博士論文

論文題目　Efficient Brain-Computer Interface based on Event-Related Potential

(事象関連電位に基づく効率的 Brain-Computer Interface に関する研究)

氏名　　徐　雅明

Efficient Brain-Computer Interface based on Event-Related Potential

A Thesis Presented for

The Doctor of Philosophy

Degree

the University of Tokyo

Yaming Xu

March 2014

Acknowledgements

I express my sincere gratitude to my supervisor Dr. Yoshikazu Nakajima for all his mentoring, support, and patience throughout the work of this dissertation. I would like to thank the members of my doctoral committee, Drs. Shin Ishii, Mamoru Mitsuishi, Mitsuru Uesaka, Shu Takagi, and Etsuko Kobayashi, for their suggestions and support. I would also like to thank all the professors in the Department of Bioengineering for their guidance throughout these years. In addition, I feel tremendous gratitude towards my previous and current colleagues at Nakajima laboratory, in particular to Jack Liang and Riaz Raufy for their tenacious efforts to polish my scientific writing. Furthermore, the work in this thesis would not have been possible without the support of all my colleagues, most notably, Jue Jiang, Lassad Ben Younes, Nobuyuki Tanigaki, and Zijian Cao, as well as Joon-Hwan Kim, and Masaki Chino.

Special thanks to Dr. Xiaolei Qu and Mitsuo Matsunaga for being my tutors and helping me through the administrative tasks during my first year at the University of Tokyo. I had the privilege to collaborate with Dr. Toki Saito, Takehito Doke, Takaaki Takeshita, Min-Kyu Kim, Ibrahim Furkan Ince, Akira Bekku, Herol Gaibin, Ryuzo Goto, Ryo Yoneda, Hwaseon Jang, Minoru Onoda, Masanori Ito, Michifumi Tomita, and kenji Akimoto and I thank them for their encouragement and constant friendship throughout these years. In addition, I would like to thank lab secretary, Hiromi Hosokawa, who's kindness in the lab always put a smile on our faces no matter how frustrated we were with research. It was a pleasure to be part of the Graduate

School of Engineering and the Department of Bioengineering at the University of Tokyo where I was given the opportunity to pursue my higher education at one of the world's greatest research centers.

Finally, I would like to thank the China Scholarship Council Fund, Center for Medical System Innovation, and Graduate School of the University of Tokyo for the financial supports that were provided to me in order to complete my doctoral degree.

Abstract

Million of people around the world suffer from degenerative diseases that impair the normal neural pathways that control muscle movements. At late-stages of these diseases, people lose all voluntary muscle control and become completely "locked in" to their bodies, unable to communicate through normal or traditional means. Brain-computer interfaces (BCIs) can allow these patients to communicate again by creating a new communication channel directly from the brain to an external output device. Recent studies on amyotrophic lateral sclerosis (ALS) patients showed that event-related potential (ERP) based BCI technology could create new pathways of communication for paralyzed people.

Although these technical demonstrations are encouraging, practical applications of ERP-based BCI technology that meet the needs of people with severe disabilities are significantly impeded primarily by two issues. First, ERPs as the control signals in ERP-based BCIs have an unfavorable ratio between signal (pure ERP components) and noise (neural background activity and various artifacts), which leads to unreliable ERP classification. Second, current epoch-based ERP classifications are still inefficient because they do not exploit full potentials of ERP signals. Mainly due to these two issues, current ERP-based BCI systems have produced impressive laboratory demonstrations but no device of appreciable clinical value.

This dissertation sets out to address these problems, in order to work toward an ERP-based BCI system that can advance the practicality. To overcome the first issue of low signal to noise ratio (SNR) of the ERP signals, the classical Farwell-Donchin

paradigm is optimized in stimulus size and inter-stimulus distance. Moreover, the row/column (RC) stimulus pattern is replaced by a single character (SC) stimulus pattern. To conquer the second issue of inefficient ERP classification based on noisy EEG signals, we model the ERP signals considering a so-called overlap effect which has never been solved before, utilize all informative ERP signals, and further incorporate prior natural language knowledge to compensate low SNR of ERP signals.

The principal results demonstrate that the significantly larger EPRs are elicited by spatial optimization of the classical Farwell-Donchin paradigm. Dramatically improved online performances in terms of accuracy, information transfer rate, and speed are achieved by further incorporating a statistical language model into the optimized stimulus paradigm. Significantly higher single-trial ERP classification accuracy is achieved by fusing prior language knowledge (obtained from the statistical language model) into an enhanced ERP probabilistic model in a Maximum A Posterior (MAP) approach. The current dominating approach uses binary ERP classifier to classify single ERP epochs and this is essentially different from the proposed method, which utilizes all ERP epochs evoked by a stimulation trial in a naive Bayesian approach. Moreover, for the first time the overlap effect is greatly reduced.

In summary, the results presented in this dissertation encompass two advances that are critical to the successful translation of brain-computer interface from their current state of primarily laboratory demonstrations into practical communication devices for the paralyzed.

Contents

List of Tables **xiii**

List of Figures **xv**

1 Introduction **1**

1.1 Communication Option For The Paralyzed 1

1.2 Brain-Computer Interfaces (BCIs) 2

1.2.1 Neurophysiological Background 2

1.2.2 Neuroimaging Techniques 4

1.2.3 Brain Signals Used to Drive BCIs 6

1.2.4 Extracting Features from Brain Signals 8

1.2.5 Classification and Translation 8

1.2.6 Applications of BCIs 9

1.3 Event-Related Potential (ERP) based BCIs 9

1.4 Purposes of the Dissertation 10

1.5 Organization and Contents 11

2 State-of-the-Art ERP-based BCI Techniques **13**

2.1 Neurophysiological Background of ERPs 14

2.1.1 P300: P3a and P3b 15

2.1.2 Visual Evoked Potential (VEP): N1/N200 19

2.2 ERP-based BCI Systems 21

2.2.1 Stimulus Paradigm: Encoding Intention 23
2.2.2 Machine Learning: Decoding Intention 26
2.3 Performance Measures of ERP-based BCIs 32
2.3.1 Online vs. Offline 32
2.3.2 Application-Specific Measures: Accuracy and Speed 33
2.3.3 Theoretical Measures: Information Transfer Rate 34
2.3.4 Measures Considering Error-Correction: Written Symbol Rate 35
2.4 Technical Challenges in ERP-based BCIs 35
2.4.1 Independence of ERP-based BCIs 35
2.4.2 Discriminability of Single-Trial ERPs 37
2.4.3 Information Transfer Rate 39
2.4.4 Practical Usability 41
2.5 Two Themes toward Efficiency in the Dissertation 42
2.5.1 THEME I: Better ERPs Encoding 42
2.5.2 THEME II: Better ERPs Decoding 42

3 THEME I: Two-Level Predictive (TLP) Matrix Paradigm 43
3.1 Motivation: to Elicit More Discriminable Signals and Encode More Information Bits 43
3.2 Previous Studies to Optimize Stimulus Paradigm 44
3.2.1 Physical Properties of Stimuli 46
3.2.2 Physiological and Psychological States of Subject 50
3.2.3 Mental Tasks 50
3.3 Spatial Manipulation of Matrix Paradigm 52
3.3.1 Inter-Symbol Distance 52
3.3.2 Symbol Size 54
3.4 The Proposed TLP Matrix Paradigm 54
3.4.1 Two-Level 3×3 Paradigm: to Elicit Better "0/1" Signals . . . 56
3.4.2 Hypothesis I: 3×3 Matrix Paradigm outperforms 8×8 One . 57

3.4.3 Single Cell (SC) Pattern versus Row/Column (RC) Pattern . 58
3.4.4 Hypothesis II: SC Pattern outperforms RC One 58
3.4.5 Predictive Step: to Encode More Information Bits 59
3.4.6 Hypothesis III: No Significant Difference between *Step-1*, *-2*, and *-p* 65
3.4.7 Hypothesis IV: TLP Paradigm outperforms Classical RC paradigm 65
3.5 Summary 66

4 THEME II: Trial-based ERP Classification with Natural Language Processing 67

4.1 Motivation: to Exploit Temporal Information in ERP Trials and Fuse Language Knowledge to Improve ERP Classification 68
4.2 Features for ERP Classification 69
4.2.1 Spatio-temporal Feature 69
4.2.2 Other Features 70
4.3 Straightforward Trial-based ERP Classification 70
4.3.1 Drawbacks of Trial-based Approach 72
4.4 State-of-the-Art Epoch-based ERP Classification 74
4.4.1 Conventional Binary-class ERP Model 75
4.4.2 Linear Classifiers 78
4.4.3 Non-linear Classifiers 82
4.4.4 Drawbacks of Epoch-based Approach 82
4.5 The Proposed Trial-based ERP Classification 83
4.5.1 Proposed Multi-class ERP Model 83
4.5.2 Feature-dimension Reduction by Linear Projection 85
4.5.3 ERP-decoding by Maximum A Posteriori (MAP) Estimation . 87
4.5.4 N-gram Language Model 90
4.5.5 From Single-trial to Multi-trial 91

4.5.6 Hypothesis V: Conditional Independence of Epoch-Projections 92
4.5.7 Hypothesis VI: Proposed Method outperforms Conventional Methods . . . 92
4.6 Summary . . . 93

5 Experiments, Results, and Discussion 94
5.1 Subjects and Apparatus . . . 94
5.1.1 Subjects . . . 94
5.1.2 Hardwares . . . 95
5.1.3 Softwares . . . 95
5.2 Study I: Initial Investigation of Stimulus Size and Inter-Stimulus Distance Effects . . . 96
5.2.1 Stimulus Setup . . . 96
5.2.2 Experimental Procedure . . . 97
5.2.3 Data Acquisition and Processing . . . 98
5.2.4 Event-Related Potentials . . . 98
5.2.5 Discussion . . . 100
5.3 Study II: Effects of Spatial Manipulation and Stimulus Pattern on ERPs and Offline Performance . . . 101
5.3.1 Experimental Procedure . . . 101
5.3.2 Data Acquisition, Pre-processing, and Classification . . . 103
5.3.3 Behavioral Data . . . 104
5.3.4 Event-Related Potentials . . . 104
5.3.5 Single-trial Binary Classification Accuracy . . . 107
5.3.6 Offline Accuracy, Information Transfer Rate (ITR), and Written Symbols Rate (WSR) . . . 110
5.3.7 Discussion . . . 110
5.4 Study III: Online Validation of the Proposed TLP Paradigm . . . 113
5.4.1 Experimental Procedure . . . 113

5.4.2 Data Acquisition, Pre-processing, and Classification 115
5.4.3 Behavioral Data 116
5.4.4 Online Accuracy, ITR, and Time to Complete the Task 116
5.4.5 Discussion 119
5.5 Study IV: Prediction Impact on Online Performance of TLP Paradigm 121
5.5.1 Dataset Description 121
5.5.2 Comparison between *Step-1*, *-2*, and *-p* 121
5.5.3 Discussion 122
5.6 Study V: Validation of Conditional Independence Assumption 123
5.6.1 Experimental Procedure 123
5.6.2 Data Acquisition and Pre-processing 123
5.6.3 Results of Pearson's χ^2 Independence Test 124
5.6.4 Discussion 124
5.7 Study VI: Validation of the Proposed Trial-based ERP Classifier . . . 126
5.7.1 Dataset Description 126
5.7.2 Pre-processing and Classifier Training 127
5.7.3 Single-trial Classification Accuracy of the Proposed Method . 128
5.7.4 Proposed Method versus Conventional Ones 131
5.7.5 Error Distributions 135
5.7.6 Classifier Outputs 136
5.7.7 Learning Curves 137
5.7.8 Discussion 139

6 Conclusions and Perspectives 141
6.1 Conclusions 141
6.2 Perspectives 143

Bibliography 145

A Event-related potentials (ERPs) **172**

A.1 ERP responses of 8×8 RC, 3×3 RC, and 3×3 SC at all 9 electrodes 172

A.2 ERP responses of *Step-1*, *Step-2*, and *Step-p* at all 9 electrodes 173

B Questionnaire **175**

C Validation of Gaussian Normality of ERP Model Proposed in Chapter 4 **177**

List of Tables

3.1 Previous studies devoted to optimizing stimulus paradigm 47

4.1 Supervised learning methods exploited in ERP-based BCIs 75

5.1 30 symbols (10 runs $\times$ 3 symbols/run) for calibration experiments for 8×8 and 3×3 matrices, respectively. 103

5.2 P300 peak latencies of 8×8 RC, 3×3 RC, and 3×3 SC over electrodes Fz, Cz, and Pz for each of S10 105

5.3 N200 peak latencies of 8×8 RC, 3×3 RC, and 3×3 SC over electrodes O1 and O2 for each of S10 109

5.4 Peak-to-peak amplitudes of non-target responses of 8×8 RC, 3×3 RC, and 3×3 SC over electrodes Fz, Cz, and Pz for each of S10 . . . 109

5.5 Independence violation times for all the 20 subjects. 'M' and 'SD' denote mean and standard deviation, respectively. 126

C.1 p-values of Kolmogorov-Smirnov (KS) tests for validating normality of the proposed multi-class ERP model 178

List of Figures

1.1 Typical BCI framework 3

1.2 EEG electrode placement according to 10-20 international system . . 6

2.1 ERP component example and nomenclature 15

2.2 Schematic illustration of variants of oddball paradigm and the P300 (P3a and P3b) component 16

2.3 ERP-based BCI Framework and Challenges 22

2.4 Farwell-Donchin Speller Matrix 24

2.5 A stimulus trial of a 3×3 RC paradigm 25

2.6 Human brain cortex distribution and an example of P300 and N1/N200 elicited by Donchin and Farwell paradigm 27

2.7 An example describing the difficulty of single-trial ERP classification 31

2.8 A simplified ERP-based BCI operation line demo 32

2.9 An example describing the overlap effect in the classical RC paradigm 39

3.1 Stimuli, subject, and task relationship 46

3.2 The effect of the distance to the center on visual acuity 53

3.3 A virtual keyboard implemented based on the traditional 8×8 RC paradigm 55

3.4 Two-level strategy in matrix paradigm 56

3.5 Ryan's predictive speller 60

3.6 Kaufmann's predictive text entry paradigm 61

3.7 An example for describing predictive step (*Step-p*) in the TLP paradigm 62
3.8 Spelling simulation based on seven sample texts 64

4.1 An example of 9-class ERP-trials 72
4.2 Basic macroscopic model of EEG generation 76
4.3 Visualization binary-class Gaussian distributions of ERP-epochs . . . 78
4.4 An example shown ERP component contamination of non-target ERP-epochs 84
4.5 Feature dimension reduction by LDA extension 86
4.6 A text spelling scene using the 3×3 TLP by a subject 87

