}$ . Since temporal and frequency domain based classifiers derive their decisions on different aspects of EEG raw signals, the simplest assumption is that these two decisions are statistically independent. the fused likelihood ratio  $\lambda$  can be derived as:

$$\lambda_{fusion}(u_1, u_2) = \prod_{k=1}^2 \lambda(u_i) = \prod_{k=1}^2 \frac{P(u_k|\mathbf{H}_1)}{P(u_k|\mathbf{H}_0)} = \prod_{k=1}^2 \left(\frac{P_{dk}}{P_{f_k}}\right)^{u_k} \prod_{k=1}^2 \left(\frac{1-P_{dk}}{1-P_{f_k}}\right)^{1-u_k} \quad (4.12)$$

The optimal fused decision rule uses the fusion likelihood ratio as a classification decision variable and then compares it to threshold  $\beta$  for decision  $u$  ([44]):

$$u = F(u_1, u_2) = \begin{cases} 1, & \text{if } \lambda_{fusion}(u_1, u_2) \geq \beta \\ 0, & \text{if } \lambda_{fusion}(u_1, u_2) < \beta \end{cases} \quad (4.13)$$

The chosen threshold  $\beta$  determines the operating point on the fused ROC curve. By varying the threshold  $\beta$  value, five operating points can be derived based on one operating point chosen on the temporal-domain classifier's ROC curve and another operating point chosen on the frequency-domain classifier's ROC curve (illustrated in Fig. 4.5).

The decision fusion classifier probability of detection and false alarm can be calculated based on equation (5.10) and (5.11):

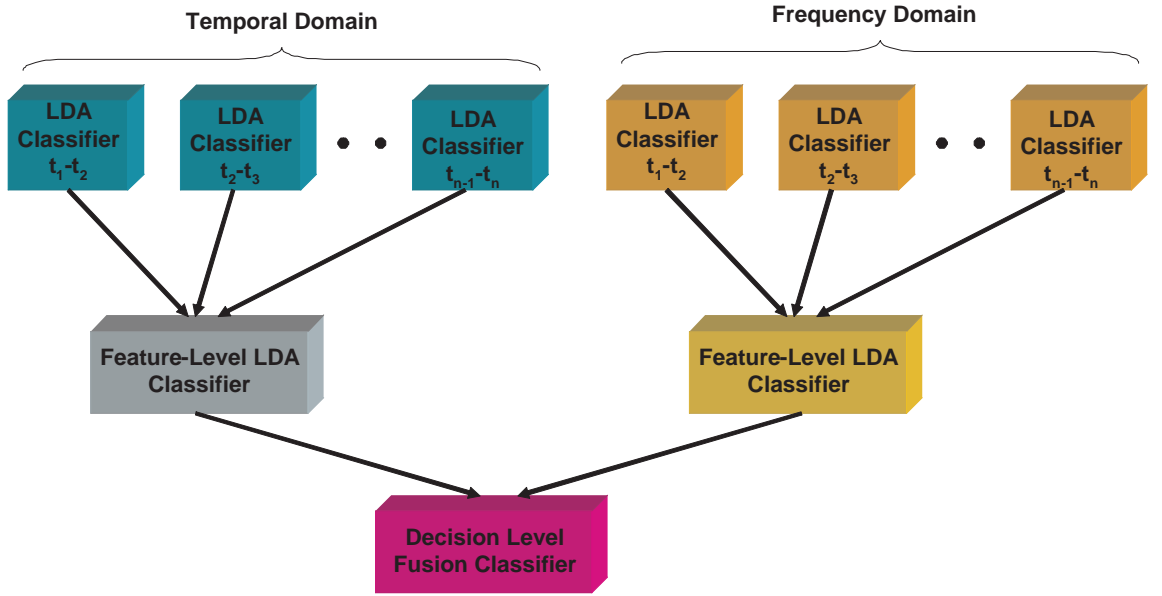
$$Pd_{fusion}(\beta) = \sum_{\lambda_{fusion} \geq \beta} P(\lambda = \lambda_{fusion} | H_1) \quad (4.14)$$

$$Pf_{fusion}(\beta) = \sum_{\lambda_{fusion} \geq \beta} P(\lambda = \lambda_{fusion} | H_0) \quad (4.15)$$

The entire decision fused ROC curve can be obtained by deriving a series of  $Pd_{fusion}(\beta)$  and  $Pf_{fusion}(\beta)$  using multiple combinations of operating points on temporal and frequency domain based classifiers' ROC curves. Fig. 4.6 illustrates the process of generating a fused ROC curve by fusing two ROC curves. There are two ROC curves; each of them has multiple operating points with different  $P_d$  and  $P_f$  (shown on subplot (a)). Every time, one operating point from one ROC curve can be fused with one operating point from another ROC curve as illustrated in subplot (b) and Fig. 4.5. Five fused operating points can be generated based on where we set the threshold to compare with the fused likelihood ratio in equation (5.8). Since each ROC curve can have multiple (e.g. 100) operating points, a large number of fused operating points (as illustrated in subplot (c)) can be generated to exhaust all operating point combinations. To evaluate the optimum improvement achievable by the decision-level fusion, the convex hull of these points (as illustrated in subplot (d)) is used

as the fused ROC curve so that this ROC curve stands for the best possible Area under ROC curve ( $A_z$ ) for the fused result.

In practice, to derive the decision-level fused ROC curve, the decisions made by the temporal-domain and frequency-domain feature-level classifiers should be compared with the ground truth using a training set so that  $P(u_1, u_2|\mathbf{H}_0)$ ,  $P(u_1, u_2|\mathbf{H}_1)$  and  $\lambda_{fusion}(u_1, u_2)$  can be evaluated empirically. Using a training set, we can obtain ROC curves and choose the operating points associated with the temporal-domain feature-level classifier, the frequency-domain feature-level classifier and the decision-level classifier.

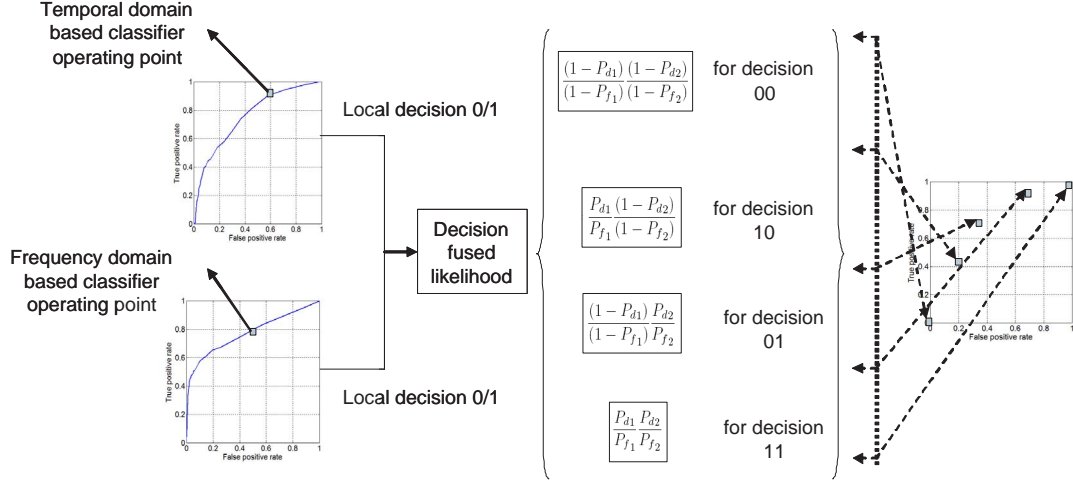


**Figure 4.4:** Decision-level fusion to combine the temporal-domain classification results and the frequency-domain classification results; the first-level LDA classifiers apply to different temporal windows; the second-level LDA classifiers are feature-level classifiers; the decision-level fusion classifier fuses the decisions made by the feature-level classifiers.

## 4.3 Results and Discussion

### 4.3.1 Training and Test Runs

The same training set of stimulus images was used for four subjects. The number of target and non-target/distractor trials in the training set are listed in Table 4.1. The training



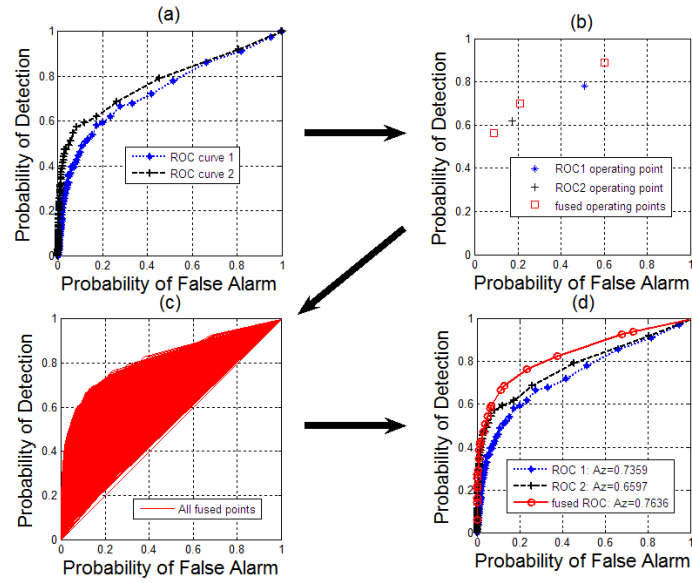
**Figure 4.5:** Fusion of two operating points from two separate ROC curves, five fused operating points can be generated depending on where the threshold is set against the four fused likelihood ratios. The corresponding decision rules are: (1) never declare “detection” regardless the decisions made by both classifiers; (2) declare “detection” if at least one classifier declares “detection”; (3) declare “detection” if the better classifier (in terms of joint likelihood ratio) declares “detection”; (4) declare “detection” only if both classifiers declare “detection” (consensus); (5) always declare “detection” regardless the decisions made by both classifiers.

**Table 4.1:** Number of target and non-target/distractor images in the training set.

Non-target/distractor images	Target images
2415	256

set was used for training the LDA classifiers applying to different 50-millisecond temporal windows, the temporal based feature-level LDA classifier, the power (frequency) based feature-level LDA classifier and the decision-level classifier (Fig. 4.4).

The classification performance is evaluated based on multiple test sets. There were two different test sets used in the experiments. The number of target and non-target/distractor images of the two test sets are listed in Table 4.2. Four subjects experimented with both test sets.



**Figure 4.6:** Illustration of fusing two ROC curves: (a) there are multiple operating points on each ROC curve; (b) five fused operating points can be generated by fusing any two operating points from each ROC curve; (c) a large number of fused operating points can be generated by exhausting all combinations of individual ROC curves' operating points; (d) the convex hull of all the fused operating points can be used to represent the optimum improved ROC curve for decision fused results.

**Table 4.2:** Number of target and non-target/distractor images in the test sets.

Test Set	Non-target/distractor images	Target images
1	3341	156
2	3464	27
sum	6805	183

**Table 4.3:** Signals used for different types of classifiers.

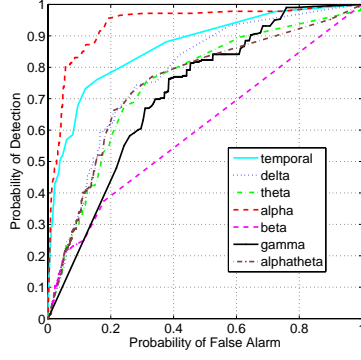
Classifiers	Signals used for the classification
voltage	64 channels of EEG voltage amplitude
delta frequency band	64 channels of delta band power amplitude
theta frequency band	64 channels of theta band power amplitude
alpha frequency band	64 channels of alpha band power amplitude
beta frequency band	64 channels of beta band power amplitude
gamma frequency band	64 channels of gamma band power amplitude
theta/alpha frequency band	128 channels of theta/alpha band power amplitude

### 4.3.2 Temporal-domain and frequency-domain classification results

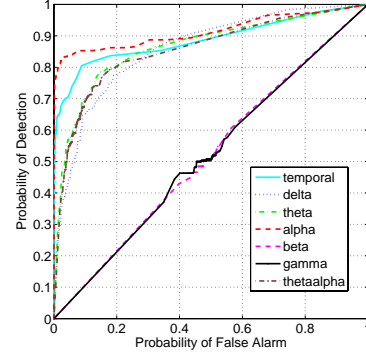
Classification results are computed based on raw EEG voltage in the temporal domain and power amplitude at different frequency bands (Table 4.3). The average ROC curves (graphical plots of correct detection rate versus false alarm rate for a binary classifier system) of two test sets for every subject are plotted on Fig. 4.7. The classification performance can be evaluated using the area under ROC curve ( $A_z$ ) as the metric. Operating points on an ROC curve with larger  $A_z$  have either a higher probability of detection (Pd) and/or a lower probability of false alarm (Pf) compared to operating points on an ROC curve with lower  $A_z$ . The average  $A_z$  for all four subjects using two test sets are listed in Fig. 4.8.