5.1 EEG Apparatus and electrode placement 95
5.2 Experimental setup for investigating stimulus size (SS) and inter-symbol distance (ISD) effects on ERPs 97
5.3 ERP results obtained for stimulus size (SS) investigation 99
5.4 ERP results obtained for inter-stimulus distance (ISD) investigation . 100
5.5 Comparison experiment setup 102
5.6 Grand P300 waveforms of the 8×8 RC, 3×3 RC, and 3×3 SC . . . 106
5.7 Grand N200 waveforms of the 8×8 RC, 3×3 RC, and 3×3 SC . . . 107
5.8 Single-trial binary classification accuracy of the 8×8 RC, 3×3 RC, and 3×3 SC 108
5.9 Offline performances of the 8×8 RC, 3×3 RC, and 3×3 SC 111
5.10 Online operation scene of the 8×8 RC paradigm 114
5.11 Online operation scene of the 3×3 TLP paradigm 115
5.12 Online accuracy of the 8×8 RC, 3×3 SC, and 3×3 TLP 117
5.13 Online information transfer rate (ITR) of the 8×8 RC, 3×3 SC, and 3×3 TLP 118
5.14 Total time to spell (with error correction) an English sentence with 57 characters 119
5.15 Offline comparison between *Step-1*, *-2*, and *-p* 122

5.16 Pairwise p-values of Pearson χ^2 test under condition #1 to condition #9125
5.17 Grand mean single-trial classification accuracy map 129
5.18 Single-trial classification accuracy map of subject #1–#10 130
5.19 Single-trial classification accuracy map of subject #11–#20 131
5.20 Offline ERP decoding accuracies for subject #1–#4 132
5.21 Offline ERP decoding accuracies for subject #5–#8 133
5.22 Offline ERP decoding accuracies for subject #9–#12 133
5.23 Offline ERP decoding accuracies for subject #13–#16 134
5.24 Offline ERP decoding accuracies for subject #17–#20 134
5.25 ERP-decoding error distributions 136
5.26 Outputs of ERP-decoding classifiers of subject #1 138
5.27 Learning curves of 8 representative methods 139

A.1 Grand mean ERPs of the 8×8 RC, 3×3 RC, and 3×3 SC at all electrodes 173
A.2 Grand mean ERPs at of the *Step-1*, *Step-2*, and *Step-p* in the TLP at all electrodes 174

B.1 Questionnaire for Study II and Study III 176

Chapter 1

Introduction

1.1 Communication Option For The Paralyzed

Dr. Stephen Hawking is one of the most famous theoretical physicists alive on the earth; however, he has been suffering a degenerative nerve disease called amyotrophic lateral sclerosis (ALS) for decades. The ALS disease has left the scientist confined to a wheelchair and unable to move even his fingers. To write or speak, he currently uses an infrared sensor system mounted on his eyeglasses. The sensor reads his cheek twitches to control the curser of wheelchair-mounted computer to select letter or numbers on a digital keyboard. It can take a half-hour for Hawking to navigate the keyboard and write a couple of sentences in response to a question. Not only Hawking, nearly nine thousands ALS patients are diagnosed in Japan each year, according to statistics from Japanese Ministry of Health, Labour, and Welfare [1]. Moreover, patients over 70 years of age account for nearly half of the total amount of ALS patients [2]. Since Japan is an aging country, the situation will become more critical in the coming years. ALS is only the "tip of an iceberg," because there are numerous other degenerative diseases such as brainstem stroke, brain or spinal cord injury, cerebral palsy, muscular dystrophies, multiple sclerosis, etc. They affect about one million people in Japan [3], and far more around the world [4]. These paralyzed patients at late-stage of

diseases may be totally locked in their bodies and have lost all voluntary muscle control, including eye movement and even respiration. Modern medical technology can prolong their lives; however, a viable means of communication for these patients has not been produced and this is a serious matter because more than ninety percent of ALS patients reject to prolong their lives due to loss of communication [5]. Therefore, restoring communication abilities of neuromuscular patients is critically meaningful and important. Brain-computer interfaces provide a potential way to "unlock" these "locked-in" patients.

1.2 Brain-Computer Interfaces (BCIs)

Brain-computer interfaces (BCIs) are communication systems, which allow people to interact with computers just by using their brain signals [6]. In other words, BCI operation does not depends on normal output pathways of peripheral nerves and muscles. BCI builds a new communication channel between the primate brain and the external world of the primate body. A typical BCI framework is presented in Figure 1.1. Human can voluntarily modulate their brain states to encode their intentions. The modulated brain state can be measured by various neuroimaging technologies that gather information from electrical, magnetic, or metabolic signals. These signals can be digitalized and transmitted to computers, where human intentions can be decoded by analysing the modulated signals. At the same time, the decoded result is mapped to control external devices, as operation feedback received by the users.

1.2.1 Neurophysiological Background

Human brain contains hundreds of billions of neurons, each of which form synapses with an average of ten thousands other neurons. Electrical signals transmitted across the huge neuron network reflect brain activity. The state of the neuron network is the so-called brain state. Why does the brain state reflect human intention? Most

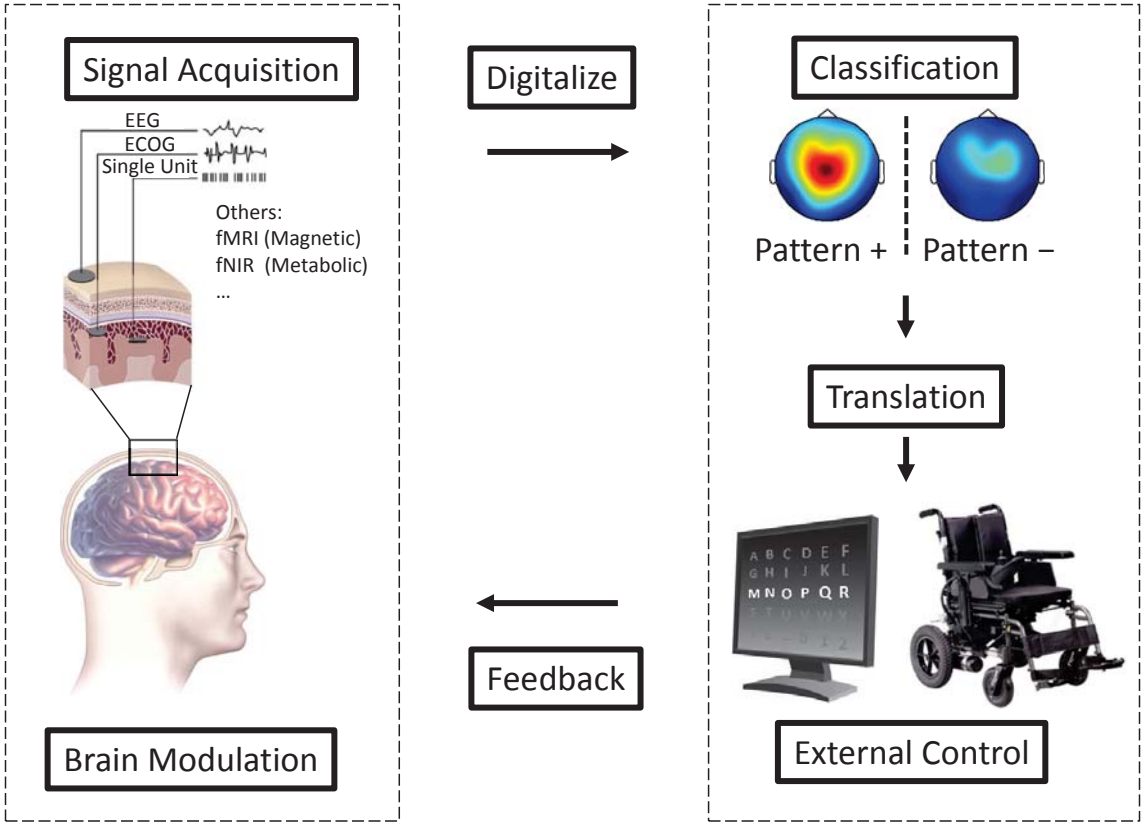


Figure 1.1: Typical BCI framework. Brain state can be modulated (by performing mental tasks) to encode human's intention. The brain state can be measured by various neuroimaging technologies, and then acquired signals are digitalized and transmitted to a computer. By analyzing (classification) the modulated brain signals, the real intention can be decoded to interact with external devices. At the same time, the user can receive interaction feedback (adapted from [7]).

neuroscientists agree that mental and neural activities are potentially correlated [8]. That means different emotions and actions we experience are associated with different patterns of brain activity, represented by the state of neuron network. Theoretically, any mental experience (intention) has a corresponding electrical signature in the brain.

However, fully acquiring the electrophysiological signature of brain state is extremely difficult. This is due to the limitations of current neuroimaging technologies that are incapable of recording each neuron's state due to the huge quantity of brain

neurons. Also, brain states recorded with low spatial fidelity may not detect tiny difference between similar mental experiences. Even if the activity of very neuron was quantified, the challenge of decoding every possible human thought or intention seems impossible. With this in mind, BCI systems attempt to classify very limited amount of brain signatures that can be used to control an interface and are insensitive to spatial resolution.

1.2.2 Neuroimaging Techniques

The neuroimaging techniques can be divided into three classes based on the extent of invasiveness.

1. Invasive

Intracortical Neuron Recording [9]: This is an invasive recording modality where microelectrodes are implanted inside the brain cortex to capture spike signals and local field potentials from neurons. This method can acquire highest quality signals; however, it is prone to scar-tissue build-up, causing the signals to become weaker or even lost as the body reacts to a foreign object in the brain.

2. Partially Invasive

Electrocorticography (ECoG) [10]: Electrodes are placed on the surface of the brain instead of implanting them into brain. Hence, this way is classified as a partial invasive method. Brain activities recorded by the ECoG suffer from signal strength loss because the signal deteriorates when traveling through the brain tissues toward the surface. Moreover, signals recorded by each ECoG electrode are combination of multiple neurons leading to reduce spatial resolution. Compared with totally invasive neuron recording, the ECoG has a lower risk of surgery; but it can only acquire middle quality signals.

3. Non-invasive

Non-invasive neuroimaging is more attractive because of the low risk to patients. Besides electrical recording way, magnetic and metabolic modalities have also been implemented in non-invasive neuroimaging.

Electroencephalography (EEG) [11]: Electrodes are placed on scalp according to specific standard placement (*e.g.*, 10-20 system [12], see Figure 1.2). The small disc-shaped electrodes are stuck to the scalp with contact gel. The electrodes and gel can easily be removed or washed off after the recording session. The EEG signal is usually recorded at many locations simultaneously by one electrode at each position (the term *channel* is often used to refer to a recording position). Comparing with the direct neuron recording and ECoG, EEG has no risk of surgery. However, acquired signals are worse than the above two modalities. The signals are distorted by brain tissue, skull, and scalp, which are served as a low-pass filter. Hence, high frequency information is lost by EEG recorded signals. Fortunately, EEG signals have high temporal resolution, which is particularly important for event-related potential recording. EEG is still the most popular measuring method for BCIs, since it is non-invasive, economic, portable, and has high temporal resolution. In this study, we use EEG as measuring method.

Magnetoencephalography (MEG) [14]: The magnetic fields induced by electrical currents in the brain can be measured directly over the head in a shielded room. Beside the magnetic modality, MEG signals have high spatial resolution compared with EEG signals. Unfortunately, the MEG is very expensive and technically demanding since it is prone to be distorted by movement artifacts and therefore needs to be recorded in a shield room.

In addition, near-infrared spectroscopy (NIRS) and functional magnetic resonance imaging (fMRI) capturing metabolic activities are also exploited for BCI signal measurement. However, they all suffer from low temporal resolution.

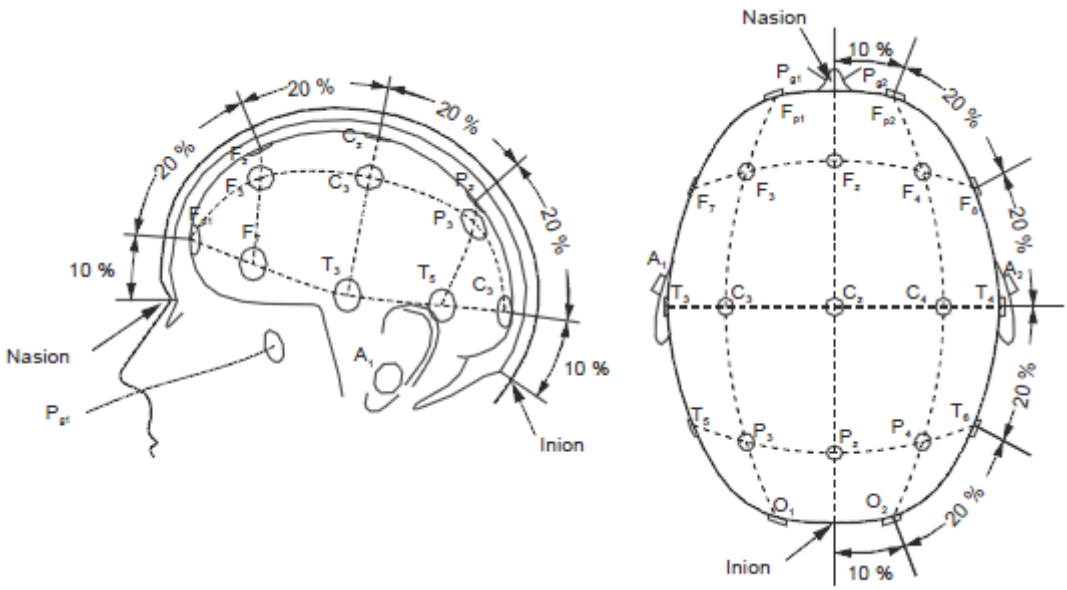


Figure 1.2: Side and top views of the 10-20 international system. The system is the standard naming and positioning scheme for EEG applications. Electrodes are placed based on an iterative subdivision of arcs on the scalp starting from craniometric reference points include nasion, inion, left and right pre-auricular points. The intersection of the longitudinal and lateral is named the vertex. (original figure from [13]).

1.2.3 Brain Signals Used to Drive BCIs

As we described in subsection 1.2.1, not any mental activity can be used as control signal for BCIs. Neuroscientists found that several neurophysiological signals can be used to encode human intention. These signals can be divided into two main categories.

1. Spontaneous signals

The signals are voluntarily generated by users and follow an internal cognitive process such as motor imagery, arithmetic operation, etc. The most used signals are sensorimotor rhythms. In addition, slow cortical potentials and non-motor cognitive signals have also been exploited.

μ/β **rhythm**: The μ/β rhythm over sensorimotor cortex can be modulated by subject motor imagery [15–17]. For instance, hand-grasping imagery can reduce contralateral sensorimotor μ/β rhythm; the phenomenon is called event-related desynchronisation (ERD). Correspondingly, hand relax imagery may increase contralateral sensorimotor μ/β rhythm, which is named event-related synchronisation (ERS).

There are two approaches to exploit ERD/ERS in BCIs. The first approach was pioneered by Dr. Wolpaw from Wadsworth Center, New York State Department of Health, which allows users to obtain the skills to voluntarily control the amplitude/power of μ/β rhythm through a long-term operant conditioning. The second approach is that machine learning methods that are trained to classify ERD/ERS, in stead of human training. The current trend is that both brain and computer (algorithm) are adaptively trained.

Slow Cortical Potentials (SCP): The SCP are very slow variations of cortical activity, which can last from hundreds of milliseconds to several seconds [18]. Users can learn to modulate the SCP to be positive or negative by a relatively long-term operant conditioning.

Non-motor cognitive signals: Various non-motor cognitive tasks have been exploited to modulate specific frequency bands of brain signals, for instance, mental arithmetic computation, mental rotation of 3D objects, etc. [19].