Each individual frequency band has its own unique classification performance evaluated by ROC curves, and they are different from the classification result using voltage amplitude (in temporal domain). For subjects 1, 2 and 4, the classification results using the alpha band spectral power information outperform the classification result using raw voltage amplitude signal in terms of the area under the ROC curve. For all four subjects, classification results using the beta band power have very poor performance (low  $A_z$ ) while delta band, theta

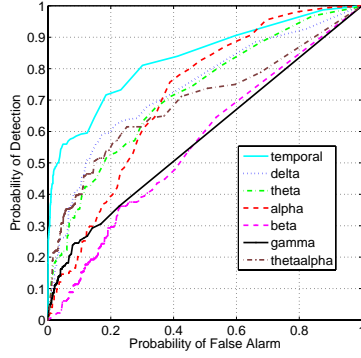




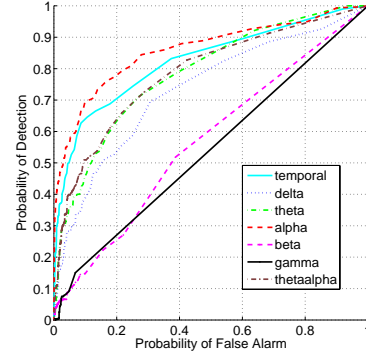
(a) Subject 1 single band ROC curves



(b) Subject 2 single band ROC curves

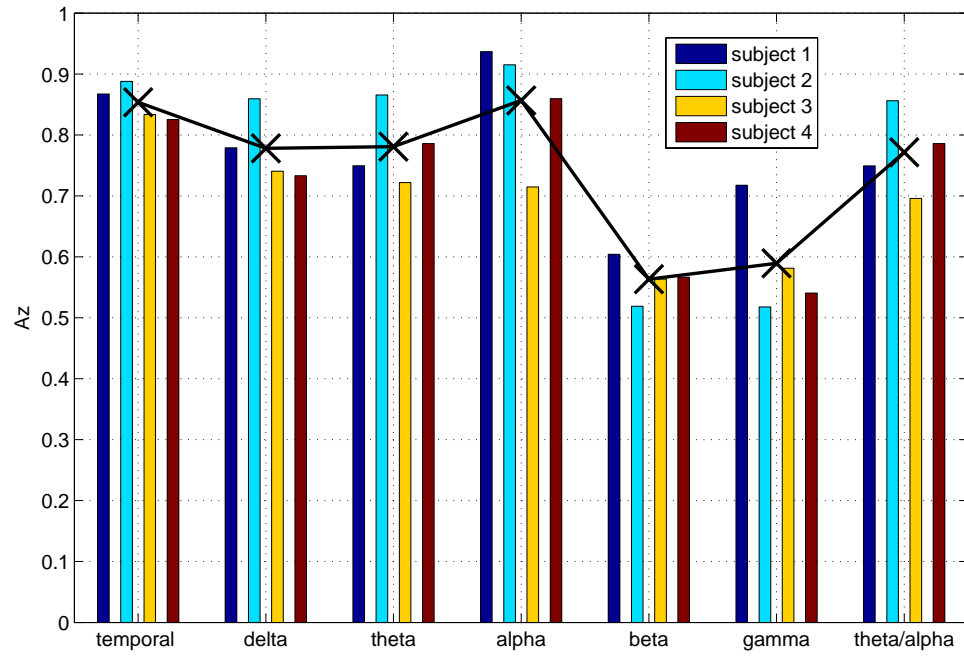


(c) Subject 3 single band ROC curves



(d) Subject 4 single band ROC curves

**Figure 4.7:** ROC curves derived based on temporal domain voltage amplitude and the frequency spectral powers located at different frequency bands (average performance over two test sets for all four subjects).



**Figure 4.8:** Average performance of all subjects using temporal domain based method and frequency spectral band power based method (the average Azs are represented by X marks).

band, alpha band perform well.

### 4.3.3 Decision fusion of temporal domain classification and frequency domain based classification results

Since the temporal-domain and the frequency-domain classification results contribute different discriminating components, we apply decision-level fusion to combine the classification results using powers from different frequency bands and the classification result using voltage amplitude (fusion scheme is illustrated in Fig. 4.4).

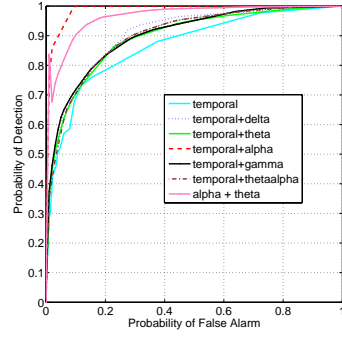
As we can observe from Fig. 4.9, the fused decisions show improved performance in terms of the area under ROC curve  $A_z$  for all four subjects over two test sets. The average  $A_z$  for all four subjects using two test sets are compared in Fig. 4.10.

For all four subjects, the classification results by fusing the temporal-domain decision and the alpha band spectral power based decision have the best performance compared to the results by fusing the temporal-domain decision and the delta/theta/gamma band frequency domain decision. Compared to other frequency band powers, alpha band power based decision seems to maximally complement the temporal domain decision in terms of the fusion gain.

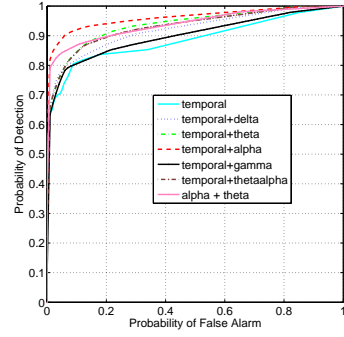
By fusing the alpha-band and beta-band information using two different methods: a combined 128-channel LDA classification method and a decision fusion method, we can see that the decision fusion method gives us better performance (Fig. 4.11). These results illustrate the importance of choosing the proper fusion methods.

## 4.4 Summary

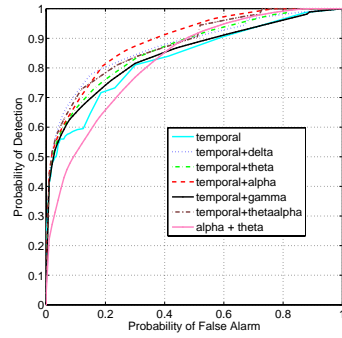
This chapter presents a method of improving the EEG-based single-trial visual detection performance by fusing classification decisions drawn from two separate EEG classification methods: a temporal-domain method based on raw voltage amplitude and a frequency-domain method based on frequency spectral power amplitude located at different bands.



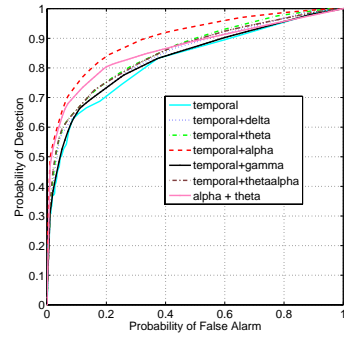
(a) Subject 1 fused ROC curves



(b) Subject 2 fused ROC curves

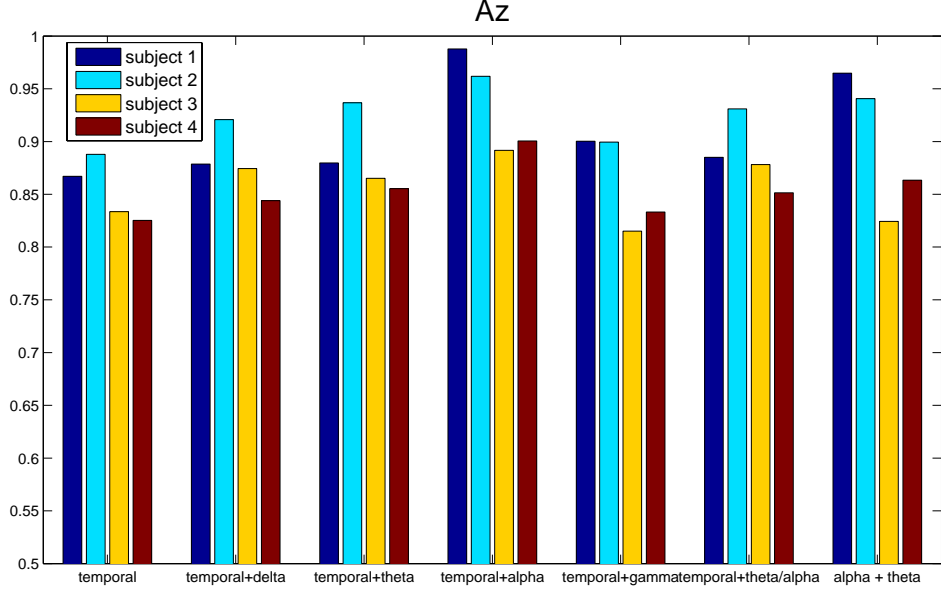


(c) Subject 3 fused ROC curves



(d) Subject 4 fused ROC curves

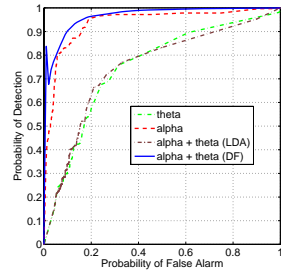
**Figure 4.9:** ROC curves for fused decisions (average over two test sets for all four subjects).



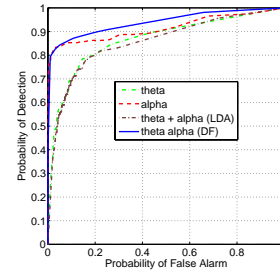
**Figure 4.10:** Average performance of all subjects by fusing temporal domain based method and frequency domain based method (the average Azs are represented by X marks).

Oscillatory activity is pervasive and ubiquitous in physiological processes. There are also a wide variety of engineering tools available to benchmark oscillatory processes (e.g. continuous wavelet transform). Spatio-temporal patterns in certain oscillatory bands was found to have the role of binding selectively distributed neuronal elements into functional assemblies for cognitive/motor tasks.

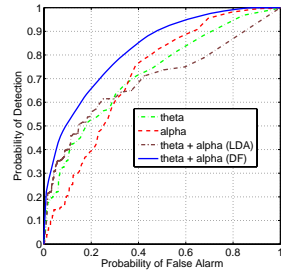
In this chapter we apply the multi-level LDA classifiers on the EEG spectral power at the delta/theta/alpha/beta/gamma frequency bands in addition to applying it on the raw EEG voltage amplitude data. Feature-level LDA classifiers are used to derive the voltage amplitude (temporal-domain) based global pattern and the spectral band power (frequency-domain) based global pattern across multiple temporal windows by fusing multiple local classifiers' outputs. We observe that the voltage amplitude based feature-level classifier and the spectral power amplitude based feature-level classifier show different visual target recognition signatures (spatio-temporal patterns) in terms of the area under ROC curve performance. The classification performance based on spectral powers located at certain



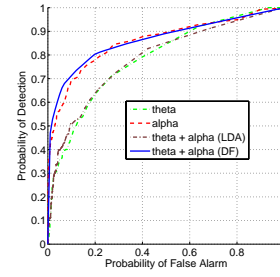
(a) Subject 1 fused ROC curves



(b) Subject 2 fused ROC curves



(c) Subject 3 fused ROC curves



(d) Subject 4 fused ROC curves

**Figure 4.11:** ROC curves for LDA classifier result based on alpha band , theta band, fused alpha/theta result using LDA classifier and fused alpha/theta result using decisions fusion (average over two test sets for all four subjects).

frequency bands (e.g. alpha band) can outperform the voltage-based (temporal-domain) classification result.

The temporal-domain decision and frequency-domain decision are then fused at decision-level. The decision-level fusion results show significant improvement over the temporal-domain results and the frequency-domain results for multiple subjects across multiple test sets.

## Chapter 5

# Decision Level Fusion of EEG-Based and Pupillary Response Based Single Trial Visual Recognition Detection

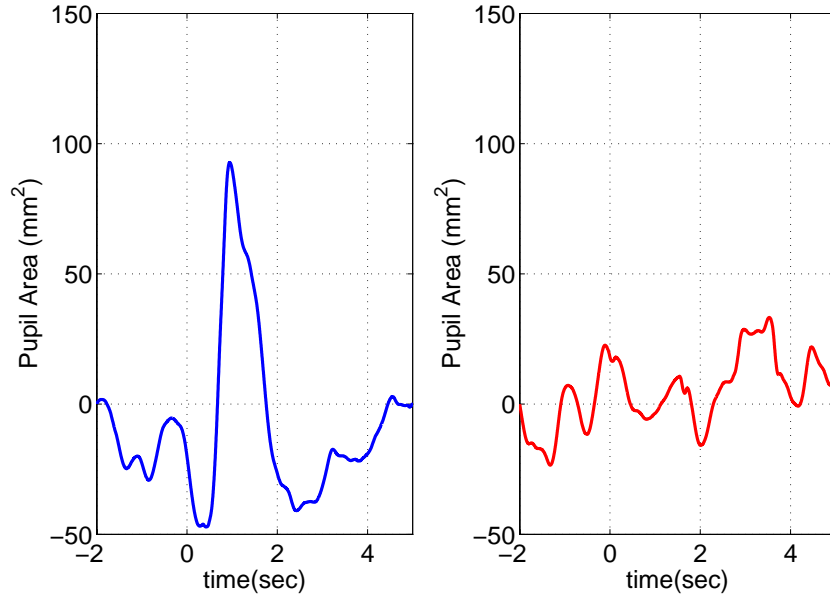
### 5.1 Motivation

Pupil response is another psychophysiological modality we can use as a basis for single-trial image triage tasks. The pupil is the opening in the iris of the eye that expands and contracts, allowing light to pass to the retina. The expansion and contraction of the pupil are mediated by the sympathetic and parasympathetic nervous systems (both are parts of the peripheral nervous system). Task-evoked pupil dilations was found to be a function of the cognitive workload and attention required to perform the task ([3], [4]). It also has long been known that the pupil dilates in response to emotion evoking stimuli ([37]). Experiments by Privitera et al. ([65]) found that the pupil responds with dilation for visual recognition events. For example, a comparison between a time course of pupil diameter change after a presentation of a target versus a non-target/distractor stimulus image is shown in Fig. 5.1. The average pupil diameter changes more drastically (dilates) after a target image display than a non-target/distractor image display.

In this chapter, we develop a pupillary feature construction and selection procedure to extract the useful features that perform best under an LDA classifier. The LDA classifier is applied on this group of pupillary response features to detect the target recognition signature.

At the same time, we apply a two-level LDA classification scheme ([61]) on cognitive task related EEG response. While the EEG method tries to capture the evolving patterns of brain dynamics across multiple pathways and functional groups, the pupillary response





**Figure 5.1:** Comparison of average pupil area change time course (after removing the baseline) after a target stimulus image display (left) and after a non-target/distractor stimulus image display (right) for an experiment subject. The pupil dilates more drastically after a target stimulus image is displayed.

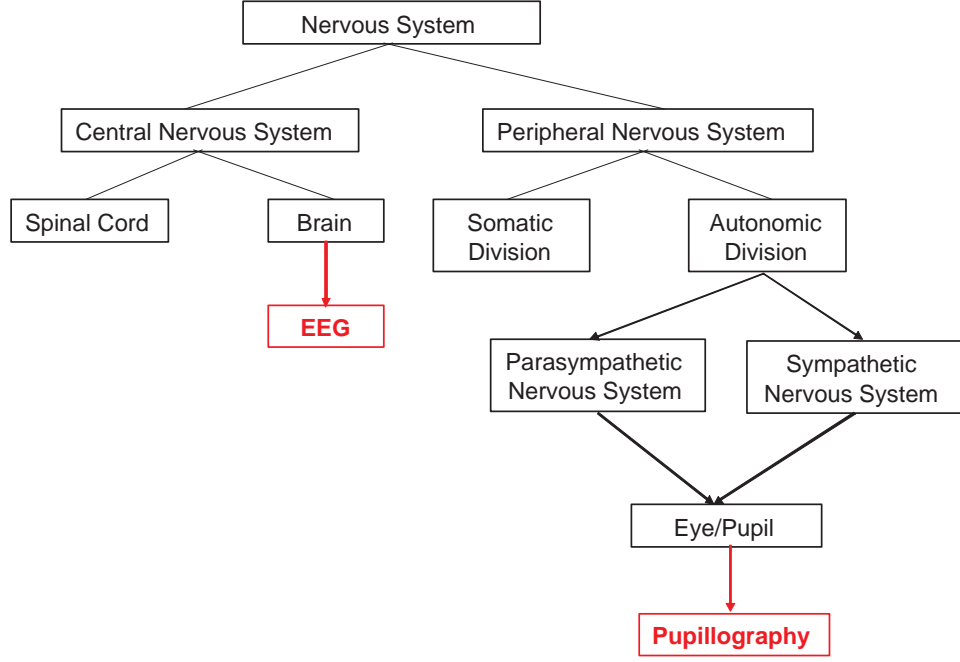
provides us a direct indicator of emotional arousal or alertness triggered by the visual detection sensory stimulus.