2. Evoked signals

Users unconsciously generate signals when they perceive a specific external/internal stimulus. The non-spontaneous signals or evoked signals rely on external or internal stimulus events. The evoked signals can be further divided into two categories: endogenous signals and exogenous signals [20].

Endogenous signals: The signals are "invoked" by an internal cognitive process when the subject perceives a specific stimulus, such as event-related potential (P300).

Exogenous signals: The exogenous signals such as visual evoked potentials (VEPs) are "evoked" by pre-attentive processing of incoming visual stimuli.

In this study, we focus on evoked signals, hence this topic will be described in detail later.

1.2.4 Extracting Features from Brain Signals

Generally, raw brain signals are given in form of multi-channel temporal signal segments. In order to facilitate further decoding processing, specific features containing discriminative information should be extracted from raw brain signals.

Feature extracting methods should be selected based on the specific brain signals. Spontaneous signals usually are featured in frequency bands (second-order feature); hence, band power features [21], power spectral density (PSD) features [22], and autoregressive (AR) coefficients [23] are proper choices. While, evoked signals are featured in temporal domain (first-order feature); therefore, raw data, signal peak amplitudes [24], area under signal curve [24], or refined spatio-temporal samples [25, 26] are generally used.

Moreover, some advanced feature selection (dimension reduction) methods such as independent component analysis (ICA) [27] and common spatial pattern (CSP) [28] can also be used.

1.2.5 Classification and Translation

Features extracted from different brain states should have different patterns. These different features with different patterns should be recognized by a classification procedure. Then, they can be translated (mapped) to execute pre-defined tasks.

Besides signal modulation procedure (intention encoding), the signal classification is another critical procedure in BCIs. The procedure sometimes is also called intention decoding, which is a reverse procedure of intention encoding. Both supervised [29] and unsupervised [30] classifications have been applied by BCIs. Currently,

supervised learning is dominated in BCIs. Moreover, different learning models have been exploited, such as generative model [26], discriminative model [31], linear, and non-linear models [29].

1.2.6 Applications of BCIs

BCIs offer their users new communication and control channels without any intervention of peripheral nerves and muscles. Hence, many researchers focus on building BCI applications, in the hope that this technology could be helpful for those with severe motor disabilities. Generally, BCI applications can be divided into medical uses and non-medical uses.

Medical Applications: These applications focus on restoring users' communication and control abilities; for example, text speller [18, 24], mouse controlling [17], wheelchair locomotion [32], environment controlling [33], and rehabilitation [34].

Non-Medical Applications: Diversity of non-medical applications have been developed, such as recently emerged brain painting [35], computer games [36], neuro-ergonomic interface for product design [37], and monitoring [37].

1.3 Event-Related Potential (ERP) based BCIs

We devoted to improving event-related potential (ERP) based BCIs. The details of ERP-based BCIs will be described in chapter 2. In this section, two major reasons are given for why we focus on ERP-based BCIs.

1. Most human can voluntarily and easily evoke ERPs (majorally P300) brain signals by an oddball paradigm, in contrast to other brain signals (*e.g.*, ERD/ERS, which will takes humans relatively long time to learn to voluntarily modulate. To make matters worse, some humans could not master the ability even after spending a long-time training [38].

2. ERPs can be reliably acquired by EEG. EEG may be the most popular brain signal measuring way for BCIs, since it is cheap, portable, non-invasive, and with high temporal resolution, in spite of with low spatial resolution and low signal-to-noise ratio (SNR). High temporal resolution is vital for ERPs measuring, because ERPs are a series of time-sensitive brain voltage signals.

In summary, ERPs are economical to be evoked and recorded. Thus, we believe ERP-based BCI is a promising way to achieve cheap, easy and fast interfaces.

1.4 Purposes of the Dissertation

Despite promising achievements have already obtained in the literature, the ERP-based BCI field is still very young and there is still much to do in order to advance current laboratory demonstrations into practical communication devices for the paralyzed. The work presented in this dissertation belongs to the framework of ERP-based BCI research. More precisely, it focuses on the stimulus paradigm optimization for eliciting better ERP signals and study of ERP classification in order to advance practicability of ERP-based BCI.

1. To elicit better ERP signals by optimizing stimulus paradigm. The main obstacle impedes practical using of ERP-based BCIs is that single-trial ERP-signal has an unfavourable ratio between signal (ERP) and noise (neural background activities and various artifacts). Stimulus paradigm plays an important role for eliciting ERPs. The classical (Farwell-Donchin) paradigm has been dominating the field for more than 25 years; however, it has not been optimized based on spatial manipulation and information theory. **THEME I** of the dissertation devotes to optimize the paradigm in order to elicit more discriminable ERPs and to encode more information bits.

2. To improve single-trial ERP classification accuracy. Current ERP-based BCIs suffer from low information transfer rate, in contrast to conventional human-computer interfaces. This is majorally attributed to unreliable single-trial

ERP classification. Although low-quality ERP signals bear part of the responsibility, current state-of-the-art ERP classification methods also have not exploited full potentials of ERP signals. **THEME II** of the dissertation proposes a novel ERP classification strategy in order to improve the robustness of single-trial ERP classification.

1.5 Organization and Contents

This dissertation consists of six chapters. The chapter 1 explains why are BCIs important and gives a general introduction of brain-computer interface including the neurophysiological underlying of BCIs, control signals for BCIs, signal processing and classification methods, and its applications. Then, the reason why we focus on EPR-based BCIs is explained. Finally, the purposes and two themes in this dissertation are stated.

In chapter 2, the typical ERP components are first introduced, as well as the experimental paradigms used to elicit them. In addition, we discuss the factors which may influence ERP generation. Then, we review the first ERP-based BCI, including its designated stimulus paradigm and intention decoding algorithms. Next, we introduce the performance measures for evaluating BCIs. At last, we point out several technical challenges, which impede the practical using of ERP-based BCIs.

The THEME I is fully expanded in chapter 3 to address the challenge of low discriminability of target and non-target ERPs. We review the previous studies on optimizing stimulus paradigm. Surprisingly, the classical Farwell-Donchin paradigm has not been comprehensively optimized in terms of spatial manipulation. Consequently, we propose a two-level predictive paradigm based on spatial manipulation and information theory. Finally, several hypotheses are made that need to be validated by designated experiments in chapter 5.

Chapter 4 focuses on THEME II to address the challenge of unreliable single-trial ERP classification. We start from straightforward classification method and point out

why has this strategy never been applied by any ERP-based BCI. Then, we review state-of-the-art approaches and show that they do not fully exploit the potential of ERP signals. Learnt strengthes and overcome drawbacks of the conventional methods, we propose a novel ERP classification strategy based on a hypothesis. At last, we hypothesize that the proposed method could improve the robustness of single-trial ERP classification.

In Chapter 5, we design six experiments to validate our proposed hypotheses. The subjects, apparatuses, experiment protocols, and data processing are presented. Statistical tests are used to test the statistical significance level.

Finally, we make conclusions and perspectives in chapter 6.

Chapter 2

State-of-the-Art ERP-based BCI Techniques

In this chapter, neurophysiological background of event-related potential (ERP) is firstly introduced. The endogenous P300 potential and exogenous N1/N200 potential are emphatically described, as well as experimental paradigms introduced by researchers from neurophysiological fields used to elicit them. Moreover, we discuss the comprehensive factors which influence the event-related components.

With the developing of neuroimaging, digital signal processing, and machine learning, the first ERP-based BCI emerges. The researchers designed a dedicated stimulus paradigm to encode user's intention based on the neurophysiological background. Next, their intention decoding algorithms are also introduced in detail. More importantly, we indicate the dominance of supervised linear learning methods for ERP-based BCIs. Then, we introduce several performance measures for comparing ERP-based BCIs.

Although the ERP-based BCIs have shown their wide success during the past more than 20 years, there are still several technical challenges existed, which hinder the practical using of the ERP-based BCIs. These challenges lead to the two themes in the dissertation.

2.1 Neurophysiological Background of ERPs

With the discovery of the electroencephalogram (EEG) in 1929 [11], the electrical activity of the human brain could be measured by placing electrodes on the scalp. Hans Berger [11] observed that the electrical voltages could be influenced by external events that stimulated the senses. Later, researchers found the event could be either be external/exogenous, such as visual or auditory input [39], or internal/endogenous, for instance, motor imagery [40] or other mental tasks [41]. The potential voltage changes time-locked to a specific event are named event-related potentials (ERPs) [42]. According to the most widely accepted model [42], ERPs are signals generated by neural populations which are activated time-locked regarding a specific event, and they reflect brain processing of the event. Generally, one denotes by a stimulus the event the brain reacts to, and by a response the activity action of the brain.

Since the EEG reflects thousands of simultaneously ongoing brain activity, brain response to a single stimulus or event is not usually visible in the EEG recording of a single trial. To reveal the event-related potential/response, many trials should be conducted, and then their EEG results should be averaged together, canceling the ongoing random activities, see Figure 2.1.

ERP waveform (an example shown in Figure 2.1) consists of a series of positive and negative voltage deflections, which are related to a set of underlying components. These components are characterized by a letter followed by a number. The letter refers to whether N or P according to whether the component is mostly negative (N) or positive (P) during the time window under investigation, and the number refers to either the latency in milliseconds or the component's ordinal position in the waveform.

Next, two kinds of ERP components most relevant in this study will be described in details.

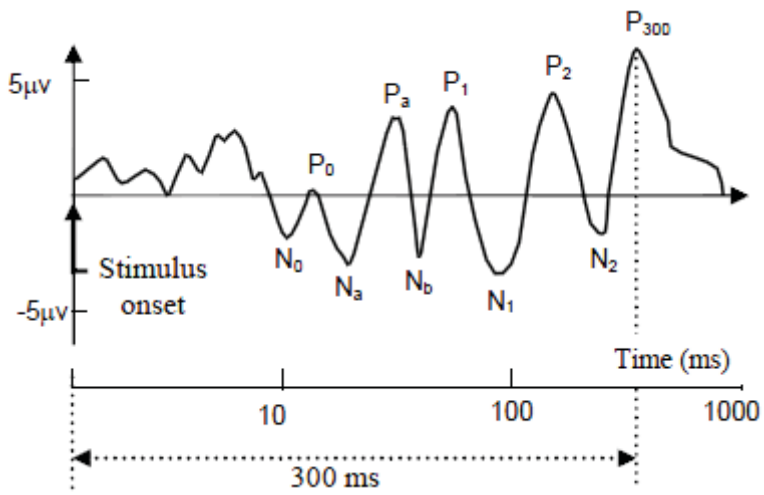


Figure 2.1: A waveform showing several ERP components, including P300, N1/N200, etc., was obtained by averaging hundreds of trials. The components are referred to by a letter (N/P) indicating polarity (negative/positive), followed by a number indicating either the latency in milliseconds or the component's ordinal position in the waveform (original figure from [43]).

2.1.1 P300: P3a and P3b

According to the ERP component nomenclature introduced above, the P300 [44] is a positive deflection elicited about 300 milliseconds after the onset of a stimulus or an event. It was first reported in 1965 by Sutton [45], and the P300 component was evoked by a so-called oddball paradigm [46, 47].

Infrequent stimuli or events randomly occur among a sequence of frequent stimuli or events. The sequence of the stimuli constitutes an oddball paradigm or task. For example, a typical oddball paradigm sequence is shown in the middle of Figure 2.2. Two different stimuli are presented in a random sequence, with one occurring less frequently than the other. The infrequent stimulus is denoted as 'T' with the meaning "Target", while the frequent one is denoted as 'S' with the meaning "Standard". In the oddball task, subjects are instructed to attend to the "Target" stimulus, whereas ignoring the "Standard" one. Then, the P300 component will be elicited in response only to the infrequent "Target" stimulus, whereas it will not be elicited by

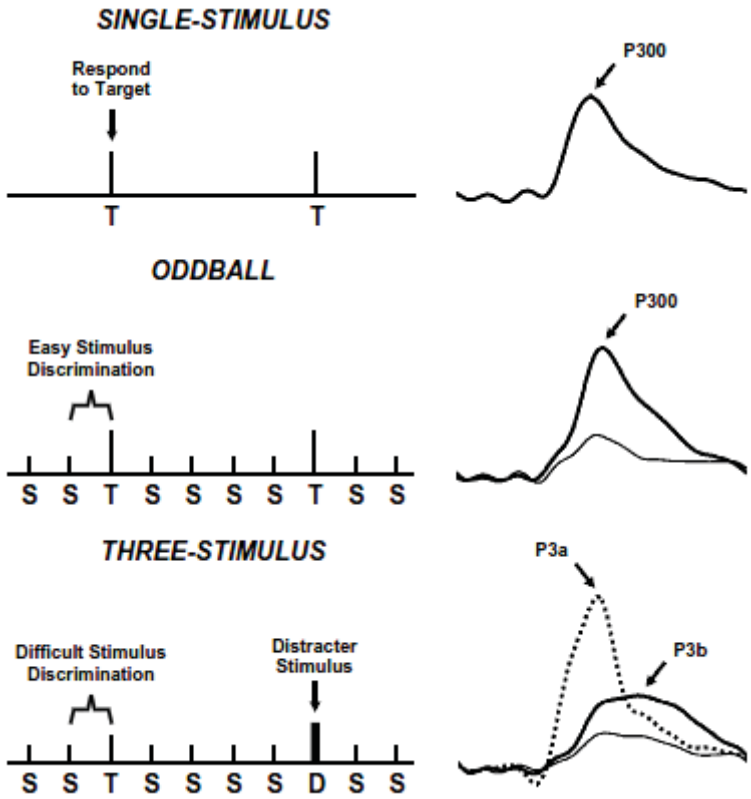


Figure 2.2: Illustration of single-stimulus (top), oddball (middle), and three-stimulus (bottom) paradigms, with the corresponding ERPs plotted at the right of each paradigm. The infrequent target stimulus is denoted by 'T', the frequent standard stimulus is denoted by 'S', and 'D' denotes a distracter stimulus in the three-stimulus paradigm. Note that the distracter elicits a P3a, while the target elicits a P3b (generally refer to the P300) (original figure from [48]).

the "Standard" stimulus. The spatial distribution of the P300 potential is symmetric over parietal cortex (see Figure 2.6). It is largest at the parietal sites and is attenuated as the recording sites moving to central and frontal locations [49, 50].

Actually, the P300 potential may also be evoked by variants of the oddball paradigm [51]. For instance, a single-stimulus paradigm may also evoke a P300-like potential (see top of Figure 2.2). The so-called single-stimulus paradigm is constituted by hiding the "Standard" stimulus of the oddball paradigm. By introducing a

distracter (D) stimulus (with infrequent occurring rate) into the standard oddball paradigm, a three-stimulus paradigm is built (see bottom of Figure 2.2). During the three-stimulus task, subjects are given same instruction as the oddball paradigm that subjects should attend to the "Target" stimulus, whereas ignoring others including both the "Standard" and "Distracter" stimuli. A P300-like component is elicited by the task-relevant "Target" stimulus; meanwhile, a similar component is elicited by the "Distracter" stimulus. Scientists denote the first task-relevant subcomponent as P3b and it is related to the task-relevant stimuli [43]. While, the second subcomponent is denoted as P3a and it is related to the task-irrelevant novel stimuli [43]. The P3b component amplitude is largest over parietal sites as same as the classical P300 potential and thus it is more generally just referred to as P300 in BCI literature [24, 52–57]. The P3a component is more pronounced over the frontal-central sites [49]. However, the P3b is most relevant in this study and we will still refer to it as P300.