Since the brain activities measured by the EEG signals are part of the central nervous system activity while the pupil size is controlled by the peripheral nervous system activity (Fig. 5.2), these signals provide significantly cognitive complementarities which may serve to significantly improve classification performance. The fusion results show significant improvement over classification performance using either of the single modality (EEG or pupillary).

## 5.2 Methods

### 5.2.1 Pupil dilation feature construction

Unlike the EEG system, which has multiple channels around the scalp, the pupillary response recording only gives us one source of information. To characterize the pupillary

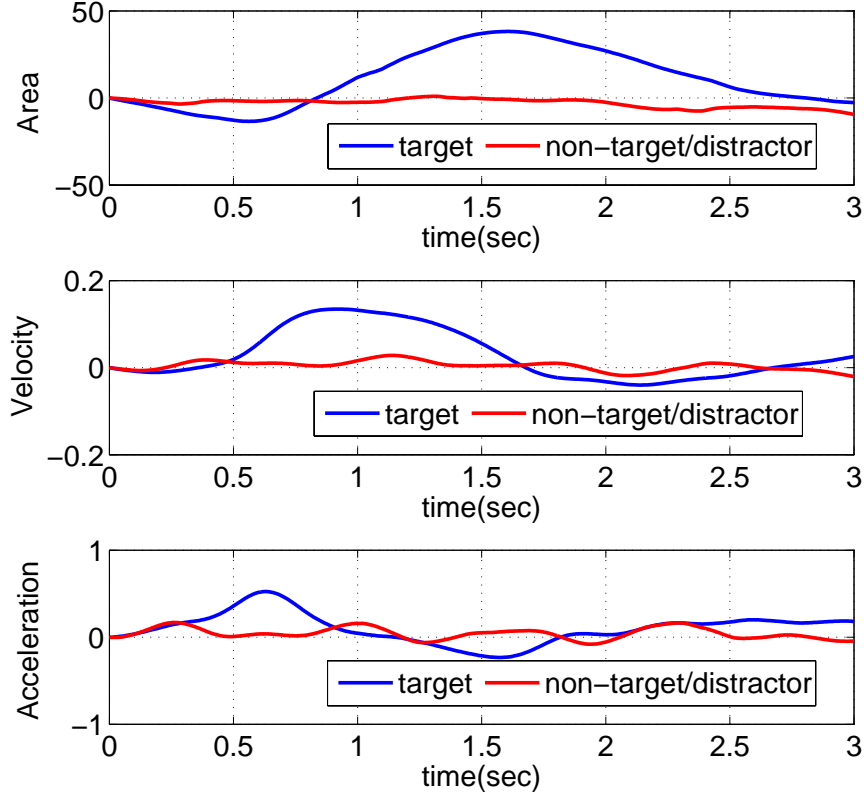


**Figure 5.2:** Nervous system organization: the brain activities measured by the EEG signals are part of the central nervous system activity while the pupil size is controlled by the peripheral nervous system activity.

response more thoroughly and improve the signal-to-noise (SNR) ratio, we identify a set of features related to the pupillary response and derive them from the original pupil diameter data. For example, the average velocity and acceleration of the pupil size change of multiple subjects are illustrated in Fig. 5.3. While the significant pupil area difference between average target response and non-target/distractor response starts to show one second after stimulus onset and peaks at around 1.5 seconds after stimulus onset, the significant velocity and acceleration differences start to reveal themselves earlier after the stimulus onset.

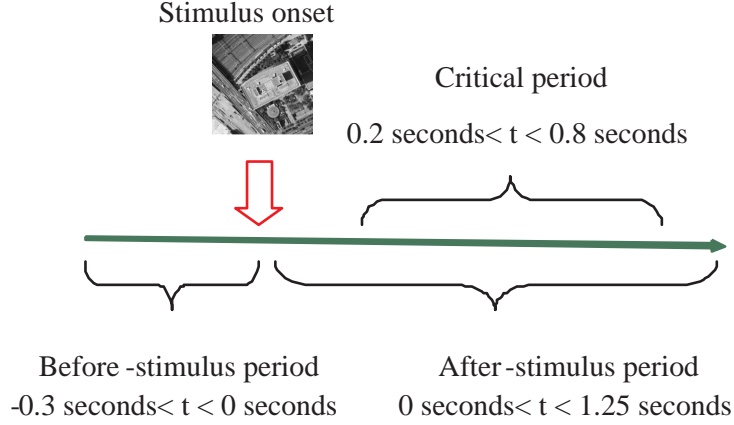
To extract pupil dilation features, the pupil diameter  $D(n)$ , pupil diameter rate of change  $V(n)$ , acceleration  $A(n)$  and pupil area  $AR(n)$  are derived for all time samples inside the time window where  $n = \{-6, \dots, -1, 0, 1, \dots, 25\}$ . The time period of  $-0.3$  seconds  $< t < 0$  seconds is defined as the before-stimulus period  $T_{before}$ . The time period of  $0$  seconds  $< t < 1.25$  seconds is defined as the after-stimulus period  $T_{after}$ . In addition, a smaller time interval between  $0.2$  seconds and  $0.8$  seconds after stimulus onset ( $0.2$  seconds

$< t < 0.8$  seconds) is regarded as the most likely time period during which the pupillary response is most salient (we call this time period “critical period”  $T_{critical}$ ) in terms of the peak velocity and peak acceleration. With 20Hz sampling frequency, the before-stimulus window contains six samples; the after-stimulus window contains 25 samples; the critical time period contains 12 samples.



**Figure 5.3:** Comparison of the average pupil area/velocity/acceleration time course after non-target/distractor image display and after target image display for all subjects. The velocity feature shows earlier differentiations between target and non-target/distractor responses compared to the area feature. The acceleration feature shows even earlier differentiations between target and non-target/distractor responses compared to the area and velocity feature. We achieve more thorough characterization of the pupil responses by deriving velocity and acceleration features.

We identified a total of 22 features that could be derived from raw pupil dilation data (listed in Table 5.1 in this chapter and the mathematics details are described in the ap-



**Figure 5.4:** Time windows defined for pupil dilation feature extraction: The time period of  $-0.3 \text{ seconds} < t < 0 \text{ seconds}$  is defined as the before-stimulus period  $T_{before}$ . The time period of  $0 \text{ seconds} < t < 1.25 \text{ seconds}$  is defined as the after-stimulus period  $T_{after}$ . A smaller time interval between 0.2 seconds and 0.8 seconds after stimulus onset ( $0.2 \text{ seconds} < t < 0.8 \text{ seconds}$ ) is regarded as the most likely time period during which the pupillary response is most salient (we call this time period “critical period”  $T_{critical}$ ) in terms of the peak velocity and peak acceleration.

pendix). Some of them are general features such as maximum dilation velocity while the others was studied in [66].

### 5.2.2 Pupil feature selection

Sequential forward floating selection (SFFS) method introduced in [67] is used for feature selection. Given the 22 candidate features described in previous section, a subset of  $d$  features,  $d < 22$ , is selected that performs the best under the specific classifier we choose to use (e.g. a linear discriminant classifier). The benefits of feature selection procedure are two-fold: it reduces the computational cost of classification by reducing the number of features that need to be calculated, and it improves classification accuracy by fitting a

simpler model based on finite number of training samples.

Sequential forward selection (SFS) method starts from an empty feature subset and sequentially selects one most significant feature at a time and adds it to the feature subset to maximize the cost function  $J$  until a predefined feature number is obtained ([68]). The classification error over a training set is used as the cost function  $J$ . Sequential backward selection (SBS) starts from a subset with all  $d$  features and selectively deletes one least significant feature at a time until a predefined feature number is obtained ([69]). Both SFS and SBS methods have the so-called nesting problem: once a feature is added/deleted, it cannot be deleted/added anymore.

SFFS method avoids nesting problem by correcting earlier ‘mistakes’ by backtracking: first enlarge the feature subset by adding  $l$  most significant features using SFS, then delete  $r$  least significant features using SBS. And the  $l$  and  $r$  are determined dynamically (“floating”) so as to approximate the optimal solution ([67]).

We choose to use SFFS for feature selection. Only the selected features are used for further pupil-based classification.

### 5.2.3 EEG-based classification

#### Linear discriminant analysis classifier for EEG

The high density EEG system simultaneously record surface potentials in multiple electrodes encompassing most of the head surface. The spatial diversity can be explored by integrating information over space to improve resolution and signal-to-noise ratio (SNR).

By linearly combining EEG channels, an aggregate representation of the data can be obtained ([31]). Let  $\mathbf{x}(t)$  be the observed vector of multidimensional EEG amplitude reading (after signal preprocessing) at time  $t$ , an optimal projection weighting vector  $\mathbf{w}_{\text{amp}}$  can be derived based on a training set and so that a one-dimensional projection  $y_{\text{amp}}(t)$  can be

derived:

$$y_{amp}(t) = \mathbf{w}_{amp}^T \mathbf{x}(t) = \sum_{i=1}^N w_{amp_i} x_i(t) \quad (5.1)$$

where  $N$  is the number of EEG electrodes (channels). For the visual detection task, we can find an optimal projection weight matrix  $\mathbf{w}_{amp}$ , which can generate a  $y_{amp}(t)$  that maximally discriminates at time  $t$  corresponding to target and non-target/distractor trial condition. It had been found that the linear discriminators trained on certain crucial short temporal windows have robust single trial discrimination performance when these discriminators were applied to analyze the spatio-temporal changes in neural processing ([31]) during visual cognitive task.

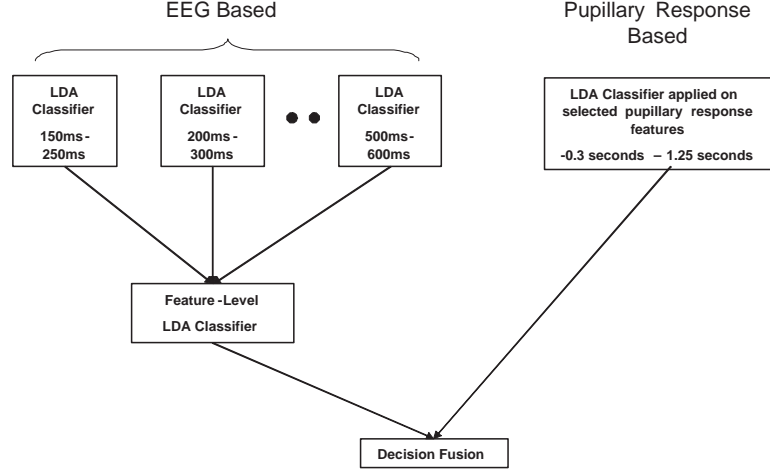
A more robust output can be derived by averaging over the  $T$  samples in the short temporal window  $\bar{y}_{amp} = T^{-1} \sum_{t=1}^T y_{amp}(t)$ . The hypotheses are  $\mathbf{H}_0$  (non-target/distractor in the image) and  $\mathbf{H}_1$  (target in the image). The posterior likelihood that a trial belongs to the target class is assumed to follow a logistic function distribution  $p(\mathbf{H}_1|\bar{y}_{amp}) = \frac{1}{1+e^{-\bar{y}_{amp}}}$  ([39]).

ROC curves can be obtained using  $p(\mathbf{H}_1|\mathbf{x})$  and comparing it with a threshold  $\theta$ . Threshold  $\theta$  can take on values within the range  $[0, 1]$ . The decision rule can be  $\begin{cases} p(\mathbf{H}_1|\mathbf{x}) \geq \theta, & u_{amp} = 1 \\ p(\mathbf{H}_1|\mathbf{x}) < \theta, & u_{amp} = 0 \end{cases}$  or vice versa where  $u_{amp} = 1$  represents a classifier's decision to declare a target detection and  $u_{amp} = 0$  represents a classifier's decision to declare a non-target/distractor detection.

In this chapter, LDA classifiers are applied on eight overlapping 100-millisecond temporal windows with onset time ranging from the 150 milliseconds following the stimulus onset to 500 milliseconds following the stimulus onset in steps of 50-millisecond. The outputs of local classifiers are fused at the feature-level in a multiple-level classification scheme.

## Multiple-level classification scheme for EEG

While LDA classifiers can be applied over multiple short temporal windows after stimulus onset, the likelihood output values of these first-level LDA classifiers (called local classifier)



**Figure 5.5:** Fusion scheme: two independent classification results are derived using EEG and pupil signals; thereafter they are fused at the decision-level.

can be used as inputs to a feature-level classifier (called global classifier) as illustrated in Fig. 5.5. The multi-level LDA classifier approach had been used on EEG amplitude data in ([31] [22]) and we apply the same approach to capture the linear pattern across multiple local windows.

By linearly combining local classifier outputs, an aggregate representation of global pattern can be obtained. Let  $\mathbf{y}$  be the observed vector of local classifier output, a weighting vector  $\mathbf{w}_{\text{window}}$  can be derived based on training data to generate a one-dimensional projection  $z$ :

$$z = \mathbf{w}_{\text{window}}^T \mathbf{y} = \sum_{i=1}^M w_{\text{window}i} y_i \quad (5.2)$$

where  $M$  is the number of local classifiers. The projection  $z$  serves as an estimate of global pattern. The likelihood of a trial belonging to the target class is assumed to follow a logistic function distribution  $p(\mathbf{H}_1|z) = \frac{1}{1+e^{-z}}$ .

ROC curves can be obtained using  $p(\mathbf{H}_1|z)$  and comparing it with a threshold  $\eta$ .  $\eta$  can take on values ranged in  $[0 \ 1]$ . The decision rule can be  $\begin{cases} p(\mathbf{H}_1|z) \geq \eta, & u = 1 \\ p(\mathbf{H}_1|z) < \eta, & u = 0 \end{cases}$  or vice versa where  $u = 1$  represents a classifier's decision to declare a target and  $u = 0$  represents

a classifier's decision to declare a non-target/distractor.

### 5.2.4 Linear discriminant analysis for pupillary response features

By linearly combining multiple pupillary-based features, an aggregate representation of the data can be obtained ([31]). Let  $\mathbf{d}$  be the observed vector of selected feature, an optimal projection weighting vector  $\mathbf{w}_{\text{pupil}}$  can be derived based on a training set and so that a one-dimensional projection  $y_{\text{pupil}}$  can be generated:

$$y_{\text{pupil}}(t) = \mathbf{w}_{\text{pupil}}^T \mathbf{d} = \sum_{i=1}^D w_{\text{pupil}_i} d_i. \quad (5.3)$$

where  $D$  is the number of pupillary features selected using the SFFS method. The projection  $y_{\text{pupil}}(t)$  can be assumed to follow some distributions of the exponential family and is regarded as a better estimate of neurophysiologic activity than any individual pupillary response feature.

ROC curves can be obtained using  $p(\mathbf{H}_1|\mathbf{d})$  and comparing it with a threshold  $\theta_{\text{pupil}}$ .

$\theta_{\text{pupil}}$  can take on values within the range  $[0, 1]$ . The decision rule can be  $\begin{cases} p(\mathbf{H}_1|\mathbf{d}) \geq \theta_{\text{pupil}}, & u_{\text{pupil}} = 1 \\ p(\mathbf{H}_1|\mathbf{d}) < \theta_{\text{pupil}}, & u_{\text{pupil}} = 0 \end{cases}$

or vice versa where  $u_{\text{pupil}} = 1$  represents a classifier's decision to declare a target and  $u_{\text{pupil}} = 0$  represents a classifier's decision to declare a non-target/distractor.

### 5.2.5 Decision fusion

While classification methods (e.g. linear discrimination analysis) can be applied on EEG and pupillary response separately and achieve significant classification performance, decision level fusion can be used to improve the overall classification performance further by combining EEG and pupil-response decisions.

Here, the goal is to enhance classification confidence through the use of multiple modalities. In addition, the signal-to-noise ratio can be improved by combining two complemen-



tary modalities. For example, while the EEG signal can be obscured by muscle movement artifact, pupillary modality is more robust to this effect. On the other hand, while pupillary signal is lost during an eye blink, EEG signal is still available even though it might be slightly corrupted by the eye blink artifact.

The decision-level fusion optimally fuses two decisions according to the operating points on their ROC curves at which each of the decisions was made with certain probability of detection ( $P_d$ ) and probability of false alarm ( $P_f$ ). The EEG amplitude observation vector has joint probability density function  $P(\mathbf{x}|\mathbf{H}_j) = P(x_1, \dots, x_N|\mathbf{H}_j)$  under hypothesis  $\mathbf{H}_j$ , for  $j = 0$  (non-target/distractor), 1 (target) and  $N = 64$ . The pupillary response feature vector (assuming the feature vector length is  $D$  after the feature selection) has joint probability density function  $P(\mathbf{d}|\mathbf{H}_j) = P(d_1, \dots, d_D|\mathbf{H}_j)$  under hypothesis  $\mathbf{H}_j$ , for  $j = 0, 1$  and  $D < 22$ . Individual detectors (EEG-based and pupil-based) receive as inputs the  $N$ -dimension EEG observation vector  $\mathbf{x}$  and the  $D$ -dimension pupillary feature observation vector  $\mathbf{D}$ , and output the decisions using the LDA classifiers. The decisions drawn by individual classifiers are denoted as  $u_k$ , for  $k = 1$  (EEG-based) and  $k = 2$  (pupillary-based), where  $u_k = 0$  if  $k_{th}$  classifier's decision is  $\mathbf{H}_0$  and  $u_k = 1$  if  $k_{th}$  classifier's decision is  $\mathbf{H}_1$ . Individual classifier's decision  $u_k$  depends only on the observation vectors  $\mathbf{x}$  and  $\mathbf{d}$ .

$$u_1 = \alpha(\mathbf{x}) = \begin{cases} 0, & \text{EEG-based classifier decides } \mathbf{H}_0, \\ 1, & \text{EEG-based classifier decides } \mathbf{H}_1. \end{cases} \quad (5.4)$$

$$u_2 = \beta(\mathbf{d}) = \begin{cases} 0, & \text{pupillary response based classifier decides } \mathbf{H}_0, \\ 1, & \text{pupillary response classifier decides } \mathbf{H}_1. \end{cases} \quad (5.5)$$

The performance characteristics of  $k_{th}$  classifier can be specified by  $P(u_k|\mathbf{H}_j)$ , where  $P(u_k=1|\mathbf{H}_0) = P_{f_k}$  = the probability of false alarm (false positives) and  $P(u_k=1|\mathbf{H}_1) = P_{dk}$  = the probability of detection (true positives). Using these probabilities, the likelihood

ratio value of a binary decision variable has a simple form as

$$\lambda_{decision}(u_k) = \frac{Pr(u_k|\mathbf{H}_1)}{Pr(u_k|\mathbf{H}_0)} = \begin{cases} \frac{P_{dk}}{P_{fk}} & \text{if } u_k = 1 \\ \frac{(1-P_{dk})}{(1-P_{fk})} & \text{if } u_k = 0 \end{cases} \quad (5.6)$$

The decision-level fusion classifier receives the decisions of the individual classifiers as its inputs. The decision at the fused level,

$$u = \varphi(u_1, u_2) = \begin{cases} 0, & \text{global decision } \mathbf{H}_0, \\ 1, & \text{global decision } \mathbf{H}_1. \end{cases} \quad (5.7)$$

depends only on local decisions and their probability of detection  $P_{dk}$  and probability of false alarm  $P_{fk}$ . Since EEG and pupillary based classifiers draw their decisions on different modalities, the simplest assumption is that these two decisions are statistically independent.

Assuming the decisions are statistically independent, the fused likelihood ratio  $\lambda$  is

$$\lambda_{fusion}(u_1, u_2) = \prod_{k=1}^2 \lambda(u_k) = \prod_{k=1}^2 \frac{P(u_k|\mathbf{H}_1)}{P(u_k|\mathbf{H}_0)} = \prod_{k=1}^2 \left(\frac{P_{dk}}{P_{fk}}\right)^{u_k} \prod_{k=1}^2 \left(\frac{1-P_{dk}}{1-P_{fk}}\right)^{1-u_k} \quad (5.8)$$

The optimal fusion rule uses the fusion likelihood ratio as the classification decision variable and compares it to threshold  $\beta$  for decision  $u$  ([44]):

$$u = F(u_1, u_2) = \begin{cases} 1, & \text{if } \frac{P(u_1, u_2|\mathbf{H}_1)}{P(u_1, u_2|\mathbf{H}_0)} \geq \beta \\ 0, & \text{if } \frac{P(u_1, u_2|\mathbf{H}_1)}{P(u_1, u_2|\mathbf{H}_0)} < \beta \end{cases} \quad (5.9)$$

The chosen threshold  $\beta$  determines the operating point on the ROC curve. By varying the threshold  $\beta$  value, the entire decision fusion ROC curve can be obtained by deriving a series of  $Pd_{fusion}(\beta)$  and  $Pf_{fusion}(\beta)$ . The decision fusion classifier probability of detection and false alarm can be calculated based on equation (5.10) and (5.11):

$$Pd_{fusion}(\beta) = \sum_{\lambda_{fusion} \geq \beta} P(\lambda = \lambda_{fusion} | H_1) \quad (5.10)$$

$$Pf_{fusion}(\beta) = \sum_{\lambda_{fusion} \geq \beta} P(\lambda = \lambda_{fusion} | H_0) \quad (5.11)$$

In practice, to derive the realistic fused ROC curve, the decisions made by the EEG-based classifier and the pupil-based classifier should be compared with the ground truth so that  $P(u_1, u_2 | \mathbf{H}_0)$ ,  $P(u_1, u_2 | \mathbf{H}_1)$  and  $\lambda_{fusion}(u_1, u_2)$  can be evaluated based on the real classifier decisions based on individual modality and compare the classifier decision vectors with the actual ground truth. Using a training set, we can obtain ROC curves and choose the operating points associated with the specific modality based classifier thresholds  $\theta_k$ , and the specific global decision threshold  $\beta$ .

The process of generating a fused ROC curve by fusing two individual ROC curves was illustrated in section 3.2.3.

## 5.3 Results and discussion

### 5.3.1 Training and test runs

The same training set of stimulus images was used for five subjects. The number of target and non-target/distractor trials in training set for five subjects are listed in Table 5.2. The training set was used for training the EEG-based local LDA classifiers located at different local time windows, the EEG-based feature-level LDA classifier, the pupil-based LDA classifier, the feature selection procedure and the decision fusion classifier.

Classification performance was evaluated based on two test sets (one set has larger target image density per sequence than the other set). Each subject was tested on both test sets. The number of target and non-target/distractor images for each subject are listed in Table 5.3.

### 5.3.2 Fusion results

The EEG-based, the pupil-based and the decision-level fusion classifier performance was evaluated using the area under the ROC curve as the metric. The results for subjects 1 to 5 over two test sets are shown in Fig. 5.6.

For the pupillary response based classification, the classification performance on test sets with lower target density per sequence are consistently better than the classification performance on test sets with higher target density per sequence across all five subjects. This observation is consistent with the stimulus probability effect: a less frequently presented stimulus elicits a stronger pupillary response while a more frequently presented stimulus elicits a weaker pupillary response and was also reported by Privitera et al. ([65]).

The results shown in the figures are analyzed for significance using a T-test method. The results show that fusion results are significantly better than EEG-based results alone ( $p < 0.003$ ) and pupil-response based results alone ( $p < 0.001$ ). Significant classification performance improvements can be achieved by fusing the EEG-based classification and the pupillary response classification results at the decision level.

## 5.4 Summary

Cognitive task related pupillary response provides a complementary modality that we can use as a basis for single-trial image triage tasks. In this chapter, we developed a pupillary response feature construction and selection procedure to extract features that perform best under a specific classifier (linear discriminant analysis classifier). At the same time, we apply a two-level linear analysis classification scheme on cognitive task-related EEG response. The classification results based on both modalities are further fused at the decision level. Here, the goal was to support increased classification confidence through inherent sensor complementarities. The fusion results show significant improvement over classification performance using a single modality.

**Table 5.1:** Pupillary features used for classification.

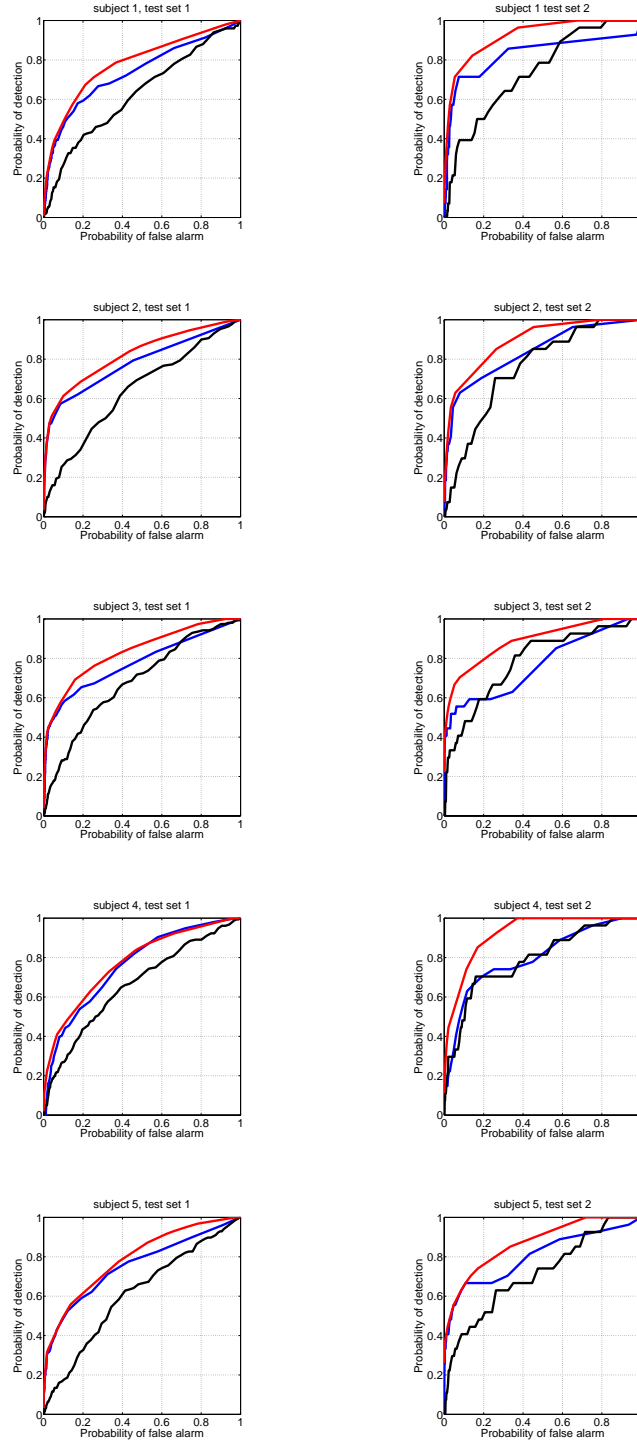
Features	Feature Description
1	maximum accumulated dilation velocity in the $T_{after}$
2	maximum accumulated dilation velocity change in the $T_{before}$ and the $T_{after}$
3	maximum dilation velocity in $T_{after}$
4	maximum dilation acceleration in $T_{after}$
5	maximum velocity in the $T_{after}$ minus mean velocity in the $T_{before}$
6	maximum velocity in the $T_{after}$ minus maximum velocity in the $T_{before}$
7	maximum pupil area after 0.2s following stimulus onset minus mean pupil area before 0.2s following stimulus onset
8	maximum pupil velocity in $T_{critical}$
9	product of features 7 (measure of area change) and feature 8(measure of velocity change)
10	maximum diameter in the $T_{after}$
11	maximum diameter in the $T_{after}$ minus mean diameter in the $T_{before}$
12	absolute difference between maximum normalized (4-sample) diameter moving averages in the $T_{after}$ and $T_{before}$
13	difference between maximum normalized (5-sample) diameter deviations from moving average in the $T_{after}$ and $T_{before}$
14	difference between maximum normalized (5-sample) diameter slope deviations from moving average in the $T_{after}$ and $T_{before}$
15	difference between maximum normalized (5-sample) diameter gradient from moving average in the $T_{after}$ and $T_{before}$
16	difference between maximum normalized absolute value slopes (using a 4-sample interval) in the $T_{after}$ and $T_{before}$
17	difference between maximum normalized diameter gradient moving sums in the $T_{after}$ and $T_{before}$
18	difference between the percentage of samples in the $T_{after}$ and $T_{before}$ that have larger diameter than the diameter at the stimulus onset time
19	difference between the percentage of samples in the $T_{after}$ and in the $T_{before}$ that have diameter values smaller than the diameter at the stimulus onset
20	average energy in the $T_{after}$ and $T_{before}$ calculated as the mean square value for histogram-based probability density function of normalized diameter
21	power ratio of the peak samples in a window centered around the stimulus onset versus all the samples in the window
22	difference between contour energy in the $T_{after}$ and $T_{before}$

**Table 5.2:** Number of target and non-target/distractor images in the training set.

Non-target/distractor images	Target images
2415	256

**Table 5.3:** Number of target and non-target/distractor images, target versus non-target/distractor ratios in test sets for five subjects.