The P300 is considered as an endogenous response, since it is caused by internal cognitive processes ("Target" recognition) rather by some automated pre-attentive processing of incoming stimuli. Hence, the P300 could be elicited by various stimuli presented in different sensory modalities, such as visual, auditory, somatosensory, olfactory or even taste stimulation [58]. This is a particularly important properties of the P300 which gives the possibility that P300 could be exploited in various sensory domains for BCI implementation [24, 56, 59].

P300 measurement [43] is based on P300 peak, as follows: P300 peak is defined as the maximal positive peak within a time-range (generally between 250 and 500 milliseconds) after stimulus onset. P300 amplitude is measured as the P300 peak offset relative to a pre-stimulus baseline average. The P300 latency is the time between stimulus onset and the peak of the P300 response. These P300 characteristics are crucial factors for building an efficient P300-based BCI system, especially the P300 amplitude. We will discuss them in details in section 2.2. Next, we will discuss which factors affect these P300 characteristics.

Factors that influence P300 characteristics

Since the P300 generation is related to human subject cognitive processing of stimuli, the characteristics of the P300, such as the amplitude, latency, temporal waveform shape, and its spatial distribution are influenced by several factors which can be divided into following three classes:

1. Physical properties of stimulus paradigm:

Several stimulus's physical properties have been found to affect P300 generation. A well-known determinant of P300 amplitude is target stimulus probability, whether local [60, 61] or global [47, 62–64]. These studies all show that targets that are more probable evoke smaller P300 potentials. The inter-stimulus interval (the temporal interval between the end of a stimulus and the beginning of the following one) was found having positive relationship with P300 amplitude [24]. Longer inter-stimulus interval results in larger P300 potential. Moreover, P300 component elicited by a stimulus represented in visual sensory domain is generally larger in amplitude than auditory [65] or tactile [59] P300.

However, P300 response elicited by a visual stimulus is generally longer in latency than auditory P300 [65].

Spatial scalp distribution of P300 is relatively similar across modalities [39].

2. Physiological and psychological states of subjects:

Human's long-term and short-term physiological and psychological states have been shown to influence their cognitive ability. Hence, these states may finally affect P300. The P300 characteristics including amplitude, latency, and scalp distribution generally varies between subjects and even within subjects over different sessions due to the variation of subjects' cognitive abilities [8].

Generally, P300 characteristics are genetically transmitted according to the study [66]. Another study has also found there is significant correlation of measured P300 between inter-family members [67]. Human's long-term physiological and psychological states affect P300. P300 amplitude has been shown to decrease with

normal aging [50, 68] and cognitive deterioration due to illness [69]; however, P300 latency increase. Individuals with shorter P300 latency have been shown to have superior cognitive functioning [70]. Human gender also has influence to P300 which was revealed by the studies [71, 72].

Meanwhile, human's short-term physiological and psychological states also affect P300 generation. Several findings support this hypothesis. Subjects who are tired show a longer P300 latency than normal human do [73]. Food intake has also been shown to affect P300 measurement. Individuals who have eaten recently show a higher P300 amplitude and shorter P300 latency than those who have not [74]. Recent nicotine, caffeine, and alcohol consumption affect P300 generation [75–77]. Moreover, current motivation [78] and alert state [79] also have been found to affect P300 potential.

3. Mental task demanding:

Mental task involved into the oddball paradigm has two sub-tasks: a perceptual task and an action task [20]. The perceptual task is what analysis of the stimuli a subject needs to do upon their presentation; while, the action task is what reaction should be done by the subject after analysis of the stimuli. In traditional oddball paradigm, the perceptual task consists of distinguishing target stimuli from standard ones; while, the action task is generally mental counting or physically pressing a button.

The overall task difficulty has already been revealed to influence P300 amplitude and latency. The P300 component evoked by harder tasks showed smaller amplitude and longer latency [62]. However, a recent study found a interesting effect that more engaging action task evoked larger P300 [20].

2.1.2 Visual Evoked Potential (VEP): N1/N200

N1/N200 is an early visual evoked component that is part of normal response to visual stimulation [80]. It is a type of visual event-related potential – a series of voltage

deflections observed in response to visual onsets, offsets, and changes. N1/N200 is so named to reflect the polarity and typical timing of the component. The "N" indicates that the polarity of the component is negative with respect to an average of a pre-stimulus baseline average. The "1" indicates that it is the first negative-going component, and "200" indexes the typical peak of this component, which is around 150 to 200 milliseconds post stimulus. The N1/N200 may be detected over occipital, parietal, and frontal sites (see Figure 2.6). Interestingly, N1/N200 component is evoked contralateral to the visual field of stimulus due to the underlying of visual processing [81].

Similar with the P300 component, the N1/N200 is measured based on the N1/N200 peak detection. The N1/N200 peak was defined by the most negative deflection point between 150 and 200 milliseconds. Amplitude of the N1/N200 is the peak offset relative to the pre-stimulus average baseline; while latency of the N1/N200 is the time delay of the peak from the visual stimulation onset.

The characteristics of visual N1/N200 is very sensitive to the performance of visual ERP-based BCIs. Numerous experiments have been conducted to investigate possible factors that affect the characteristics of the N1/N200.

Factors that influence N1/N200 characteristics

Research has suggested that the amplitude of N1/N200 is affected by stimulus luminance, which is positively correlated with the size of N1/N200 amplitude [82]. The amplitude of N1/N200 is also greater in response to attended stimuli than unattended stimuli [83, 84]. Recent studies on N1/N200 have expanded into the processing of socially-relevant stimuli. Attention is particularly relevant to the processing of emotional stimuli because emotional stimuli are believed to receive preferential attention and perceptual processing. A study have found that smiling or sad faces evoked greater N1/N200 than the neural faces [85]. These findings are guiding researcher to introduce emotional stimuli into the field of visual ERP-based BCIs.

Similar with the task difficulty effect on the latency of the P300, latency of the N1/N200 seems to increase during tasks that are significantly complex or difficult [86]. Meanwhile, the intensity of visual stimulation has inverse correlation with the N1/N200 latency. The stimuli with more brightness elicit N1/N200 with shorter latency [87].

2.2 ERP-based BCI Systems

Brain-computer interfaces (BCIs) exploit the fact that human can (learn to) voluntarily modulate their brain signals by performing some mental tasks. The modulation may be used to encoding human's various intentions. In other words, the modulated brain signals are the "vehicles" of human's intention. Modern neuroimaging and communication technology provides numerous "pathes" (described in section 1.2.2 of chapter 1) for the intention "vehicles" to drive from inside of human brain to external computers, such as EEG, MEG, etc. When the "vehicle" reaches the computer, the real intention could be decoded from the "vehicle" by modern signal processing and machine learning technologies (described in section 1.2.4 and section 1.2.5 of chapter 1). Then, the computer could execute predefined action according to the human's intention. Hence, there are at least three challenges (see Figure 2.3) should be overcome before implementing an ERP-based BCI:

1. As shown in section 2.1, some event-related potentials, such as P300 and visual N1/N200, are very good candidates to be used as "vehicle" for encoding intentions of human. Because either P300 or visual N1/N200 could be elicited by voluntary attention manipulation during a visual oddball task. More importantly, most people can elicit them [38]. In contrast, some other possible "vehicles" (e.g. motor-imagery based event-related desynchronization/synchronization (ERD/ERS), described in section 1.2.3 of chapter 1) need humans to take relatively more efforts to learn to voluntarily elicit. To make matters worse, some people could not master the ability to modulate ERD/ERS, even after a long-time training [38].

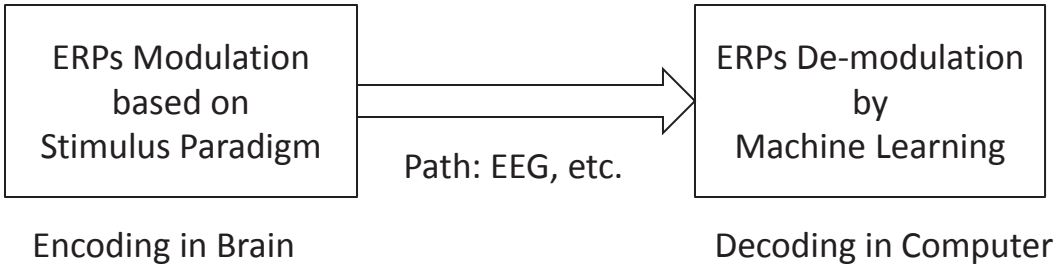


Figure 2.3: The core concept/framework of the ERP-based BCIs. The three components are intention encoding mechanism (ERPs modulation relying on stimulus paradigm), ERPs transporting "path" from brain to computer, and intention decoding mechanism (ERPs demodulation by machine learning methods).

2. ERPs can be reliably acquired by electroencephalography (EEG) which could be a suitable "path" between brain and computer. EEG is the most popular brain signal measuring way in BCI field, since it is cheap, portable, non-invasive, and with high temporal resolution, in spite of with low spatial resolution and low signal to noise ratio (SNR). Due to ERPs (P300 and visual N1/N200) have relatively large spatial scalp distribution, they are not sensitive to spatial resolution. EEG with macroelectrodes is enough to record ERPs. Moreover, ERPs need to be measured in high temporal resolution, since ERPs are a series of time-sensitive brain voltage signals. Therefore, EEG is exactly in line with the requirement of ERPs recording.

3. Modern digital signal processing and computer science provide the last block for building an ERP-based BCI system. EEG signals can be digitalized and then be transmitted to a modern computer. Later, comprehensive signal processing and machine learning methods can be used for real-time decoding the modulated ERPs.

Based on the fact that the three core techniques are prepared, Farwell and Donchin proposed the first ERP-based BCI named P300 speller in 1988 [24]. The so-called P300 speller provides a virtual keyboard that allow people to type texts just using their brain signals. This provides a new communication way for healthy people. More importantly, it provides a possible communication way for neuromuscular patients.

User communication intention is encoded in ERPs which are modulated by a 6×6 row/column stimulus paradigm. ERP signals are acquired using EEG technique. ERPs decoding is performed by using several sophisticated machine learning methods. In the next two subsections, we will describe the intention encoding paradigm and the intention decoding methods. However, how to acquire and transport brain signals is out of scope of this study, thus we will not describe them in depth.

2.2.1 Stimulus Paradigm: Encoding Intention

Farwell and Donchin designed a 6×6 matrix filled with alphanumeric symbols and a "space" symbol, shown in Figure 2.4 (Note that the matrix shown here is not the originally first matrix used by Farwell and Donchi in [24]. However, it was shown in Donchin's second paper on the P300 speller [88]. It is widely known as Farwell-Donchin matrix and has been dominated the field for more than 20 years). The matrix is visually presented on a monitor, served as a virtual keyboard. The matrix background is black, while the 36 symbols are gray-colored in normal state and white-colored in intensified state. During spelling, the rows and columns of the symbols are intensified successively and randomly to constitute an oddball paradigm (see section 2.1). Any given symbol has a rare (1/6) chance of being intensified. The intensification of row and column containing the given symbol can be served as the "Target" stimuli/events in the standard oddball paradigm, while the intensification of other rows and columns can be served as the "Standard" stimuli/events.

Users can communicate any symbol by just attending to it, and mentally identify its intensification. Their attending to the target symbol makes the intensifications of the row/column containing the target symbol task-relevant, and the intensifications are thus served as "Target" stimuli. Hence, onset of the target stimuli may elicit a clear P300 potential inside the users' brain according the neurophysiological background of P300 (see section 2.1). However, other stimuli generally called non-target stimuli in the ERP-BCI field may not elicit P300 potential, theoretically. At

Figure 2.4: The classical Farwell-Donchin Speller Matrix. During spelling, subject is instructed to attend to the desired symbol while the rows and columns of the symbol matrix intensifying randomly and successively. This classical 6×6 row/column (RC) paradigm has been dominated the ERP-BCI field for more than 20 years (original figure from [88]).

the same time, an enhanced visual N1/N200 potential may also be elicited by onset of the target stimuli, since selective attending to visual stimuli could enhance visual N1/N200 component over the occipital cortex of the users (see section 2.1). The Farwell-Donchin matrix paradigm is well known as 6×6 row/column (RC) paradigm.

For ease of explanation, a 3×3 RC stimulus paradigm is presented in Figure 2.5. Before describing in detail, a group of terminologies used in the ERP-based BCI field are introduced below:

Subject: human who participates in the ERP-based BCI experiment;

Stimulus: row/column symbols intensification. Note that stimuli can also be presented in auditory, tactile, or other sensory domains;

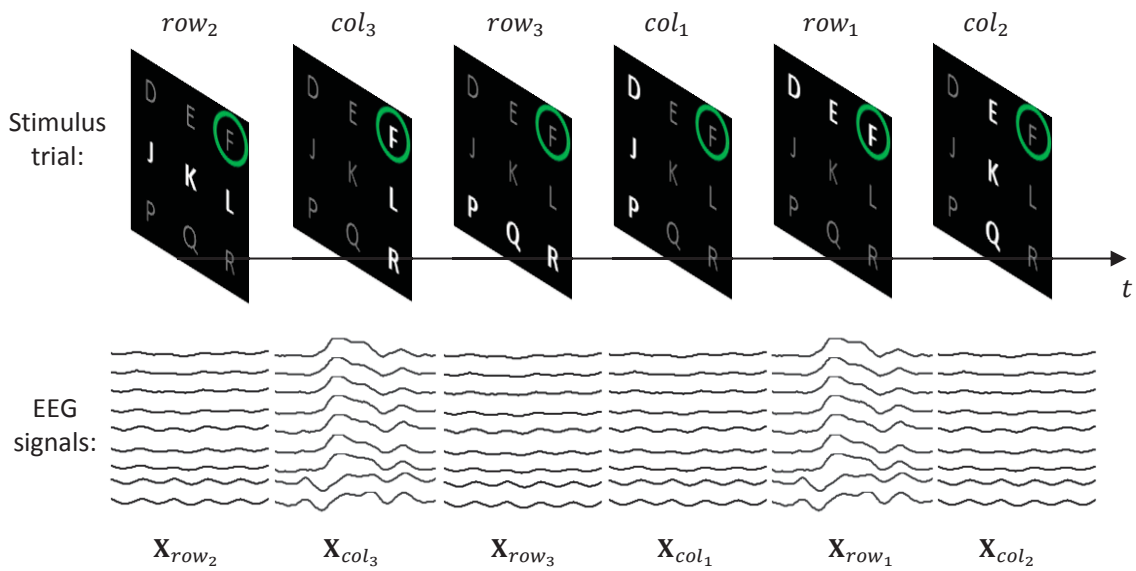


Figure 2.5: A trial of stimulation of an ERP-BCI based on a 3×3 visual matrix. The rows/columns are intensified successively and randomly to constitute an oddball paradigm. The green circle denotes the target which subjects (are asked to) attend to. The green circle will not be shown during experiment. Below each row/column stimulus, the corresponding ERP response is exhibited.

Target stimulus: the stimulus which subject (is asked to) attend to (*e.g.*, stimuli row_1 and col_3 containing 'F' in Figure 2.5, given the assumption that subject is asked to focus on 'F');

Non-target stimulus: the stimulus which subject ignores (*e.g.*, stimuli row_2, row_3, col_1, and col_2 in Figure 2.5);

Stimulus trial (or stimulus sequence): each of the 3 rows and 3 columns in the 3×3 RC paradigm intensifies once and exactly once;

Stimulus onset asynchrony (SOA): the temporal interval between the beginnings of two adjacent stimuli (*e.g.*, the interval between stimuli row_2 and col_3 in Figure 2.5. SOA is usually set to around 100–200 milliseconds in order to achieve efficient BCI performance);

Target-to-target interval (TTI): the temporal interval between the beginnings of two adjacent target stimuli (*e.g.*, the interval between stimuli col_3 and row_1. Apparently, TTI is not constant since the randomness involved in the RC paradigm);

ERP response: a temporal segment of brain response (multi-channel EEG signals) time-locked to a specific stimulus originated from subject's brain (*e.g.*, 9-channel EEG signal segment shown below each stimulus is time-locked to the beginning of the corresponding stimulus and lasts for 800 milliseconds. So the duration of the ERP response here is 800 milliseconds);

Target response: the ERP response corresponding to the target stimulus. Generally, target response contains some modulated ERP components, such as P300, N1/N200, etc. For example, EEG signal segments $\mathbf{X}_{col_3}$ and $\mathbf{X}_{row_1}$ are target responses to target stimuli col_3 and row_1. The target response is plotted in detail in Figure 2.6. Clear P300 components are elicited over the frontal and parietal cortices, while clear N1/N200 components are elicited over the occipital cortex;

Non-target response: the ERP response corresponding to non-target stimulus. Ideally, non-target response only contains background brain signals (*e.g.*, ERPs $\mathbf{X}_{row_2}$, $\mathbf{X}_{row_3}$, $\mathbf{X}_{col_1}$, and $\mathbf{X}_{col_2}$ are non-target responses to non-target stimuli row_2, row_3, col_1, and col_2).

2.2.2 Machine Learning: Decoding Intention

In theory, subject's intention (*e.g.*, selecting 'F' from the 3×3 matrix in Figure 2.5) is fully encoded in a single-trial of ERP responses (*e.g.*, $\mathbf{X}_{row_1}$, $\mathbf{X}_{row_2}$, $\mathbf{X}_{row_3}$, $\mathbf{X}_{col_1}$, $\mathbf{X}_{col_2}$, and $\mathbf{X}_{col_3}$ in Figure 2.5). Hence, given a single-trial of ERP responses $\mathbf{X}_i \in \mathbb{R}^{M \times T}, i = 1, \ldots, n$ with M being the number of channels, T being the number of sampled time points, and n being number of stimuli in the single trial, the target symbol which subject (is asked to or voluntarily) attend to could be decoded theoretically by determining the presence or absence of P300 and N1/N200 in these ERP responses.

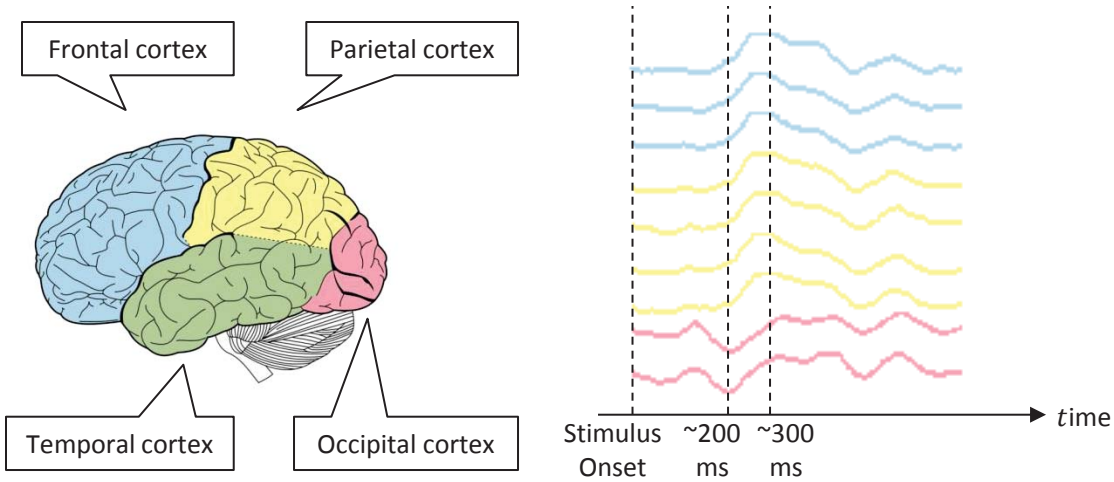


Figure 2.6: Left: Human brain cortex distribution: blue-, yellow-, green-, and pink-colored cortices are frontal, parietal, temporal, and occipital cortices. Central sulcus is located between the frontal and parietal cortices; Right: a target-response sample from Figure 2.5 are colored following the color map of their recording sites shown left. The first seven signal channels recorded at the frontal, central, and parietal cortices show clear P300 components, while clear N1/N200 components are presented in the last two channels recorded at the occipital cortex.

In Farwell and Donchin's first implementation [24], they recorded the brain responses by only one EEG electrode (or channel) at parietal site. A 600-ms ERP response was extracted after each stimulation onset. Thus, only P300 components were expected to be contained in the extracted ERP responses. Then, they applied four different algorithms to detect P300 component [24]:

1. Peak picking: The P300 window was defined as the time range within which the average attended wave form was positive. The window they selected was typically between 220 and 500 milliseconds. The amplitude of peak was computed as the difference between the lowest negative point prior to the P300 window and the highest positive point in the P300 window. Thus, each ERP response can be scored by its peak amplitude. ERP response containing the P300 (target response) will receive a big and positive score, whereas other responses (non-target response) may be scored by smaller values.

2. Area: They summed up all data points within the P300 window to form a measure. Same as the "peak picking", target response may be scored higher than non-target ones.

3. Covariance: A P300 template was computed as the average of the target responses obtained through a calibration experiment (described later). Each ERP response was scored by the covariance (or correlation) with the P300 template. The covariance measures the similarity between any given ERP response and the P300 template. Target response with higher similarity with the P300 template may receive higher covariance score.

4. Stepwise discriminant analysis (SWDA): A stepwise-fit procedure was conducted on calibration data (labeled data) to train a group of weights for each ERP response data point, providing more weight to points that were more critical. Summation of weighted ERP response gives the distance between the ERP response and a hyperplane separating target and non-target responses, which is virtually a binary classifier.

Based on their experimental results, none of the four algorithms is best for all subjects. However, the SWDA is generally more efficient than the others, since it involves modern "training-test" supervised learning schema. It has better generalization ability and also fully exploits the covariance information between ERP data features. Nevertheless, other algorithms do not exploit statistical information of ERP data points.

In 2000, Dochin and colleagues further developed SWDA's potentiality for ERP classifcation [88]. From then on, intention decoding of P300 speller has been majorally relied on ERP classification.

Let's review the task given at the beginning of this subsection. The task is to determine the presence or absence of P300 and N1/N200 in ERP responses. Hence, the scenario of discrimination of ERP responses can be cast into a binary classification problem:

$$f(\mathbf{x}) = \mathbf{w}^T K(\mathbf{x}) + b = \begin{cases} > 0, \mathbf{x} : \textit{Target response} \\ < 0, \mathbf{x} : \textit{Non-target response} \end{cases} \tag{2.1}$$

where $\mathbf{x}$ denotes column vector of ERP response, which is obtained by concatenation of all columns of matrix $\mathbf{X} \in \mathbb{R}^{M \times T}$ into $\mathbf{x} \in \mathbb{R}^{M \cdot T}$, $\mathbf{w}$ is the classification coefficient vector ($\mathbf{w}^T$ denotes transpose of $\mathbf{w}$), $K(\cdot)$ is a translation function, and b is the bias term. For nonlinear classifier, the $K(\cdot)$ can represent a kernel transformation that maps the features into a higher dimensional space in an attempt to create a linearly separable set; while, for linear one, $K(\cdot)$ is simply an identity transformation: $K(\mathbf{x}) = \mathbf{x}$. Generally, target responses are denoted as positive $(+)$ class, whereas non-target responses are denoted as negative $(-)$ class.

The coefficients $\mathbf{w}$ and b in equation 2.1 define a hyperplane. The hyperplane is obtained by a training procedure. The training data with known class labels (target/non-target) is generally collected through a so-called calibration experiment. During the calibration experiment, subject is generally given some symbols to attend following instructions. Meanwhile, ERP responses are recorded. Thus, the acquired ERP responses are with known class labels. The hyperplane with normal $\mathbf{w}$ and bias b is trained to separate target and non-target responses. Ideally, target responses will be located at positive side, while non-target responses will be located at negative side.

After training, any new coming trial of ERP responses from a test procedure can be used to score each symbol $c_{i,j}$ (i denotes the row index, while j denotes the column index) in the Farwell-Donchin's matrix. It can be scored by summing up the scores of its corresponding row stimulus $\mathbf{X}_{row_i}$ and column stimulus $\mathbf{X}_{col_j}$:

$$score(c_{i,j}) = f(\mathbf{x}_{row_i}) + f(\mathbf{x}_{col_j}) \tag{2.2}$$

where $\mathbf{x}_{row_i}$ and $\mathbf{x}_{col_j}$ are the vectorization of $\mathbf{X}_{row_i}$ and $\mathbf{X}_{col_i}$.

Then, the target symbol $c_{\hat{i},\hat{j}}$ which subject attending to (*i.e.*, 'F' in the Figure 2.5) can be determined by :

$$c_{\hat{i},\hat{j}} = \arg\max_{i,j} score(c_{i,j}) \tag{2.3}$$

The target detection in equation 2.3 can be regarded as a multi-class ERP classification based on binary classification of single-trial ERPs. However, single-trial ERPs are with unfavorable ratio between signal (ERP) and noise (neural background activity and various artifacts such as electromyography (EMG), electrooculogram (EOG), mains interference, etc.) [89, 90] and non-stationary (high trial-to-trial variability) [91]. The unfavorable properties of ERP trial make single-trial ERP classification be a very hard problem [26]. In other words, misclassification rate on single-trial basis is high. Farwell and Donchin [24] applied a simple averaging method to improve the SNR of ERP trial. An example is shown in Figure 2.7. By averaging multi-trials of ERP signals, the random noise could be reduced. This fact has a simple mathematical explanation provided that some simplifying assumptions are made. These includes:

1. The signals (P300 and N1/N200) are time-locked to target stimuli, and with invariable latency and shape;

2. The noises (neural background activity and artifacts) can be approximated by a zero-mean Gaussian random process of variance σ^2.

Having defined k, the trial number, m, the channel index, and t the time elapsed after stimulus onset in the kth trial, trial recorded from channel m can be written as $\mathbf{X}^k(m,t) = \mathbf{S}_{ERP}(m,t) + \mathbf{N}(m,t,k)$ where $\mathbf{S}_{ERP}(m,t)$ is the signal and $\mathbf{N}(m,t,k)$ is the noise fulfill Gaussian distribution $\mathcal{N}(0,\sigma^2)$. The average of N trials is:

$$\bar{\mathbf{X}}^k(m,t) = \frac{1}{N}\sum_{k=1}^{N}\mathbf{X}^k(m,t) = \mathbf{S}_{ERP}(m,t) + \frac{1}{N}\sum_{k=1}^{N}\mathbf{N}(m,t,k) \tag{2.4}$$

Obviously, expectation of $\bar{\mathbf{X}}^k(m,t)$ is $\mathbf{S}_{ERP}(m,t)$, while its variance is down-scaled as $\frac{\sigma^2}{N}$.

Thus, a better symbol score can be obtained based on multi-trial averaging. The equation 2.2 can be rewritten as:

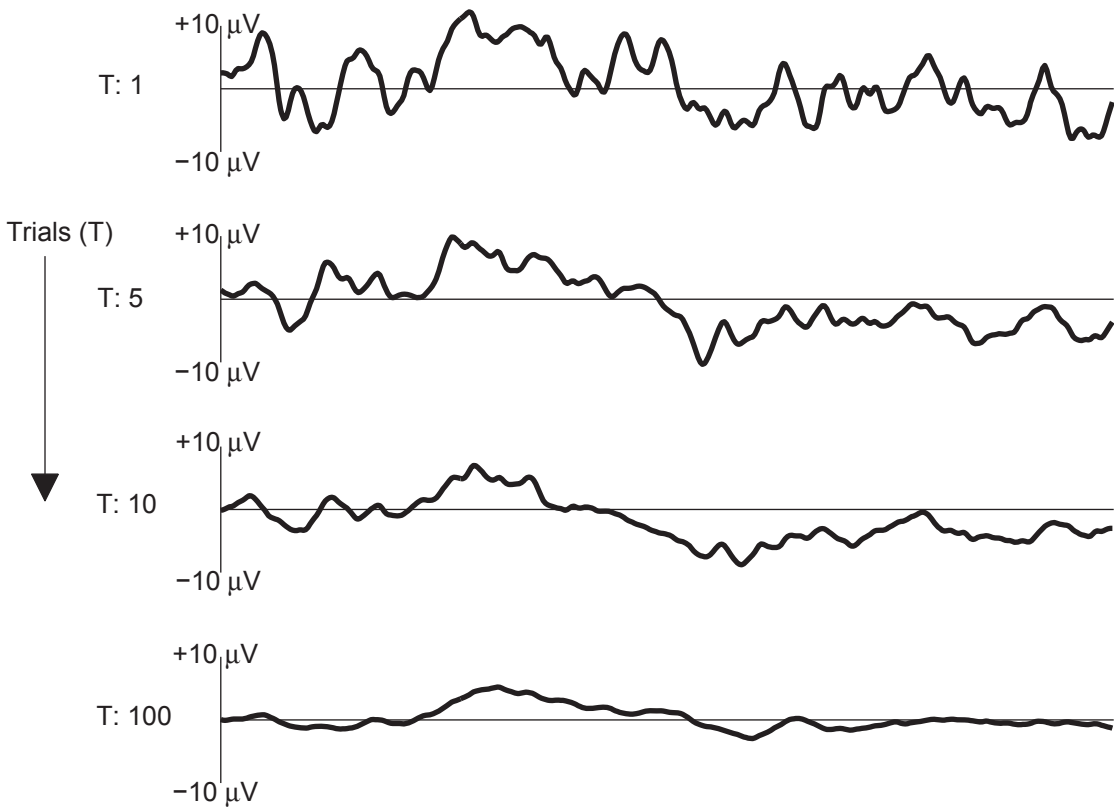


Figure 2.7: A real example describing the difficulty of single-trial ERP classification. Signals are collected by one EEG channel located at parietal site. 800-ms target responses (P300 potentials) averaging over single-trial, 5 trials, 10 trials, and 100 trials are shown from top to bottom. Single-trial ERP response shows the unfavorable ratio between P300 component and ongoing neural background activity. Averaging over more trials, more clear P300 component is revealed.

$$score(c_{i,j}) = f(\frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_{row_i}^{k}) + f(\frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_{col_j}^{k}) \tag{2.5}$$

Alternatively, averaging over single-trial scores is equivalent to equation 2.5 when linear function applied in equation 2.1.