Subject	Test set	Non-target/distractor images	Target images	Target versus non-target/distractor ratio
1	1	3341	150	1:22
	2	3464	27	1:128
2	1	3341	150	1:22
	2	3437	27	1:127
3	1	3341	150	1:22
	2	3437	27	1:127
4	1	3277	156	1:21
	2	3464	27	1:128
5	1	3121	156	1:20
	2	3374	27	1:125



**Figure 5.6:** Five subjects' EEG-based, pupillary response-based, and fused results on test set one and two; Blue curves are the ROC curves for the EEG-based results; Black curves are the ROC curves for the pupil-response based results; Red curves are the ROC curves for the fused results.

# Chapter 6

## Multi-Subject Decision Fusion

### 6.1 Motivation

For crucial image triage tasks, multiple image analysts could be asked to look at the same set of images to improve the probability of detection and reduce the probability of false positives. From the viewpoint of a cortically-coupled computer vision system, multiple subjects can be regarded as multiple sensors.

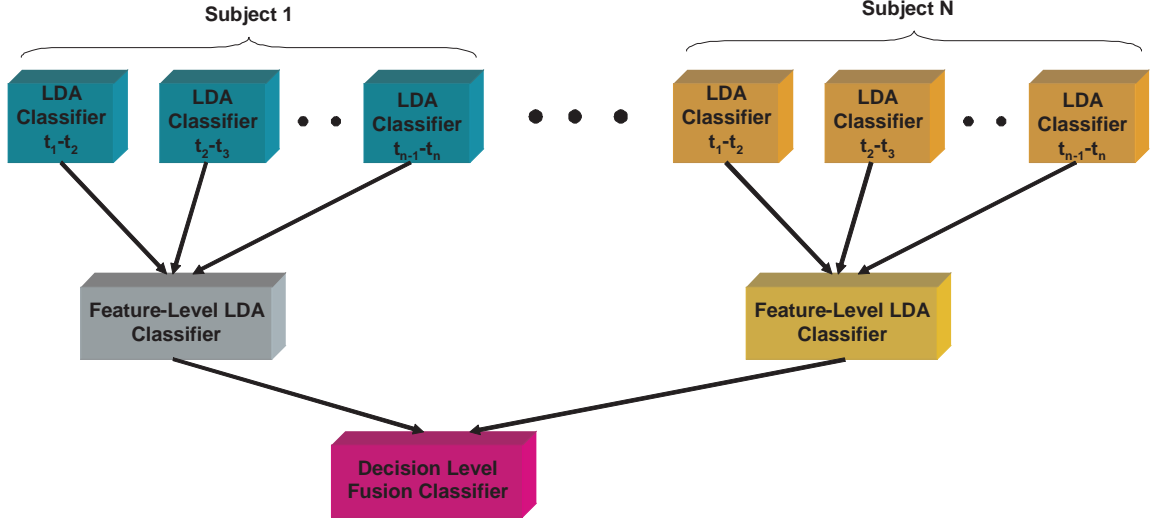
Every individual analyst differs in visual detection capabilities such as acuity and contrast sensitivity over ranges of target size and target-background contrast. And each individual might have a different level of familiarity with certain image context background. As a result, there is complementariness among decisions drawn by different analysts, and overall performance can be improved by fusing the individual decisions. In this chapter we evaluate the benefit of using multiple analysts on the same test data sets.

### 6.2 Methods

For each subject, we apply the feature-level fusion method we described in chapter three using an LDA feature-level classifier. The feature-level LDA classifier approach was applied on multiple local LDA classifier outputs ranging from 150 to 600 milliseconds in steps of 50 milliseconds. This approach fuses multiple local LDA classifier outputs in an attempt to capture the linear spatio-temporal pattern across multiple local windows.

The feature-level LDA classifier results for multiple subjects are fused at the decision-level. Since more training data are required to fuse more subjects, we only tried to fuse decisions from two or three subjects in our study. The process of generating a fused ROC curve by fusing individual ROC curves was illustrated in section 3.2.3.





**Figure 6.1:** Decision fusion scheme to combine decisions drawn by multiple subjects.

**Table 6.1:** Number of target and non-target/distractor images in the training set.

Non-target/distractor images	Target images
2415	256

## 6.3 Results and discussion

### 6.3.1 Training and test runs

The same training set of stimulus images was used for six subjects. The number of target and non-target/distractor trials in the training set for six subjects are listed in Table 6.1. The ratio of target versus non-target/distractor trials is about 1:10 (while this ratio can be smaller than 1:100 in real applications, we have to have enough target trials to train the classifiers). The training set was used to derive weights for local LDA classifiers at different time windows, weights for the feature-level LDA classifier, and to decide the operating points for the global decision fusion classifier.

The classification performance is evaluated based on four test sets. The numbers of target and non-target/distractor images of test sets are listed in Table 6.2. The ratios of target versus non-target/distractor images differ in different test sets so that we can evaluate the classification performance for different target density profiles.

**Table 6.2:** Number of target and non-target/distractor images in two test sets.

Test Set	Non-target/distractor images	Target images	Subjects executed the test set
1	3121	150	1, 2, 3
2	3437	27	1, 2, 3
3	3121	156	4, 5, 6
4	3374	27	4, 5, 6

**Table 6.3:** fusion gains for test sets by combining decisions from different subjects.

Test Set	Subject 1 and 2	Subject 1 and 3	Subject 2 and 3	subject 1, 2 and 3
1	6.21%	10.08%	7.87%	11.01%
2	13.41%	13.55%	8.68%	14%
Test Set	Subject 4 and 5	Subject 4 and 6	Subject 5 and 6	Subject 4, 5 and 6
3	1.69%	3.54%	3.17%	4.42%
4	6.46%	9.58%	6.41%	7.83%

### 6.3.2 Decision-level fusion classification performance

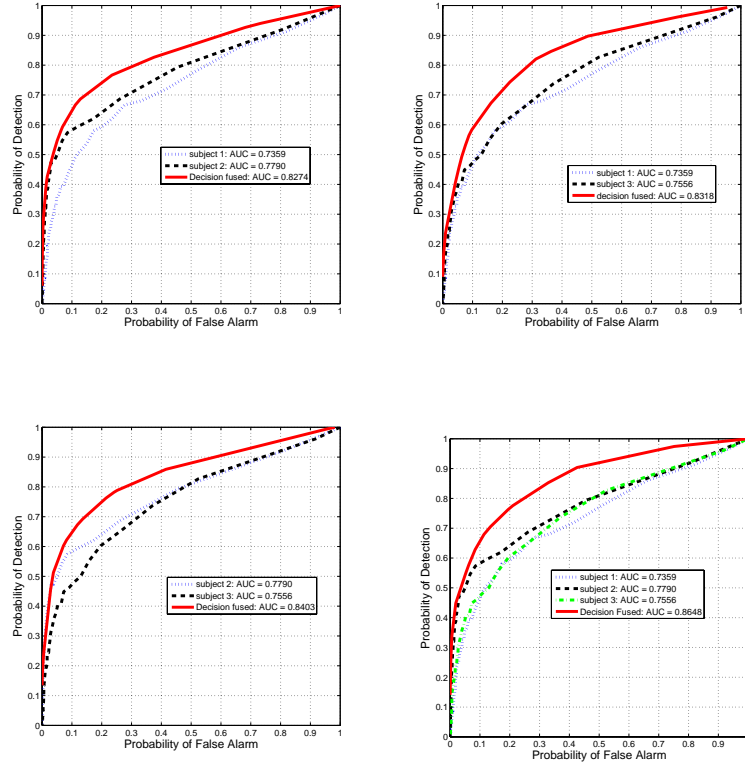
For test set 1, the results of fusing decisions for two of the three subjects (1 and 2, 2 and 3, 1 and 3) and for all three subjects (1, 2 and 3) are shown on Fig. 6.2.

For test set 2, the results of fusing decisions for two of the three subjects (1 and 2, 2 and 3, 1 and 3) and for all three subjects (1, 2 and 3) are shown on Fig. 6.3.

For test set 3, the results of fusing decisions for two of the three subjects (4 and 5, 5 and 6, 4 and 6) and for all three subjects (4, 5 and 6) are shown on Fig. 6.4.

For test set 4, the results of fusing decisions for two of the three subjects (4 and 5, 5 and 6, 4 and 6) and for all three subjects (4, 5 and 6) are shown on Fig. 6.5.

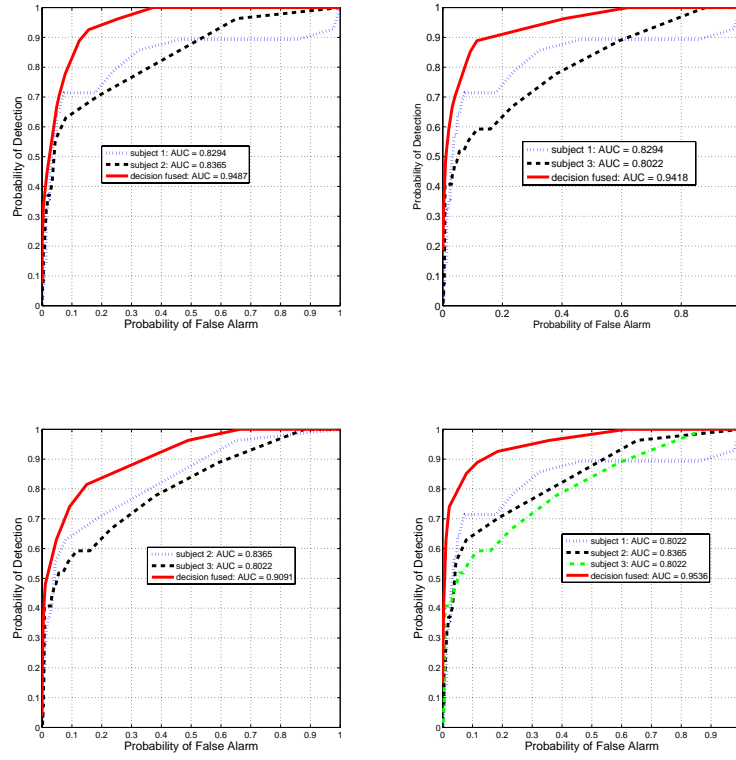
The multiple subject fusion gains are measured using percentage improvement of the area under the ROC curve for the fused decision versus the area under the ROC curve for the best single subject. The improvements are listed in Table 6.3. We can observe that for all test sets, there are significant fusion gains by combining the decisions made by multiple subjects.



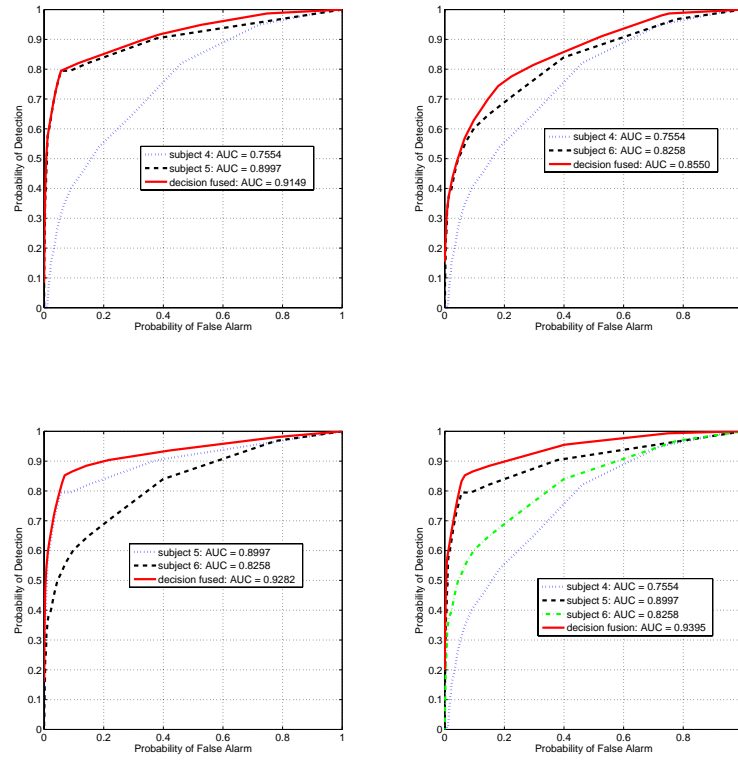
**Figure 6.2:** Decision fusion performance using subject 1 and 2, subject 1 and 3, subject 2 and 3 and using all three subjects for test set 1.

## 6.4 Summary

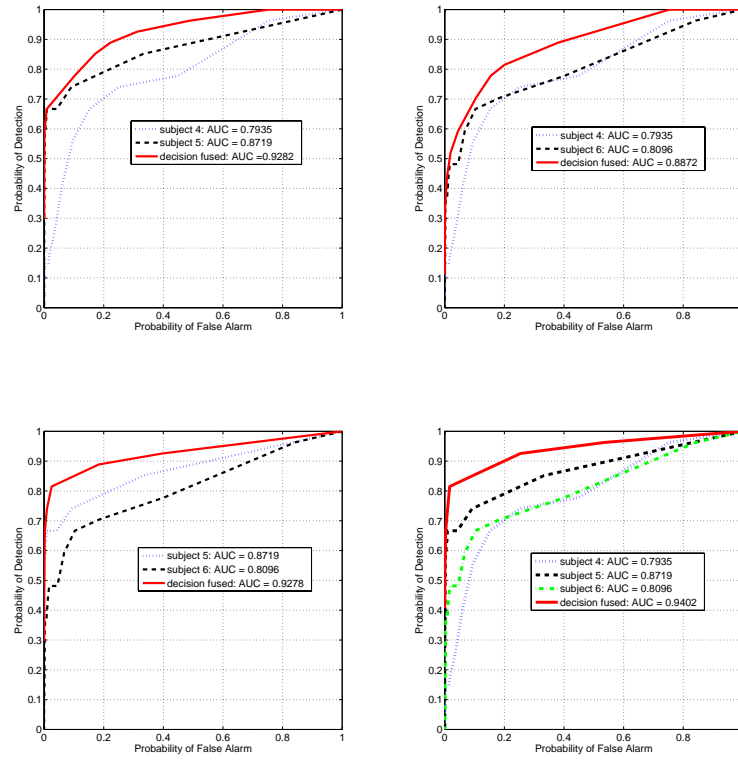
This chapter presents the method of fusing the decisions made by multiple subjects on the same set of imagery. The fusion is executed at the decision level and we can observe significant fusion gains for all four test sets using two-subject and three-subject decision fusion.