2.3 Performance Measures of ERP-based BCIs

ERP-based BCIs are one kind of human-computer interfaces (HCIs) which are communication systems between human users and computers. To assess the system performance, researchers in the field introduced both application-specific and generally theoretical performance measures [6].

2.3.1 Online vs. Offline

Before introducing performance measures, two important concepts "Online" and "Offline" should be explained. The "line" here means BCI operation line, see Figure 2.8.

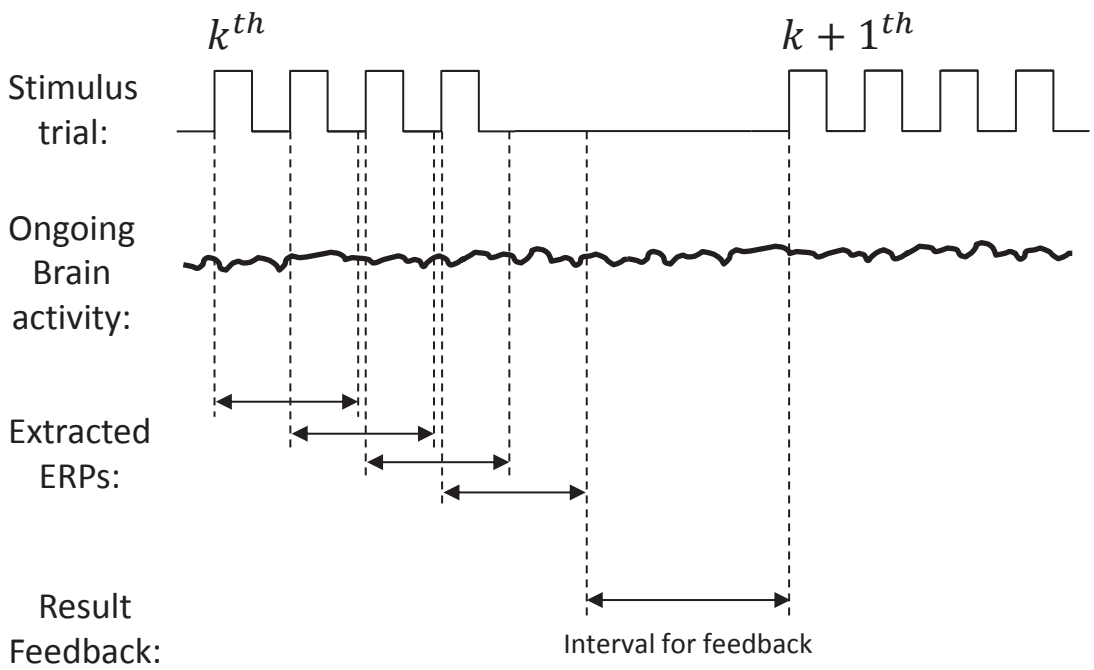


Figure 2.8: A simplified ERP-based BCI operation line demo. Assuming a selection should be made after k^{th} stimulus trial. Hence, result feedback (all ERPs analysis for the selection) should be given during the "Interval for feedback", before next $(k+1^{th})$ stimulus trial starting.

Online: Human intention decoding program is running on an ERP-based BCI operation line. Stimuli present one by one according to pre-defined frequency; meanwhile, ERP response elicited by each stimulus is extracted from ongoing brain activity, and then decoding (classification) result should be computed during "Interval for feedback" based on the collected data. That means online analysis needs to meet very strict time limitation.

Offline: Human intention decoding program is running, literally, off the ERP-based BCI operation line. Offline analysis just performs on previously collected data stored in hard disks or any other storage devices. Naturally, it does not need to fulfill any time limitation. On the other hands, offline analysis can use any recorded data, whereas online analysis can just use currently recorded data. Cross-validation is usually performed offline.

Note that online performance is the golden standard for evaluating a BCI system. In other words, only online performance can provide tangible evidence of any BCI system's validity. There are two reasons according to the work [92]. First, cross-validation does not account for the temporal dependency necessary during online classification. That is, for online classification, only prior data can be used regardless of whether the data were collected one second or one year prior to the online classification. Second, offline analysis only estimates potential online performance. Although it may show that a system is likely to work, only online testing can establish that for certain.

2.3.2 Application-Specific Measures: Accuracy and Speed

For the P300 speller, the application specific measures including typing accuracy and speed are usually used. Typing accuracy P is defined as the proportion of correct symbol selections in the total symbol selections, defined as below:

$$P = \frac{N_{correct}}{N_{total}} \tag{2.6}$$

where, $N_{correct}$ is the number of correct selections and N_{total} is the number of total selections. Speed is defined as how many symbols can be selected in an unit time (*e.g.*, one minute). Speed, sometimes, is replaced by totally consumed time (T_{total}) to complete a spelling task.

2.3.3 Theoretical Measures: Information Transfer Rate

Both accuracy and speed are certainly important. However, they are affected by the characteristics of the specific application and the success with which the system interfaces the user's control of the signal features with that application. For example, two BCIs, one with 36 symbol candidates and the other with 10 symbol candidates, achieve same accuracies. Obviously, the first BCI is better since selection from larger amount of symbol candidates is more difficult. If only giving accuracy, we can not determine which BCI is better. Thus, it makes comparisons between different studies difficult, and it does not reveal what might theoretically be done with the degree of control that the user has.

The standard method for measuring communication systems is information transfer rate (ITR), or bit rate. It is defined as the amount of information communicated per unit time [6], defined as below:

$$ITR = \frac{Bits}{T} \tag{2.7}$$

where, $Bits = log_2(N) + Plog_2(P) + (1-P)log_2(\frac{1-P}{N-1})$ denotes the information bits transferred by one selection; T is the consumed time for one selection, in minutes; and N is the number of symbol candidates to be selected (*e.g.*, $N = 36$ in a 6×6 matrix). The measure incorporates speed (alternatively time), accuracy, and number of symbol candidates (or degree of communication freedom) in a single value. It provides a better measure for comparing different works in the ERP-based BCIs field.

2.3.4 Measures Considering Error-Correction: Written Symbol Rate

As described in the above subsection, information transfer rate provides a simple way to measure and compare performances between different BCI systems. However, it does not consider error correction in spelling application. Furdea and colleagues proposed a novel measure — written symbol rate (WSR) [93], defined as below:

$$WSR = \begin{cases} \frac{2SR-1}{T} & : SR > 0.5 \\ 0 & : SR \leq 0.5 \end{cases} \tag{2.8}$$

where, $SR = \frac{Bits}{log_2(N)}$ indicates the symbol spelling rate.

The measure WSR is a more realistic measure of the actual speed of spelling communication. Assuming that a subject attempts to correct all errors, a SR of less than or equal to 0.5 would indicate that the subject will make more errors than she/he is capable of correcting, and the final typed text will contain more errors than correctly selected letters and likely be indecipherable. Therefore, the range of SR from 0 to 0.5 is assigned a WSR of zero.

2.4 Technical Challenges in ERP-based BCIs

The ultimate goal of ERP-based BCIs is to provide a communication way for neuromuscular patients, especially for totally locked-in patients. However, there are still some technical challenges needed to be solved.

2.4.1 Independence of ERP-based BCIs

Target users of the ERP-based BCIs are people with neuromuscular diseases (*e.g.*, ALS, brainstem stroke, etc.). They may lose all the voluntary muscle controlling

abilities. For example, ALS patients in their very late-stage may be totally locked-in their bodies. They even can not move their eyes and they need breath-assistance system to continue their life. Hence, the operation of BCIs can not rely on any normal output pathway.

Dependence on Peripheral Muscle

Recently, Brunner *et al.* [94] and Treder *et al.* [57] revealed that the classical Farwell-Donchin Speller relies on eye gaze. Covert attention to visual stimuli dramatically hampered the speller's accuracy [57]. Although, various gaze-independent BCIs based on non-spatial visual- [95–100], auditory- [56, 101–104], and tactile-sensory [59] have been proposed, they are still non-competitive with traditional overt visual BCIs in terms of performance. Moreover, operation feedback always visually presents to users. Therefore, current ERP-based BCIs still relies on eye controlling ability, more or less.

Synchronous Working Mode

Synchronous setup of the ERP-based BCIs is another limitation for totally locked-in patients, since the operation is paced by external stimuli. At least, the startup of the external stimuli needs caregivers to trigger. Moreover, for each selecting in the Farwell-Donchin speller, users need to focus their attentions on the desired symbol for a fixed number of stimulus trial presentation. It is tedious and make users tired. Recently, several researchers have attempted to dynamically stop stimulus presentation [103, 105–109]. Schreuder and colleagues [110, 111] studied all these dynamic strategies on ERP-based BCIs; however, they found all these methods worsen performance of low-performing subject.

Most ERP-based BCIs assume users continuously operate them; however, users may transfer to other tasks. For example, a user may type a while by an ERP speller, then switch to watch a video. Thus, the BCIs should automatically detect whether users' attention on them. If non-attention paid, the BCIs should stop processing.

Very few researcher devoted on this topic. Panicker *et al.* [112] integrated a steady-state visually evoked potentials (SSVEPs) paradigm into a conventional P300-BCI to detect whether users focus on the P300-BCI stimulation.

2.4.2 Discriminability of Single-Trial ERPs

Efficiency of ERP-based BCIs relies on high discriminability between target ERP responses and non-target ERP responses. Two reasons limit the discriminability:

Poor-quality EEG signals

1. Low Signal-to-Noise Ratio (SNR): As described in subsection 2.2.2, ERPs recorded by EEG are with unfavorable SNR. Brain signals received by EEG electrodes are hampered by brain tissues, skull, and scalp (see Figure 4.2 in chapter 4); EEG signals are prone to be contaminated by noises (EMG, EOG, mains interference, etc.); Moreover, peak-to-peak amplitude of neural background oscillation is relatively large, about 10–20 μV, compared to P300/N200 peak amplitude (about 5 μV). That means ERP components are buried in extremely noisy background signals (see Figure 2.7).

2. Non-stationary: EEG signal variation commonly exists from subject to subject, session to session, even trial to trial.

Known Effects in the Classical Farwell-Donchin Paradigm

In cognitive psychology experiments (see Figure 2.2), the SOAs applied by stimulus paradigm are usually quite long (several seconds) [20]. In this way, the ERP associated with a stimulus is clearly distinguishable from those associated with the following, or previous, stimuli. In ERP-based BCIs, however, this is not viable since such long SOAs would result in impractically slow communication rates. Therefore, the SOAs applied in current ERP-based BCIs are much shorter (around 100–200 milliseconds).

Generally, very limited (generally one or two) stimuli are presented on monitor in traditional cognitive psychology experiments. In the Farwell-Donchin paradigm

setup, however, 36 (6×6) symbols are presented in order to provide users large degree of communication freedom (DOCF).

They trigger a set of problems and effects:

1. Adjacent Distraction Effect [48, 92, 113, 114]: In the classical Farwell-Donchin paradigm, each symbol is crowded by several symbols. Unfortunately, subject's attention may easily be attracted by stimuli near to the target stimulus [115]. This may elicit a fake P300 by the non-target but adjacent stimuli. Moreover, assuming subject's attention has been attracted by current non-target stimulus, next target stimulus may be ignored by the subject if the interval between current non-target stimulus and next target stimulus is less than 500 milliseconds, since attentional blink may be triggered [116, 117].

2. Refractory Effect [92, 118]: Due to small SOAs (100–200 milliseconds) and randomness applied in the Farwell-Donchin paradigm, the temporal interval between two consecutive target stimuli may be less than 500 milliseconds. In this case, subject may not response to the second target stimulus since a so-called repetition blink may be triggered [116, 119].

3. Overlap Effect [20, 118, 120]: As mentioned above, ERPs elicited by neighbouring stimuli may be temporally overlapped. That because the interval (SOA) between two consecutive stimuli is generally smaller than the ERP durations. In Figure 2.9, the SOA is set to 200 milliseconds in a 3×3 RC paradigm; however, the whole duration of ERP response is set to 800 milliseconds, in order to fully cover ERPs (P300, N200, etc.) evoked by the current stimulus. Since the latencies of these event-related potentials are a bit lengthy, long ERP duration is unavoidable.

EEG signals are virtually with poor quality, and further deteriorated by these terrible effects in the RC stimulus paradigm. It makes the discrimination between target and non-target responses more difficult. It leads to lower BCI performance. In chapter 3, we will firstly review previous studies toward to overcome these defects, and then lead to the first theme in this study.

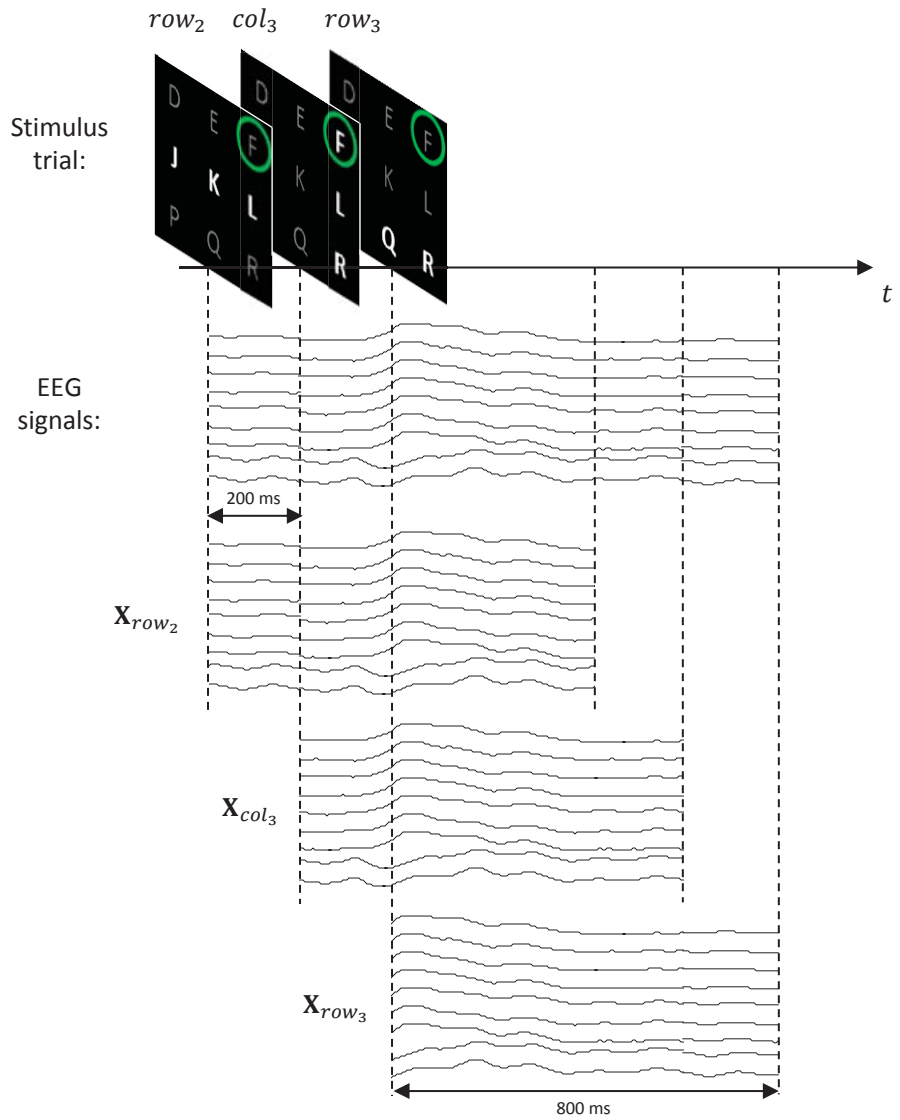


Figure 2.9: A real example describing the overlap effect in current ERP-based BCIs. The SOA is set to 200 milliseconds, while ERP duration is 800 milliseconds. Thus, the neighbouring ERP responses are temporally overlapped. Clearly, non-target ERPs $\mathbf{X}_{row_2}$ and $\mathbf{X}_{row_3}$ also (partly) contain P300 and N200 components. This figure can be regarded as a more realistic version of the Figure 2.5.

2.4.3 Information Transfer Rate

Traditional augmentative and alternative communication (AAC) devices have decent information transfer rates, *e.g.*, recent eye-tracker even reaches 10 words pre minute

[121]. However, the fastest ERP-based BCIs to date provide a typing rate with 5–7 letters per minute [92, 106]. Low information transfer rate may be attributed to following reasons:

Low Decoding Accuracy of Single-trial ERPs

Since the known effects in the RC paradigm and low SNR of EEG signals, target and non-target ERP responses obtained in the classical RC paradigm of ERP-based BCI are with unfavorable discriminability. This makes binary classification of single-trial ERPs very difficult, and makes the final decoding based on single-trial recordings unreliable.

More advanced and complex decoding algorithms perhaps may improve the ERP analysing performance. Complex algorithms are generally time-consuming which may result in low real-time ability. What's more worse, EEG signals are with high temporal sampling rate (about thousands per seconds). It may further deteriorate online ability of decoding algorithms.

Non-Competitive Decoding Speed

Due to unreliable single-trial classification performance, current ERP-based BCIs usually decode based on ensemble of multiple trials of ERP responses, *e.g.*, by averaging ERP responses evoked by same stimulus in a series of trials. Averaging is a simple but powerful way to improve SNR of ERP signals. We have introduced this technique in an intuitive Figure 2.7 in section 2.2.2. This means, before target symbol could be determined, multiple simulation trials should be presented first. This undoubtedly limits BCI's decoding speed.

To date, various machine learning algorithms have been exploited for ERP-based classification to improve BCI's information transfer rate. In chapter 4, we will have a review of previous studies on ERP classification, and then it will lead to the second theme in this study.

2.4.4 Practical Usability

Troublesome Setup

Most current ERP-based BCIs are still designated experimental demos usually used in a noise-free experimental environment, thus they may suffer in practical scenes. Moreover, EEG devices are still too bulky, and EEG recording is very troublesome since gels should be attached to users' scalp for wet-electrodes. After each recording, users need to wash their hairs and clean the electrodes.

Swartz Center for Computational Neuroscience at University of California, San Diego has made great efforts to develop mobile/portable BCI systems [122] and dry electrodes [123]. However, there is still a long way to go.

Calibration Needed for Each Use

As mentioned in section 2.1, ERPs are influenced by physiological and psychological states of subjects. Different subjects may generate different ERP patterns. Meanwhile, ERP patterns recorded from same subject may change with time, due to the changing of physiological and/or psychological states of the subject. Moreover, identical repositioning of the EEG recording apparatus is impossible. Slight deviation of electrode placement may change the ERP patterns. Therefore, the decoding algorithm (ERP classifier) is needed to be re-trained by a calibration experiment for each subject and for each time of using.

Current machine learning techniques greatly reduce the calibration experiment to about 10–20 minutes. However, adaptive decoding algorithms are still eagerly required to adapt to the ERP changing with time. Nevertheless, very few researchers are devoting to solve this problem. Mcfarland and colleagues [124] made a pioneer study on whether adaptive decoding algorithms are needed for ERP-based BCIs. Their results showed adaptive approach did not outperform non-adaptive one. We think that is because their experimental duration was very short (minute level), and the subjects' states were not dramatically changed.

2.5 Two Themes toward Efficiency in the Dissertation

In order to move toward efficient ERP-based BCIs, two key challenges among those challenges stated in section 2.4 urgently need to be addressed.

2.5.1 THEME I: Better ERPs Encoding

Single-trial ERP-signal has an unfavourable ratio between signal (ERP) and noise (neural background activities and various artifacts). Current ERP-based BCIs usually decode based on ensemble of multiple trials of ERP-signals. This obviously hampers efficiency. Naturally, we would like to improve the quality of ERP-signal. Stimulus paradigm plays an important role for eliciting ERPs. The classical paradigm has been dominating the field for more than 25 years; however, it has not been optimized based on spatial manipulation and information theory. **THEME I** of the dissertation devotes to optimize the paradigm in order to elicit more discriminable ERPs and to encode more information bits. This theme will be fully unfolded in the next chapter.

2.5.2 THEME II: Better ERPs Decoding

The low efficiency of current ERP-based BCIs is majorally attributed to unreliable single-trial ERP classification. Although low-quality ERP-signal bears part of the responsibility, current state-of-the-art ERP classification methods also have not exploited full potentials of ERP-signals. **THEME II** of the dissertation proposes a novel ERP classification strategy in order to improve the robustness of single-trial ERP classification. We propose a more reasonable multi-class ERP model than the conventional binary-class ERP model, and build the model in a robust one-dimensional space. The details are described in chapter 4.

Chapter 3

THEME I: Two-Level Predictive (TLP) Matrix Paradigm

This chapter discusses THEME I of this dissertation, which is about how to elicit more discriminable ERP signals by optimizing stimulus paradigm. First, we point out low discriminability as one of the obstacles, which prevent ERP-based BCIs from practical using. Therefore, we are motivated to optimize stimulus paradigm to elicit more discriminable target and non-target ERPs. Then, we review the previous studies on optimizing stimulus paradigm. Surprisingly, the classical Farwell-Donchin paradigm has not been comprehensively optimized in terms of spatial manipulation. Consequently, we propose a two-level predictive (TLP) paradigm based on spatial manipulation and information theory. Finally, several hypotheses are made to be validated by designated experiments in chapter 5.

3.1 Motivation: to Elicit More Discriminable Signals and Encode More Information Bits

As described in section 2.4 of chapter 2, one of the most challenges prevents the ERP-based BCIs from practical using is the low discriminability of single-trial ERPs.

From the perspective of machine learning, the low discriminability of target and non-target ERP responses elicited by target and non-target stimuli, respectively, is because that the ERP components contained in the target ERP responses are with unfavorable ratio to neural background activity. In other words, the ERP components (with discriminable information) are (partly) buried in noisy background activity. This makes target and non-target ERP responses present with indistinguishable patterns that leads to be very difficult for pattern recognition. Therefore, we are motivated to try to elicit more discriminable ERP signals. Simply speaking, we are trying to elicit larger target ERP components. ERPs are elicited through a stimulus paradigm. Naturally, comprehensive studies should be conducted to improve the stimulus paradigm, in order to elicit larger (more discriminable) ERP components.

Stimulus paradigm is not only used for ERP signal elicitation, but also serves as the user interface. However, its identity as user interface has drawn enough attentions. Current ERP paradigm is still naive and should be optimized based on information theory.

3.2 Previous Studies to Optimize Stimulus Paradigm

Stimulus paradigm is placed in an important position in the ERP-based BCI field, since it directly influences the signal generation for further decoding analysis. Good-quality signals may dramatically ease the further signal analyzing.

As reviewed in section 2.1 of chapter 2, stimulus paradigms used to elicit P300 (oddball paradigm) and N1/N200 (visual stimulation) have been comprehensively studied in the field of neurophysiology. Although the underlying mechanism of event-related potentials (ERPs) is still unclear, some factors which influence the characteristics of these ERPs (P300 and N1/N200) have been revealed. Since the stimulus paradigms were adapted to build ERP-based BCIs 26 year ago, these factors

have been greatly limiting the efficiency of the ERP-based BCIs. What's worse, customized stimulus paradigm (*e.g.*, the Farwell-Donchin paradigm) used by the ERP-based BCIs introduced a new set of factors restricting its practicality, as described in section 2.4 of chapter 2.

The classical Farwell-Donchin stimulus paradigm is not optimal. During the past 26 years, lots of researchers devoted to optimize it. However, their studies are quite diver. Some groups focused on improve stimulus presentation. Others investigated influence of physiological and psychological states of human subjects on operation performance of BCIs. Additionally, several works attempted to find better mental tasks. Up to now, there is no comprehensive review of these scattered studies. An initial summary is made in this thesis.

Generally, stimulus paradigm is narrowly regarded as stimulus presentation in literature. In my understanding, however, physical stimulus presentation, human subject, and mental task together constitute the circle of stimulus paradigm, see Figure 3.1. I refine the ERP elicitation procedure by the Farwell-Donchin paradigm according to Figure 3.1. In the classical Farwell-Donchin paradigm, the physical stimuli, the intensifications of rows or columns in a 6×6 matrix, pace subject to response to each stimulus. The response is to execute a perceptual task (first sub mental-task) — recognition of target stimuli. Then, different stimuli signal subject to execute different action tasks (second sub mental-task) — mental counting or ignoring. At the same time, the execution of different mental tasks modulates subject's brain to different states — target ERP responses or non-target ERP responses. Thus, the three components of the stimulus paradigm all contribute to the final discriminability between target and non-target ERP responses.

Past studies toward optimizing each of the three components are summarized in Table 3.1.

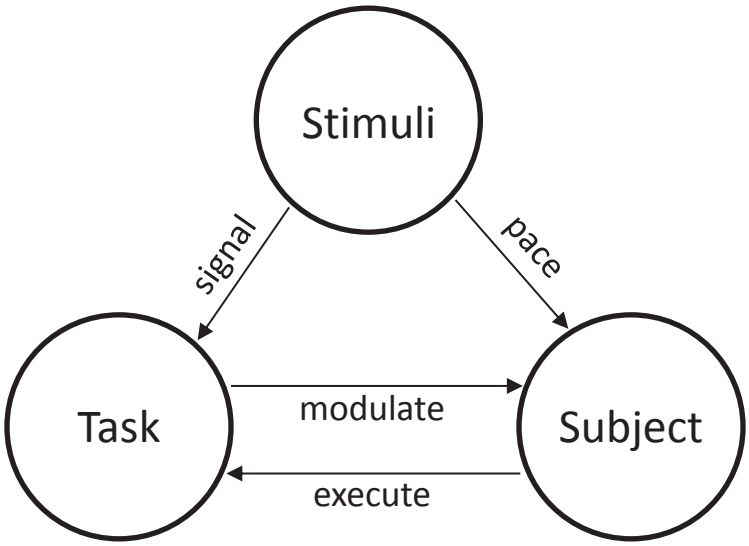


Figure 3.1: Stimuli, subject, and task relationship in the circle of stimulus paradigm. Stimuli pace subject to response to each of them; For each given stimulus, subject execute perceptual recognition task, and then different stimuli signal different action task which subject should execute. The execution of the mental task modulates subject's brain signals to different states.

3.2.1 Physical Properties of Stimuli

I divide the studies on manipulating physical properties of stimuli into six groups:

1. Temporal manipulation: Since temporal characteristics of oddball paradigm is a well-known determinant of P300 amplitude, many pioneers have tried to target a better SOA. Allison and Pineda [8, 127] did early work on the SOA impact on P300 potentials in the field of ERP-based BCIs. They found P300 amplitude is positively correlated with SOA at low target probabilities; however, this finding did not hold for higher target probabilities. Sellers *et al.* [53] compared two SOA configurations (175 and 350 milliseconds) for two RC paradigm with different sizes. As expected, RC paradigm with larger SOA elicited larger P300 potentials. However, they concluded that longer consumed time counteracted the benefit from larger P300s. Recently, McFarland *et al.* [55] compared four stimulus rates (SOA: 31.25, 62.5, 125, and 250 milliseconds, respectively) for a checkerboard paradigm. Keeping the SOA same,

Table 3.1: Previous studies devoted to optimizing stimulus paradigm

Physical properties of stimuli	Temporal-related	TTI [120, 125], SOA [53, 55, 126, 127]
	Visual stimulus appearance	(Emotional) facial stimulus [128–131] Motion-based visual stimulus [118, 132–135] Multi-stimulus [136, 137] Stimulus intensity [138] Stimulus color [139]
	Stimulus modality	Auditory [56, 102–104, 140], Tactile [59], Hybrid [141]
	Stimulus pattern	Checker board paradigm (CBP) [92] "m choose n" C(m, n) [142, 143] Single Cell (SC) [38, 144] Rapid serial visual presentation (RSVP) [96]
	Spatial-related	Matrix size [53, 145, 146] Matrix-beyond [115, 147] Inter-symbol distance [148, 149]
	Stimulus sequence	Fixed sequence [20, 150, 151]
Physiological and psychological states of subjects	Mental workload [152], Motivation [78], Attention [79]	
Mental task	Active mental recognition [20, 153, 154]	

they varied stimulus-on and stimulus-off times to investigate whether there is any effect on P300 potentials. Their result with regard to SOA effect on P300 amplitude is consistent with previous findings. Moreover, they found variation in stimulus-on and stimulus-off times did not significantly affect the P300 amplitude. Höhne and Tangermann [126] tested 14 SOA conditions in an auditory oddball paradigm. They observed a general trend, the amplitudes of the ERPs increase with larger SOA, which is in line with previous findings in visual sensory domain.

2. Spatial manipulation: Due to target probability is another well-known determinant of P300 amplitude, the size of the RC matrix paradigm has been manipulated to vary target probability. Larger-size matrix leads to smaller target probability. Sellers *et al.* [53] compared RC matrix paradigms with size 3×3 and 6×6 in terms of P300 amplitude, and they found the P300 is significantly larger for 6×6 RC paradigm than for 3×3 one; however, the accuracy is higher for the 3×3 RC paradigm than for the 6×6 one. Allison and Pineda [146] compared 4×4, 8×8, and 12×12 matrix paradigm. They found similar matrix size effect on P300 amplitude.

Additionally, different spatial layouts have also been exploited. Fazel-Rezai and colleagues [48, 147] proposed a region-based paradigm to disassociate the symbol at same row or column, in order to mitigate adjacent distraction.

3. Appearance manipulation: Stimulus appearance may dramatically affect human perceptual ability that may lead to affect ERPs. Covington and Polich [138] reported that increases in stimulus intensity produced increases in P300 amplitude and decreases in peak latency. Takano and colleagues [139] found green/blue flicker stimulus resulted in a better performance than the traditional white/gray flicker stimulus. Although they did not show ERP results, we can make a reasonable assumption that chromatic stimulus may elicit larger ERPs. Tangermann and his colleagues [137] compared some fabricated stimulus appearances, including brightness highlight, rotation, scaling, masking, and "combination". They found that the masking effect outperformed all others. This may be attributed to the masking effect is easy to be perceived. Several groups of researchers [118, 132–135] investigated motion-based stimuli for ERP-based BCIs. They found motion change (second-order) stimulus may lead to larger ERPs than traditional luminance change (first-order) stimulus. Facial stimulus also evoked significantly larger P300 than traditional luminance change approach, since higher level cognitive activities involves when subject recognizes a target facial stimulus [128–130].