**Figure 6.3:** Decision fusion performance using subject 1 and 2, subject 1 and 3, subject 2 and 3 and using all three subjects for test set 2.



**Figure 6.4:** Decision fusion performance using subject 4 and 5, subject 4 and 6, subject 5 and 6 and using all three subjects for test set 3



**Figure 6.5:** Decision fusion performance using subject 4 and 5, subject 4 and 6, subject 5 and 6 and using all three subjects for test set 4

# Chapter 7

## Classification Using Reduced Number of EEG Sensors

### 7.1 Motivation

Throughout the previous chapters, we relied on a 64-channel EEG system to obtain all the experiment data and all 64 channels were used for data analysis (for example, the LDA classifiers used all 64 raw EEG-channel data). For the purpose of developing a practical real-time EEG-based application system, sometimes we have to work with an EEG system that has fewer electrodes (channels). The limitations are due to the limited communication capability and processing power of the mobile system. For example, g-tec medical engineering GmbH (<http://www.gtec.at>) developed a portable EEG biosignal research system that makes mobile biosignal acquisition and analysis possible. The mobile system only supports 6-8 EEG channels. So a crucial question is whether the classification techniques developed in the previous chapters based on 64-channel data can achieve robust performance using a fraction of the channels.

In this chapter, we evaluate two methods of identifying the dominant small set of channels that can deliver relatively robust classification performance. One method is based on the sequential forward floating selection (SFFS) method introduced in section 5.2.2, where 64 channels are treated as the candidate features. We selected the limited number of channels that performs the best under the specific classifier we choose to use (e.g. an LDA). The other method is a method that used the top channels that had the maximum absolute weights for the LDA classifier to perform the classification.

## 7.2 Methods

### 7.2.1 Finding the inherent dimensionality of the EEG-based detection using principal component analysis

The principal component analysis (PCA) transforms a number of correlated variables into a smaller number of uncorrelated variables (principal components) using eigen analysis (also known as ‘Singular Value Decomposition’). The first principal component accounts for as much of the variability in the data as possible in the direction of the eigenvector associated with the largest eigenvalue.

PCA can be used to discover the dimensionality of the data set. In this chapter, we apply the PCA on data matrix representing the difference between the average target-trial and the average non-target/distractor trial response. The matrix is 64 channels (rows) long and each channel has multiple data samples (columns). The objective is to analyze the dimensionality of the EEG-based detection problem.

### 7.2.2 Ranking the EEG channels using a single temporal window based LDA classifier

For the single temporal window based LDA classifier we described in section 3.2.1, we can linearly combine EEG channels. Let  $\mathbf{x}(t)$  be the observed vector of multidimensional EEG amplitude reading (after signal preprocessing) at time  $t$ . An optimal projection weighting vector  $\mathbf{w}_{\text{amp}}$  can be derived based on a training set and so that a one-dimensional projection  $y_{\text{amp}}(t)$  can be derived using:

$$y_{\text{amp}}(t) = \mathbf{w}_{\text{amp}}^T \mathbf{x}(t) = \sum_{i=1}^N w_{\text{amp}_i} x_i(t) \quad (7.1)$$

where  $N$  is the number of EEG electrodes (channels). For the visual detection task, we can find an optimal projection weight matrix  $\mathbf{w}_{\text{amp}}$ , which can generate a  $y_{\text{amp}}(t)$  that



maximally discriminates at time  $t$  corresponding to target and non-target/distractor trial condition.

If we assume the number of EEG channels is 64, now the problem is that we have to identify a reduced set of channels (e.g. eight channels) that still can have robust classification performance. One approach is to adapt the method of ranking the channels based on their absolute values of projection weight value. After the 64 channels are sorted in descending order based on  $\mathbf{w}_{\text{amp}}$ , we will select the top eight if eight is the maximum number of channels we are allowed to use. The other way is to treat the 64 channels as features and use the feature selection (SFFS) algorithm to identify the top ranked channels.

### 7.2.3 Feature-level fusion of multiple local classifiers' outputs

We use the feature-level LDA fusion classifier to fuse the multiple local classifiers' outputs (as described in section 3.2.2). The key observation is that we have to use the same reduced set of channels (e.g. eight channels) across all eight local classifiers.

Using the 64-channel training data, we can derive the top eight channels for each local classifier located at different temporal intervals after stimulus image onset. For example, we might choose to use multiple local LDA classifier outputs ranging from 150 to 600 milliseconds in steps of 50 milliseconds. The temporal window length is 100 milliseconds. Altogether there are eight local classifiers located at different temporal intervals (150-250 millisecond, 200-300 millisecond, 250-350 millisecond, 300-400 millisecond, 350-450 millisecond, 400-500 millisecond, 450-550 millisecond, 500-600 millisecond). For each of the local classifiers, we can identify eight top ranked channels. They are different for different local classifiers. We use the top ranked channels from the local classifier that gives the best classification performance in terms of the Area under the ROC curve ( $A_z$ ). This reduced set of channels is then used in all eight local classifiers during both the training and test phases.

**Table 7.1:** Number of target and non-target/distractor images in the training set.

Non-target/distractor images	Target images
2415	256

**Table 7.2:** Number of target and non-target/distractor images in multiple test sets.

Test Set	Non-target/distractor images	Target images	Subjects executed the test set
1	3341	150	2, 3
2	3464	27	2, 3
3	3121	156	1, 4, 5, 6, 7
4	3374	27	1, 4, 5, 6, 7

## 7.3 Results and discussion

### 7.3.1 Training and test runs

The same training set of stimulus images was used for seven subjects. The number of target and non-target/distractor trials in the training set for seven subjects are listed in Table 7.1.

The training set was used to derive weights for local LDA classifiers at different temporal windows, to derive eight top-ranked EEG channels (using the top weight selection method or the feature selection method) and to derive weights for the feature-level LDA classifier.

The classification performance is evaluated based on multiple test sets. There are four different test sets that were used in the experiments. The number of target and non-target/distractor trials of these test sets are listed in Table 7.2. The ratios of target versus non-target/distractor trials differ in different test sets so that we can evaluate the classification performance for different target density profiles. Each subject had worked on a test set with a higher target density and a test set with a lower target density.

### 7.3.2 Principal component analysis results

For each of the seven subjects, we use the training set data (2415 non-target/distractor images and 256 target images) for PCA analysis. Using the data samples located between 150 milliseconds and 600 milliseconds after stimulus onset (54 data points given 120Hz

sampling rate), the average target trial matrix and the average non-target/distractor trial matrices are both 2-dimensional matrix of size 54 x 64 (54 sample points for each channel and 64 channels). The difference between the average target trial matrix and the average non-target/distractor trial matrix is a matrix of size 54 x 64 too. We apply the PCA on this matrix that represents the difference between target and non-target/distractor images.

The resulting eigenvalues give us the estimation of the dimensionality of this detection problem. The amplitude of the eigenvalue indicates the importance of each corresponding component in terms of the variance represented by each component. By plotting the eigenvalues on a graph (Fig. 7.1), we can observe that the points on the graph tend to level out after the first few components. Those eigenvalues after the level-out are usually very small and they can be ignored.

For each of the subjects, we can observe that the graph starts to level out after the first 5-8 components. This gives us an estimation on the dimensionality of EEG-based detection problem related to visual target trials and non-target/distractor trials.

### **7.3.3 Classification performance using reduced number of channels**

Using the two test sets for each subject, we can identify the single 100-millisecond window that gives us the best area under the ROC curve performance for each subject. For subject 1, the best performing window is located at 250-350 milliseconds after stimulus onset; for subjects 2, 4 and 6, the best performing window is located at 300-400 milliseconds after stimulus onset; for subjects 3, 5 and 7, the best performing window is located at 150-250 milliseconds after stimulus onset;

The top ranked eight channels are selected based on the best performing window. The reason is that we are using a two-level fusion scheme; there are eight local classifiers and we can identify eight channels on every individual local classifier. While we cannot optimize the classification performance of every local classifier, we can optimize the performance

for one of the local classifiers located at the ideal position so that the overall classification performance can be maintained at a satisfactory level.

For the top weight selection method, we derive the LDA weight vector using the training data set, and we choose the eight channels that have the top absolute weights. For the feature selection method, we derive the top eight channels or less using the sequential floating forward selection (SFFS) method explained in section 5.2.2. The selected eight channels (or less) are then used for training, testing and the classification performance is compared with the performance derived using a full 64-channel system.

Fig. 7.3 shows the comparison of classification performance using 64 channels versus 8 channels (using the top weight selection method and the feature selection method based on the LDA classification) for the higher target density test set in terms of the Area under the ROC curve ( $A_z$ ). Fig. 7.4 shows the comparison of classification performance using 64 channels versus eight channels (using the top weight selection method and the feature selection method based on LDA classification) for the lower target density test set in terms of the Area under the ROC curve ( $A_z$ ).

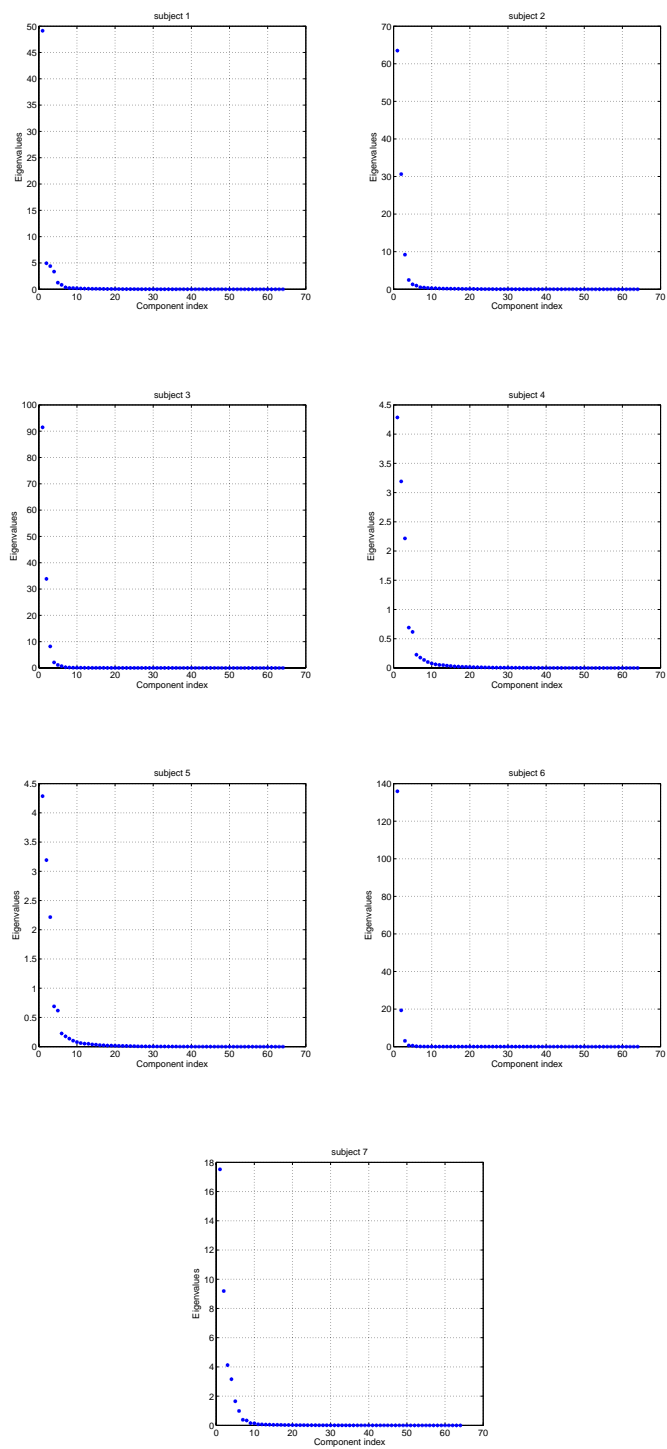
We can conclude that there is no significant performance degradation by using fewer channels (8 channels instead of 64 channels). The approaches we adapt to identify the top ranked channels are effective methods for the multi-level LDA-based classification scheme.

## 7.4 Summary

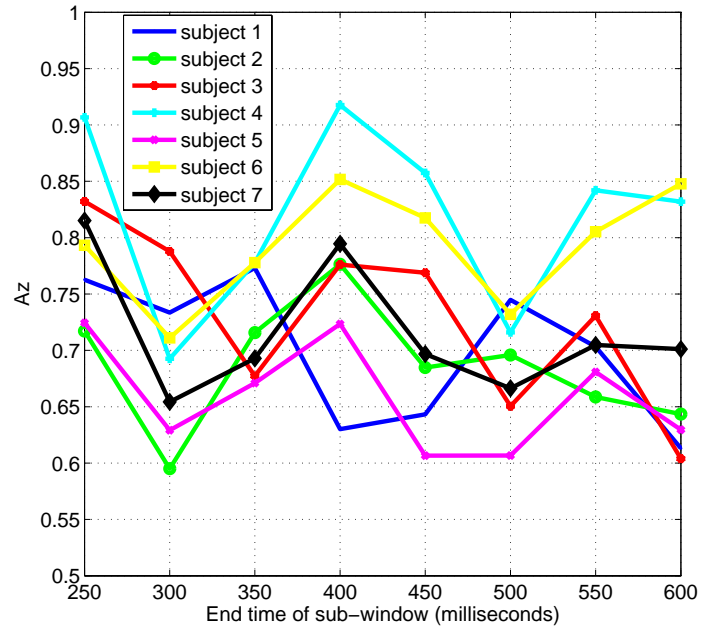
This chapter presents the method of identifying a reduced set of EEG channels that can deliver robust classification performance. We adapt both a selection method of ranking the channels based on their absolute projection weight values and a feature-selection method. By applying these methods on both a high target density and a low target density test sets using multiple subjects, we are able to show that there is no significant performance degradation by using fewer channels (8 channels instead of 64 channels).

In real application, these methods offer us an effective way to identify and position the

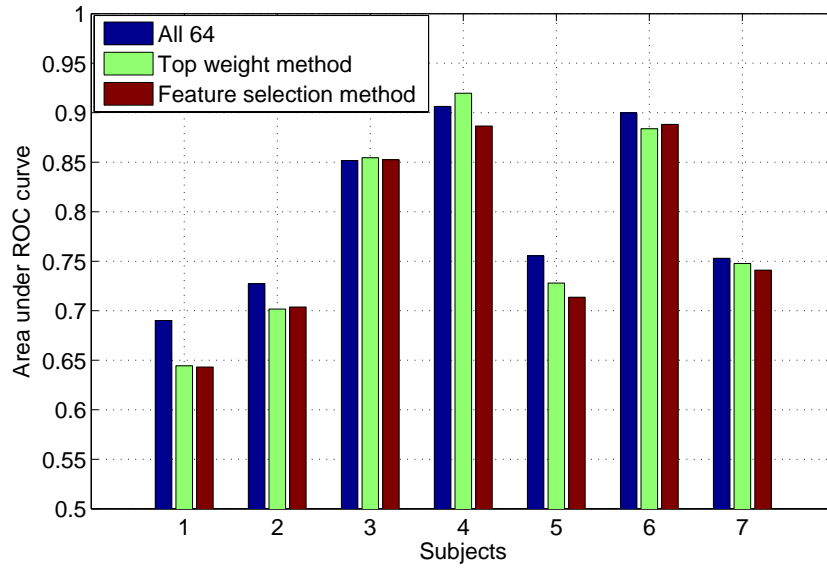
limited number of EEG electrodes on a wireless mobile EEG system.



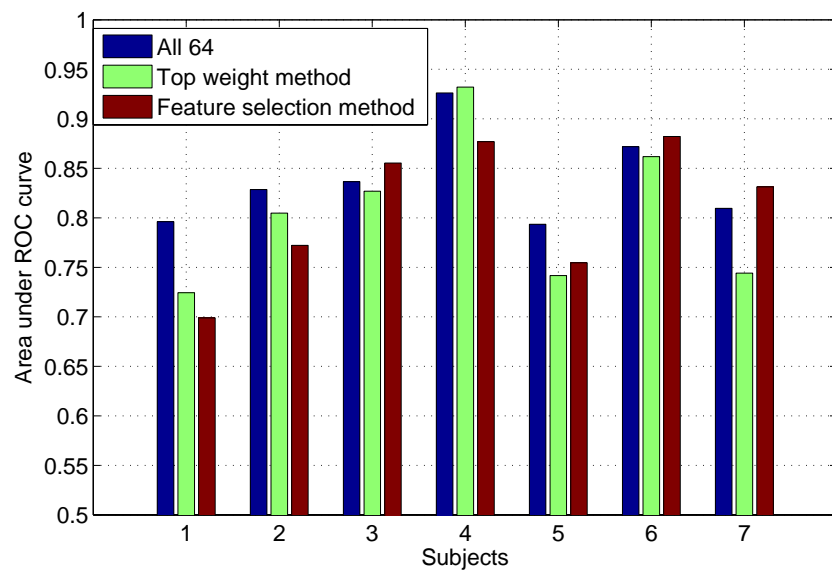
**Figure 7.1:** PCA eigenvalues for seven subjects.



**Figure 7.2:** Single window classifier performance for four subjects.



**Figure 7.3:** Comparison of classification performance using different number of channels for the higher target density test set.



**Figure 7.4:** Comparison of classification performance using different number of channels for the lower target density test set.



# Chapter 8

## Conclusions

### 8.1 Conclusions and Future Works

#### 8.1.1 Conclusions

The EEG brain-machine interface systems provide a nonmuscular communication and control method for a wide range of applications because they do not require invasive surgical implants ([70], [71] and [72]).

Conventional neuroimaging methods typically seek to determine which regions of the brain are activated (encoded) in a particular perceptual and cognitive task. The activities from tens, hundreds or thousands of locations are recorded repeatedly during the task time period and each location is analyzed separately. Those individual locations that show significant average difference under opposite cognitive states (e.g target versus non-target) can be used to determine which one of those two cognitive states reflects the current state. For example, the fusiform face area (FFA) is a region in the human ventral visual stream that responds strongly to faces than to any other object category ([73], [74]). The drawback of this method is that it is very difficult to isolate and monitor signals from those crucial locations using noninvasive (e.g. EEG) methods.

The pattern-based multivariate analysis based on the full spatial pattern of brain activity measures simultaneously at many locations so that the distributed information can be accumulated efficiently and jointly analyzed ([23], [75], [76]). The pattern-based approach also facilitates the accurate detection/classification from single-trial inputs.

In this study, we apply fusion techniques on visual target recognition applications and demonstrate the ability to improve the detection/classification accuracy further from the state-of-the-art pattern-based analysis method. In chapter three, we fuse multiple EEG-

based local classifiers' outputs, both at the feature-level and at the decision-level, to discriminate between target and non-target/distractor images, in order to capture the global neurophysiological activities located in selectively distributed sites of the brain evolving with a continuous time course. In chapter four, we apply decision-level fusion to combine the EEG-based temporal domain classification decision and the EEG-based frequency domain classification decision because they contribute different discriminating components that are established on alternative perspectives on brain cognitive dynamics. In chapter five, we developed a pupillary response feature construction and selection procedure that facilitates a pupil response based detection/classification algorithm. Further classification accuracy through inherent complementarities can be achieved by fusing the EEG-based and pupil-based classification results at the decision level.

Chapter six presents the fusion gain of combining the decisions made by multiple subjects on the same set of imagery. The reason behind the fusion is that every individual analyst differs in visual detection capabilities such as acuity, contrast sensitivity and level of familiarity with certain image context background.

Chapter seven addresses a practical aspect of applying EEG-based visual target detection method. We proposed methods of identifying a reduced set of EEG channels that can deliver robust classification performance so that we can obtain significant detection performance using a reduced set of EEG electrodes (channels).

### **8.1.2 Future works**

The future goal is to develop a fully integrated system for optimal image presentation, analysis of physiological signals, and classification of imagery into target and non-target/distractor using physiological signals (EEG and pupillary signals). The architecture can be composed of multiple input data streams such as EEG and pupil signals. The corresponding processed signals are then combined in a multi-sensor fusion module to provide a mechanism for decoding target recognition events. The techniques introduced in the previous chapters

can be incorporated into this fully integrated system.

The goal of the integrated system is to generate the classification results in real-time. Further works should be done to optimize and speed up the classification algorithms. For example, the frequency domain based classification algorithm relies on the spectral power calculation using wavelet transform that requires significant computational time. A faster solution is required for an integrated system with a real time response capability.

## APPENDIX A: Pupillary response feature extractions

$T_{before} \in \{-0.3s < t < 0s\}$  where  $t = 0$  is the stimulus onset time.  $T_{after} \in \{0s < t < 1.25s\}$  where  $t = 0$  is the stimulus onset time.  $T_{critical} \in \{0.2s < t < 0.8s\}$  where  $t = 0$  is the stimulus onset time. With 20Hz sampling frequency, the before-stimulus window contains 6 samples; the after-stimulus window contains 25 samples; the critical time period contains 12 samples.

**Table A-1:** Pupillary features used for classification.

<i>Features</i>	Feature Calculation
1	$\operatorname{argmax}_a(\sum_{i=0}^{i=a} V(t_i))$ where $a \in \{1, 2, \dots, 25\}$ for the post-stimulus period
2	$\max(\sum_{i=0}^{i=a} V(t_i)) - \max(\sum_{j=0}^{j=6} V(t_j))$ , where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$ and $b \in \{1, 2, \dots, 6\}$ for the $T_{before}$
3	$\max(V(t_a))$ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$
4	$\max(A(t_a))$ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$
5	$\max(V(t_a)) - \text{mean}(V(t_b))$ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$ and $b \in \{1, 2, \dots, 6\}$ for the $T_{before}$
6	$\max(V(t_a)) - \max(V(t_b))$ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$ and $b \in \{1, 2, \dots, 6\}$ for the $T_{before}$
7	$\max(AR(t_a)) - \text{mean}(AR(t_b))$ where $a \in \{4, 5, \dots, 25\}$ for the $T_{after}$ , $b \in \{1, 2, \dots, 6\}$ for the $T_{before}$ and $b \in \{1, 2, 3\}$ for the $T_{after}$
8	$\max(V(t_c))$ where $c \in \{1, 2, \dots, 12\}$ for the $T_{critical}$
9	product of feature 8 (measure of area change) and 9 (measure of changing velocity)
10	$\max(D(t_a))$ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$ ,
11	$\max(D(t_a)) - \text{mean}(D(t_b))$ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$ and $b \in \{1, 2, \dots, 6\}$ for the $T_{before}$
12	$ \max(\frac{\tilde{D}(a)}{\max(\tilde{D}(m))}) - \max(\frac{\tilde{D}(b)}{\max(\tilde{D}(m))}) $ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$ , $b \in \{1, 2, \dots, 6\}$ for the $T_{before}$ , $m \in \{1, 2, \dots, 31\}$ for all sample points in $T_{after} \cup T_{before}$ and $\tilde{D}(k) = \frac{1}{4} \sum_{i=k-3}^k D(i)$
13	$\max(\frac{\tilde{D}(a)-D(a)}{\max(\tilde{D}(m)-D(m))}) - \max(\frac{\tilde{D}(b)-D(b)}{\max(\tilde{D}(m)-D(m))})$ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$ , $b \in \{1, 2, \dots, 6\}$ for the $T_{before}$ , $m \in \{1, 2, \dots, 31\}$ for all sample points in $T_{after} \cup T_{before}$ and $\tilde{D}(k) = \frac{1}{5} \sum_{i=k-4}^k D(i)$
14	$\max(\frac{(\tilde{D}(a)-D(a))-(\tilde{D}(a-1)-D(a-1))}{(\tilde{D}(m)-D(m))-(\tilde{D}(m-1)-D(m-1))}) -$ $\max(\frac{(\tilde{D}(b)-D(b))-(\tilde{D}(b-1)-D(b-1))}{(\tilde{D}(m)-D(m))-(\tilde{D}(m-1)-D(m-1))})$ where $a \in \{1, 2, \dots, 25\}$ for the $T_{after}$ , $b \in \{1, 2, \dots, 6\}$ for the $T_{before}$ , $m \in \{1, 2, \dots, 31\}$ for all sample points in $T_{after} \cup T_{before}$ and $\tilde{D}(k) = \frac{1}{5} \sum_{i=k-4}^k D(i)$

**Table A-2:** Pupillary features used for classification.

15	$\max(\frac{\tilde{\nabla}D(a)}{\tilde{\nabla}D(m)}) - \max(\frac{\tilde{\nabla}D(b)}{\tilde{\nabla}D(m)})$ <p>where <math>a \in \{1, 2, \dots, 25\}</math> for the <math>T_{after}</math>, <math>b \in \{1, 2, \dots, 6\}</math> for the <math>T_{before}</math>,  <math>m \in \{1, 2, \dots, 31\}</math> for all sample points in <math>T_{before}</math> and <math>T_{after}</math>,  <math display="block">\tilde{\nabla}D(k) = \frac{1}{5} \sum_{i=k-4}^k \tilde{\nabla}D(i)</math></p>
16	$\max(\frac{\frac{\tilde{D}(a)-D(a)}{\max(\tilde{D}(m)-D(m)) - \max(\tilde{D}(t_m)-D(t_m))}}{\max( \frac{\tilde{D}(m)-D(m)}{\max(\tilde{D}(m)-D(m)) - \max(\tilde{D}(m)-D(m-3))} )} - \frac{\tilde{D}(b)-D(b)}{\max(\tilde{D}(m)-D(m)) - \max(\tilde{D}(m)-D(m-3))})$ $- \max(\frac{\frac{\tilde{D}(b)-D(b)}{\max(\tilde{D}(m)-D(m)) - \max(\tilde{D}(m)-D(m-3))}}{\max( \frac{\tilde{D}(m)-D(m)}{\max(\tilde{D}(m)-D(m)) - \max(\tilde{D}(m)-D(m-3))} )} - \frac{\tilde{D}(m-3)-D(m-3)}{\max(\tilde{D}(m)-D(m)) - \max(\tilde{D}(m)-D(m-3))})$ <p>where <math>a \in \{1, 2, \dots, 25\}</math> for the <math>T_{after}</math>, <math>b \in \{1, 2, \dots, 6\}</math> for the <math>T_{before}</math>,  <math>m \in \{1, 2, \dots, 31\}</math> for all sample points in <math>T_{after}</math>  <math>\cup T_{before}</math> and <math>\tilde{D}(k) = \frac{1}{5} \sum_{i=k-4}^k D(i)</math></p>
17	$\max(\frac{\sum_{k1=a}^{a+5} (\nabla D(k1))}{\max( \sum_{k2=m}^{m+5} (\nabla D(k2)) )}) - \max(\frac{\sum_{k3=b}^{b+5} (\nabla D(k3))}{\max( \sum_{k4=m}^{m+5} (\nabla D(k4)) )})$ <p>where <math>a \in \{1, 2, \dots, 20\}</math> for the <math>T_{after}</math>, <math>b \in \{1, 2\}</math> for the <math>T_{before}</math>,  <math>m \in \{1, 2, \dots, 26\}</math> for all sample points in <math>T_{before}</math> and <math>T_{after}</math>,  <math display="block">\tilde{D}(k) = \frac{1}{5} \sum_{i=k-4}^k D(i)</math></p>
18	$(P_1 - Q_1)/M$ <p>where <math>P_1</math> is the number of <math>D(a) &gt; D_0</math>,  <math>a \in \{1, 2, \dots, 25\}</math> for the <math>T_{after}</math>, <math>Q_1</math> is the number of <math>D(b) &gt; D_0</math>  where <math>b \in \{1, 2, \dots, 6\}</math> for the <math>T_{before}</math> and <math>M</math> is the number  of <math>D(m) &gt; D_0</math> where <math>m \in \{1, 2, \dots, 32\}</math> for all sample points  in <math>T_{before}</math> and <math>T_{after}</math>, <math>D_0</math> is the diameter size at the stimulus onset.</p>
19	$(P_2 - Q_2)/M$ <p>where <math>P_2</math> is the number of <math>D(a) &lt; D_0</math>,  <math>a \in \{1, 2, \dots, 25\}</math> for the <math>T_{after}</math>, <math>Q_2</math> is the number of <math>D(b) &lt; D_0</math>  where <math>b \in \{1, 2, \dots, 6\}</math> for the <math>T_{before}</math> and <math>M</math> is the number  of <math>D(m) &gt; D_0</math> where <math>m \in \{1, 2, \dots, 32\}</math> for all sample points  in the <math>T_{before}</math> and <math>T_{after}</math>, <math>D_0</math> is the diameter size at the stimulus onset.</p>
20	$E = \ln \frac{1}{g} \sum_{i=0}^{g-1} D(i)^2 p_i$ <p>using the probability density function (pdf)  of the normalized diameter in the <math>T_{before} \cup T_{after}</math> where <math>p_i = n_i/N</math>,  <math>N</math> is the number of the samples represented by <math>g</math> diameter  values, <math>n_i</math> is the number of samples having value <math>i</math> for <math>i \in \{0, 1, \dots, g-1\}</math>  so that <math>N = n_0 + n_1 + \dots n_{g-1}</math>.</p>
21	$P = \frac{\sum_{i=1}^M D(i)}{\sum_{j=1}^N D(j)}$ <p>where <math>N = 16</math> samples = window size,  <math>M=8</math> samples that have maximum diameter values  in the window are used to calculate the peak powers  the window is centered around the stimulus onset.</p>

**Table A-3:** Pupillary features used for classification.

22	$E_{contour} = E_{continuity} + E_{curvature} + E_{gradient} + E_{amplitude}$ <p>The <math>E_{continuity}</math> measures the spacing between the the samples along the signal: <math>E_{continuity} = \bar{d} -  v(s_j) - v(s_{i-1}) </math></p> <p>The <math>E_{curvature}</math> measures the smoothness of the signal:</p> $E_{curvature} =  v(s_{i-1}) - 2v(s_j) + v(s_{i+1}) ^2,$ <p>The <math>E_{gradient}</math> is the gradient magnitude at each point along the signal:</p> $E_{gradient} = \frac{G_{min} - \nabla(v(s_i))}{G_{max} - G_{min}},$ <p>The amplitude energy is computed as <math>E_{amplitude} = \frac{v(s)}{\max(v(s))}</math>,  where <math>v(s_i)</math> is the current point, <math>v(s_{i-1})</math> is the predecessor point to <math>v(s_i)</math>,  <math>v(s_{i+1})</math> is the successor point to <math>v(s_i)</math>,  <math>v(s_j)</math> is a point located in the neighborhood of <math>v(s_i)</math>,  <math>G = \nabla v(s)</math> is the gradient magnitude for the entire signal.  The <math>E_{continuity}</math> and <math>E_{curvature}</math> are computed for each neighborhood point and then normalized to fall within the range <math>[0,1]</math>.  The <math>E_{gradient}</math> is negative and falls within the range of <math>[-1, 0]</math>.</p>
----	--

# Bibliography

- [1] J. Allanson and S.H. Fairclough. A research agenda for physiological computing. *Interacting with Computers*, 16:857–878, 2004.
- [2] B.J. Fisch. *EEG Primer*. Elsevier, 2005.
- [3] J. Beatty. Pupillary response in a pitch-discrimination task. *Perception and Psychophysics*, 2:101–105, 1967.
- [4] J. Beatty and J.W. Brennis. The pupillary system. In John T. Cacioppo, Louis G. Tassinary, and Gary G. Berntson, editors, *Handbook of Psychophysiology*. Cambridge University Press, 2000.
- [5] P.L. Nunez and R. Srivivasan. *Electric fields of the Brain: The Neurophysics of EEG*. Oxford, 2005.
- [6] S. Luck. *An Introduction to the Event-Related Potential Technique*. MIT Press, 2005.
- [7] J. Winters. Just short of telepathy: can you interact with the outside world if you cannot even blink an eye? *Psychology Today*, 2003.
- [8] J. Hockenberry. The next brainiacs. *Wired Magazine*, August 2001.
- [9] W. Rosenstiel M. Bogdan, M. Schroder. Artificial neural net based signal processing for interaction with peripheral nervous system. *1st Int. IEEE EMBS Conf. on Neural Engineering*, 2003.
- [10] E.R. Miranda and A. Brouse. Toward direct-computer musical interfaces. *Proceedings of the 5th International Conference on New Instruments for Musical Expression (NIME06)*, 2006.
- [11] N. Birbaumer. Slow cortical potentials: Plasticity, operant control, and behavioral effects. *Neuroscientist*, 5:74–78, 1999.
- [12] Using a bci as a navigation tool in virtual environments. *Proceedings of the 2nd International Brain-Computer Interface Workshop and Training course*, pages 49–50, September 2004.
- [13] G.R. McMillan and G.L. Calhoun. Direct brain interface utilizing self-regulation of steady-state visual evoke response. *Proceedings of RESNA*, pages 693–695, June 1995.
- [14] G. Pfurtscheller, D. Flotzinger, and J. Kalcher. Brain-computer interface: a new communication device for handicapped persons. *Journal of Microcomputer Application*, 16:293–299, 1993.
- [15] C. Andeson, E. Stolz, and S. Shamsunder. Multivariate autoregressive models for classification of spontaneous eeg during mental tasks. *IEEE transactions on Biomedical Engineering*, 45(3):277–286, 1998.



- [16] J.D Bayliss and D.H. Ballard. Single trial p300 recognition in a virtual environment. *Computational Intelligence: Methods and Applications Proceedings, Soft Computing in Biomedicine*, June 1999.
- [17] C. Gugar, C. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller. Architectures of a pc and pocket pc based bci system. pages 49–50, September 2004.
- [18] S.J. Roberts, W.Penny, and I. Rezek. Temporal and spatial complexity measures for eeg-based brain-computer interfacing. *Med. and Biolog. Engineering and Computing*, 37:93–98, 1999.
- [19] H.S. Kenyon. Unconventional information operations shorten wars. *Signal Magazine, Armed Forces Communications and Electronics Association*, 2003.
- [20] R.F. Brem. The never ending controversies of screening mammography: what is the appropriate callback rate for women undergoing screening mammographic examination? *Cancer*, 100:1549–52, 2004.
- [21] L.C. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda. Single-trial detection in EEG and MEG: keeping it linear. *Neurocomputing*, 52:177–183, 2003.
- [22] Gerson A.D, L.C.Parra, and P. Sajda. Cortical origins of response time variability during rapid discrimination of visual objects. *NeuroImage*, 28:326–341, 2005.
- [23] A.D. Gerson, L.C. Parra, and P. Sajda. Cortical-coupled computer vision for rapid image search. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, June 2005.
- [24] O.D. Bruijn and R. Spence. Rapid serial visual presentation: a space-time trade-off in information pre-sentation. *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 189–192, 2000.
- [25] M. Aguilar, B. Ahrens, M. Qian, K. Zachery, S. Klein, T. Carney, Z. Li, C. Privitera, and L. Renninger. Neurotechnology for image analysts project progress, results and plans. *DARPA NIA PI Meeting*, March 2006.
- [26] E. Basar. *Brain function and oscillations II: integrative brain function, neurophysiology and cognitive processes*. 1998.
- [27] S. Makeig. Auditory event-related dynamics of the eeg spectrum and effects of exposure to tones. *Electroencephalograph and Clinical Neurophysiology*, 86:283–293, 1993.
- [28] S. Thorpe, D.Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.
- [29] D. Tan. Using electroencephalograph signals for task classification and activity recognition. *United States Patent 20070185697*, 2007.
- [30] S. Mathan, S. Whitlow, and D. Erdognuns. Neurophysiologically driven image triage: a pilot study. *Conference on Human Factors in Computing Systems*, 2006.

- [31] L.C. Parra, C.D. Spence, A.D. Gerson, and P. Sajda. Recipes for the linear analysis of EEG. *NeuroImage*, 28:342–353, 2005.
- [32] Lucas Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda. Linear spatial integration for single trial detection in encephelography. *NeuroImage*, 17, 2002.
- [33] W. Kilmesch. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29:169–195, 1999.
- [34] Schacter D. EEG theta waves and psychological phenomena: A review and analysis. *Biological Psychology*, 5:47–82, 1977.
- [35] G. Pfurtscheller and A. Aranibar. Event-related synchronization (ers): an electrophysiological correlate of cortical areas at rest. *Electroencephalography and Clinical Neurophysiology*, 83:62–69, 1992.
- [36] G. Pfurtscheller and A. Aranibar. Mapping of event-related desynchronization of and type of derivation. *Electroencephalography and Clinical Neurophysiology*, 70:190–193, 1988.
- [37] O. Lowenstein I. Loewenfeld. *The pupil: anatomy physiology and clinical applications*. Wayne State University Press, 1993.
- [38] M.W. Scerbo, J.P. Bliss, P.J. Milkulka F.G. Freeman, and S.S Robinson. Measuring task-related and task-unrelated thoughts. In Dennis K. McBride and Dylan Schmorow, editors, *Quantifying Human Information Processing*. Lexington Books, 2005.
- [39] L.C. Parra, C.D. Spence, A. Gerson, and P. Sajda. Response error correction - a demonstration of improved human-machine performance using real-time eeg monitoring. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11:173–177, 2003.
- [40] S. Makeig A. Delorme. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics. *Journal of Neuroscience Methods*, 134:9–21, 2004.
- [41] S. Thorpe and M. Thorpe. Seeking categories in the brain. *Science*, 291:261–263, 2001.
- [42] M. E. Tipping. The relevance vector machine. In S.A. Solla, T.K. Leen, and K.R. Miller, editors, *Advances in Neural Information Processing Systems*, volume 12, pages 652–658. MIT Press, 2000.
- [43] M.E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.
- [44] A.R. Reibman and L.W.Nolte. Design and performance comparison of distributed detection networks. *IEEE Transactions on Aerospace and Electronic Systems*, 23:789–797, 1987.
- [45] B.V. Dasarathy. *Decision fusion*. IEEE Computer Society Press, 1994.

- [46] E. Basar. *Brain function and oscillations I: brain oscillations, principles and approaches*. 1998.
- [47] S. Karakas E. Basar, C. Basar-Eroglu and M. Schurmann. Gamma, alpha, delta and theta oscillations govern cognitive process. *International Journal of Psychophysiology*, 39:241–248, 2001.
- [48] G. Pfurtscheller and A. Aranibar. Event-related cortical desynchronization detected by power measures of scalp eeg. *Electroencephalography and Clinical Neurophysiology*, 42:817–826, 1977.
- [49] H. Russeger T. Pachinger W. Kilmesch, M. Doppelmayr and J. Schwaiger. Induced alpha band power changes in the human eeg and attention. *Neuroscience Letters*, 224:73–76, 1998.
- [50] T. Demirap M. Schumann Basar-Eroglu, E. Basar. P300-response: possible psychophysiological correlates in delta and theta frequency channels. *International Journal of Psychophysiology*, 13:161–179, 1992.
- [51] A.K. Engel C.S Hermann, M.H. Munk. Cognitive functions of gamma-band activity: memory match and utilization. *Trends in Cognitive Sciences*, 8:347–355, 2004.
- [52] L. Ward. Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7, 2003.
- [53] M.J. Posner. *Images of mind*. *Scientific American Library*, 1997.
- [54] G. Edelman and G. Tononi. *A universe of consciousness: how Matters becomes imagination*. BASIC BOOKS, 2000.
- [55] D.O. Hebb. *The organization of behavior*. New York: Wiley, 1949.
- [56] M. Schrmann C. Basar-Eroglu E. Basar T. Demiralp, A. Ademoglu. Detection of p300 waves in single trials by the wavelet transform. *Brain Lang.*, 66:108–128, 1999.
- [57] D. Gabor. Theory of communication. *Journal of IEE*, 93:426–441, 1946.
- [58] I. Daubechies. *Ten lectures on wavelets*. Cambridge University Press, 1992.
- [59] O. Bertrand. Time-frequency digital filtering based on an invertible wavelet transform: an application to evoked potentials. *IEEE Transactions on Biomedical Engineering*, 41(1), January 1994.
- [60] A. Grossman P. Goupillaud and J. Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23:85–102, 1984.
- [61] L.Nolte M. Qian, M. Aguilar. The role of multi-level fusion in eeg-based classification for image triage tasks. *submitted to IEEE tracsactions on Cybernetics*, 2008.
- [62] I.Y. Hoballah and P.K. Varshney. Distributed bayesian signal detection. *IEEE Transactions on Information Theory*, 35:995–1000, 1989.

- [63] R.R.Tenney and N.R.Sandell. Detection with distributed sensors. *Proceedings of the IEEE Conference on Decision and Control*, 1:501–510, 1980.
- [64] Z. Chair and P.K. Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Transaction on Aerospace and Electronic System*, 98, 1986.
- [65] T.Carney S.Klein M. Aguilar C. Privitera, L. Renninger. Pupil dilation during visual target detection. *xxxx*, 2007.
- [66] E. Hasanbelliu Y. Barniv, M. Aguilar. Using emg to anticipate head motion for virtual-environment applications. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, 52, JUNE 2005.
- [67] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.
- [68] A.W. Whitney. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 20:1100–1103, 1971.
- [69] T. Marill and D.M. Green. On the effectiveness of receptors in recognition system. *IEEE Transactions on Information Theory*, 9:11–17, 1961.
- [70] J. Kaiser B. Kotchoubey A. Kubler, N. Neumann. Brain-computer communication: self-regulation of slow-regulation of slow cortical potentials for verbal communication. *Archives of Physical Medicine and Rehabilitation*, 82:1533–1539, 2001.
- [71] D.J. McFarland G. Pfurtscheller J.R. Wolpaw, N. Birbaumer. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113:761–791, 2002.
- [72] E.A. Curran and M.J Stokes. Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (bci) systems. *Brain and Cognition*, 51:326–336, 2003.
- [73] T. Allison. Face recognition in human extrastriate cortex. *Journal of Neurophysiology*, 71:821–825, 1994.
- [74] N. McDermott N. Kanwisher. The fusiform face area: a module in huamn extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17:4302–4311, 1997.
- [75] et al. J.V. Haxby. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293:2425–2430, 2001.
- [76] F. Tong Y. Kamitni. Decoding the visual and subject contents of the human brain. *Nature Neuroscience*, 8:679–685, 2005.

## Biography

Ming Qian received the Bachelor of Arts degree from Carroll College, MT, the Master of Science degree in Electrical Engineering from Virginia Tech and the Master of Science degree in Computer Science from Duke University. He worked as a design engineer with Alcatel Network Systems (now Alcatel-Lucent), as a research associate with Duke University Neurobiology department before he enrolled in PhD program at Duke University. He started as a research technical intern while he pursues his PhD and is working as a research scientist at Rockwell/Teledyne Scientific and Imaging LLC, intelligent systems and planning division.

His research interests include multi-sensor fusion, decision fusion, pattern classification, machine learning, biomedical signal processing, biomedical data analysis, physiological computing, brain-machine interface and computational neurobiology.