4. Pattern manipulation: The classical row/column (RC) stimulus pattern has dominated more than 20 years, since it achieves a decent tradeoff between target

probability and duration of stimulus trial. According to the classical neurophysiology literature [62], target probability should be "rare" to constitute an oddball paradigm. However, the duration of stimulus trial can not be too long by considering practical usage. The RC pattern is not optimal, since it introduces some new errors, as described in section 2.4 of chapter 2. Rapid serial visual presentation (RSVP) [95, 96] takes an extreme. It presents symbol stimulus one by one instead of row/column approach. It further reduces target probability; however, the stimulus duration is much longer. Townsend *et al.* [92] proposed a checkerboard paradigm (CBP), which ensures at least 6 non-target stimuli occurring between two consecutive target stimuli, in order to reduce "double flashes" (refractory effect, in section 2.4 of chapter 2) introduced by the classical RC paradigm. Later, Townsend *et al.* [142] and Jin *et al.* [143] proposed a more general pattern framework — "m choose n". Totally "m" stimuli in a complete stimulus trial, while "n" out of "m" stimuli can uniquely determine a single symbol. They manipulated "m" and "n" to find a better stimulus pattern configuration.

5. Sequence manipulation: One of the characteristics of oddball paradigm introduced by neurophysiology literatures [62, 62] to elicit P300s is constructed by a random sequence. The randomness introduces refractory effect in RC paradigm. Salvaris *et al.* [20], Tangermann *et al.* [150], and Fong *et al.* [151] questioned whether the randomness is necessary. Their results showed that regular (not random) sequence can also elicit P300 potentials, despite the amplitude is smaller.

6. Modality manipulation: The visual Farwell-Donchin paradigm relies on eye gaze ability of subjects [57, 94]. To overcome the limitation, auditory, tactile, and hybrid oddball paradigms have been exploited to build ERP-based BCIs. However, ERPs are smaller for auditory [65] and tactile [59] stimuli than for visual stimuli. Researchers from Berlin BCI laboratory made pioneering works to improve auditory stimuli [56, 102, 103, 108].

3.2.2 Physiological and Psychological States of Subject

Subject's physiological and psychological states affect cognitive ability, which further affects ERPs generation. ERP morphology is specific to each subject, since different human have different cognitive abilities. There are very few studies to manipulate subject's states, in order to evaluate their effects on BCI performance. Kleih *et al.* [78] manipulated subject's motivation by money awarding, and they found highly motivated subject generated larger P300 potentials. Lakey *et al.* [79] examined the effects of a short mindfulness meditation induction (MMI) on the performance of a P300-BCI task. Their results revealed that MMI subjects produced significantly larger P300 amplitudes than control subjects at some specific electrodes. Felton *et al.* [152] investigated the mental workload during brain-computer interface training. Although they did not show ERP results, they found that subjects with higher mental workload felt stress and produced more errors.

3.2.3 Mental Tasks

Perceptual and action sub-tasks constitute the complete mental task in ERP-based BCIs. Stimulus appearance may affect perceptual sub-task, however no one studied on it. Very limited researchers pay their attentions on action sub-tasks. Hong and his colleagues [153, 154] made the pioneering work on this topic. During auditory-based BCI experiments, they instructed subjects to actively recognize speaker's gender of current auditory stimulus, rather simply mental counting. They revealed that employing more active task rather than the simple counting elicited significantly larger ERPs. Salvaris *et al.* [20] used a color-recognition task in a visual fixed paradigm. The results showed color recognition enlarged evoked potentials.

In conclusion, most studies have been conducted to optimize physical stimuli from different dimensions, including temporal, spatial, appearance, pattern, sequence, and

modality. Wherein, stimulus pattern and sequence can be classified into temporal manipulation. Some highlights are summarized for each class as below:

Temporal manipulation received most concerns, since oddball paradigm is a sequence represented temporally. More recently, a hypothesis that target-to-target temporal interval (TTI) is the "true" determinant of P300 amplitude has been broadly accepted [20, 43, 118, 120, 155]. Previously revealed SOA and target probability impacts on P300 amplitude are all attributed to TTI effects. Longer TTIs produce greater P300 potentials. Citi and colleagues [120] investigated the TTI effect. They modeled the effect of number of preceding non-target stimuli for both target stimulus and non-target stimulus. They found the overlap effect (described in section 2.4 of chapter 2) is significant for target stimulus with only zero/one preceding non-target stimulus. By applying their modelling approach, significant improvement in accuracy was achieved.

Spatial manipulation majorally focuses on matrix-size optimization. Some case studies [53, 146] made a conclusion that larger-size matrix elicits larger ERPs while keeping SOAs same. However, this may largely be attributed to TTI effect. Larger-size matrix just enlarges the average TTI which leads to larger ERPs. Cinel and colleagues [113] was the first to investigate the adjacent distraction effect in the matrix paradigm. They found neighbouring stimuli are prone to be wrongly selected. Then, Fazel-Reza *et al.* [115, 147] proposed a region-based paradigm to suppress adjacent distraction. Only two studies attempted to investigate inter-symbol distance (ISD) in the Farwell-Donchin paradigm. Sakai and Yagi [149] showed a counter-intuitive results that smaller ISD led to larger ERPs. However, Salvaris and Sepulveda [148] revealed larger ISD achieved better accuracy, even though they did not report ERP results. Therefore, there is a contradiction between their findings.

Appearance and modality manipulation showed that more engaging stimulus may elicit larger ERPs. There is still large space to exploit. Perhaps, subject-specific stimulus is preferred.

3.3 Spatial Manipulation of Matrix Paradigm

The classical matrix (or Farwell-Donchin) paradigm has been dominating the ERP-based BCIs field for about 26 years. In contrast to traditional oddball paradigm used in neurophysiological field, two significant differences are introduced by the matrix paradigm: 1. short temporal stimulus pace (small SOA); 2. crowed spatial distribution of stimulus symbols. The former leads to refractory and overlap effects (described in section 2.4 of chapter 2), which deteriorate ERPs and further hamper the BCI performance. Just because of this, the temporal properties of the Farwell-Donchin paradigm received lots of research attentions [53, 55, 120, 125–127]. In particular, Citi and colleagues [120] documented and modelled the ERPs change due to the number of preceding non-target stimuli. Their work inspires part of my study.

The later difference (spatial distribution of stimulus symbols) has not been comprehensively studied, even though it causes the adjacent distraction effect and partial refractory effect (see section 2.4 of chapter 2). Previously revealed matrix-size effect [53, 145, 146] has been majorally attributed to TTI effect. Other spatial properties including inter-symbol distance and symbol size have not been fully explored. Only two studies on inter-symbol distance, however, achieved contradictory results [148, 149]. Therefore, effects of the two factors should be explored and modelled.

3.3.1 Inter-Symbol Distance

Intuitively, the inter-symbol distance (ISD) should have relationship with adjacent distraction effect to some extent. Psychological study has already drawn human visual acuity curve against to the distance to the focused point, see Figure 3.2. It is clear that the closer the fovea, the more sensitive the visual acuity. A natural hypothesis is that the smaller distance between the focused target stimulus and adjacent non-target stimuli, the more severe adjacent distraction effect. Here are some possible evidences: Frye and colleagues [156] suppressed stimuli surrounding targets during

calibration, and then they obtained better online BCI operation performance. This may because they reduced the possibility that subject's attention attracts by adjacent stimuli during calibration experiments. Therefore, they obtained better training data which led to better classifier. Finally, the classifier achieved better online performance. Salvaris and Sepulveda [148] found matrix with larger ISD achieved a better BCI operation performance.

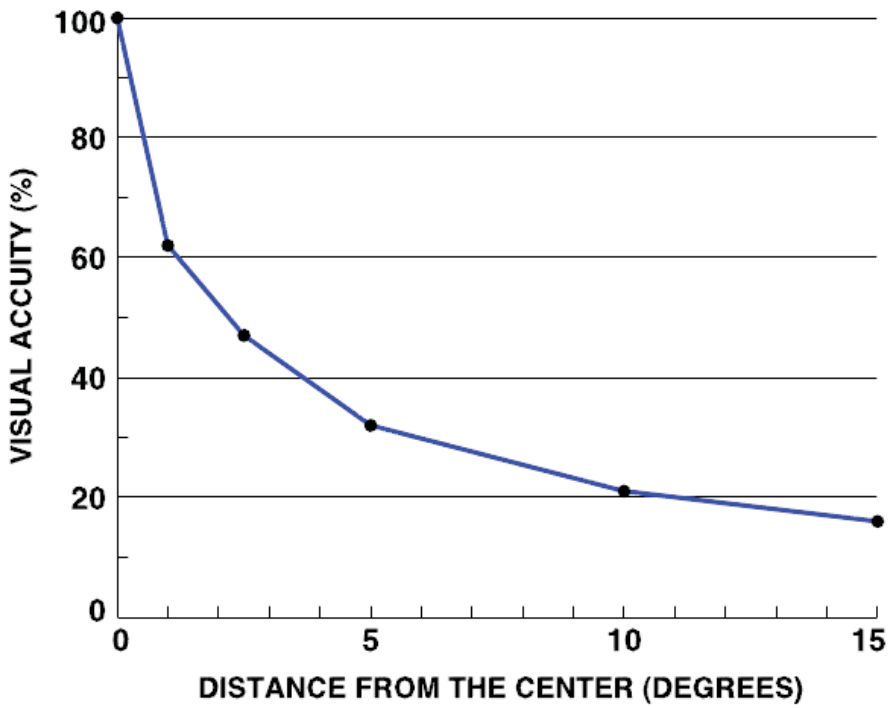


Figure 3.2: The effects of the distance (visual angle) to the center of focused point on visual acuity. This figure shows the degradation of visual acuity at increasing angles from a centered focus point. The visual acuity (expressed as the Snellen fraction [157], *i.e.*, 20/20 equals 100%) quickly declines to approximately 20% at 10 degree eccentricity. (Original figure from [158].)

Therefore, a comprehensive study of effect of the ISD on ERP morphology and BCI performance should be conducted. The effect on ERP morphology should be investigated with increased ISDs.

3.3.2 Symbol Size

Stimulus symbol size (SS) should also have impact on ERP morphology. Symbol with larger-size is more easier to be attended and perceived by subject. This may ease the perceptual sub-task as the first stage of the mental task. Theoretically, this may enlarge P300 amplitude while decrease P300 latency according to the neurophysiological background introduced in section 2.1 of chapter 2. Moreover, larger stimulus may evoke larger visual evoked potential — N1/N200. Therefore, the effect of SS on ERP morphology should also be investigated.

An initial experimental study for investigating ISD and SS effects is conducted. The results are partly support our above discussion. Check subsection 5.2 of chapter 5 for more details.

3.4 The Proposed TLP Matrix Paradigm

Based on the spatial manipulation discussed in the above section and initial experimental results in subsection 5.2 of chapter 5, a two-level predictive (TLP) paradigm is proposed in this section. The TLP paradigm is designed based on two considerations:

1. To elicit better "0/1" signals: The TLP paradigm should be optimized to elicit more discriminable target and non-target ERP signals, compared with the classical Farwell-Donchin paradigm (or RC matrix paradigm). The paradigm optimization is implemented by spatial manipulation discussed in the above section.

2. To encode more information bits: Traditional Farwell-Donchin paradigm has never been optimized based on information theory. However, human communication text can be highly compressed based on statistical language model. In other words, an elaborate BCI paradigm can encode more information bits for each selection. Hence, faster typing speed can be expected.

Before describing the proposed TLP paradigm, we firstly introduce an 8×8 RC paradigm which is used as control paradigm in our studies. The 8×8 RC is depicted in Figure 3.3, which is functioned as a virtual keyboard with sixty four commands. It is a naive extension of the classical 6×6 RC paradigm (see Figure 2.4) that contains more symbols. However, the addition of symbols may further increase inherent spatial distraction[48, 92, 113, 114] since a reduced adjacent spatial interval between symbols increases crowding around the retina fovea. Based on the above discussed spatial manipulation, we favour a matrix with larger inter-symbol distance.

Figure 3.3: A virtual keyboard implemented based on the traditional 8×8 RC paradigm. Sixty four commands are provided by the virtual keyboard, including twenty six English letters, ten arabic numbers, eight punctuation marks, six arithmetic operators, six special characters, and eight functional commands, wherein "←" denotes the delete function. Subjects select any command by gaze attending to it.

3.4.1 Two-Level 3×3 Paradigm: to Elicit Better "0/1" Signals

Inevitably, matrix with larger inter-symbol distance will reduce its size that leads to lower degree of communication freedom (DOCF). A multi-level strategy can be used to compensate the loss of DOCF. In this study, we devised a two-level 3×3 matrix paradigm, see Figure 3.4. Similar multi-level strategy was firstly used in "Hex-o-Spell" speller proposed by Berlin BCI group [159]. The "Hex-o-Spell" speller is driven by motor imagination, and it achieved great success on CeBIT (world's largest IT fair) in 2006. Fazel-Reza *et al.* [115] introduced this strategy into ERP-based BCIs. Later, Berlin BCI group also applied this strategy to their gaze-independent ERP-based BCIs [57, 97].

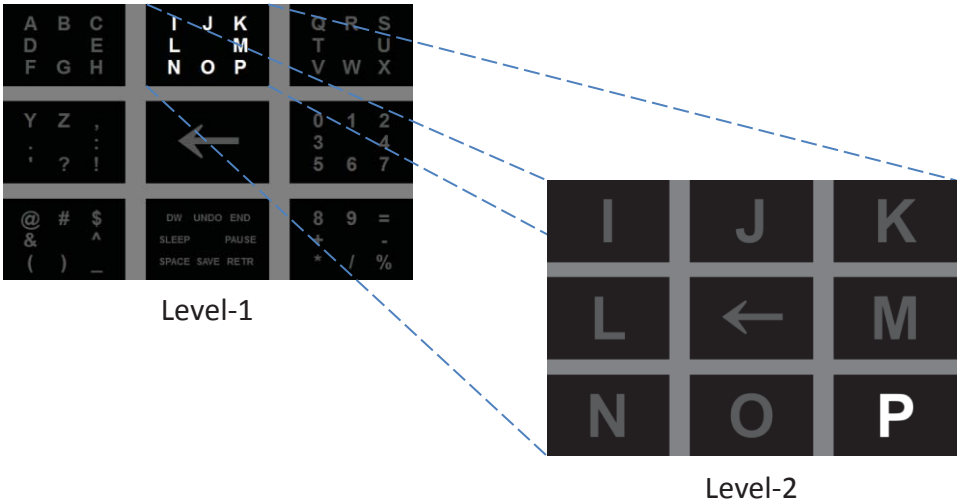


Figure 3.4: Two-level strategy in matrix paradigm. The level-1 matrix is initially presented to subjects. The matrix is consisted of nine cells/groups. Each of the groups contains eight symbols, excluding the middle one which just contains a "←" symbol with delete function (All these sixty four symbols are kept same as symbols in the 8×8 matrix paradigm in Figure 3.3). When subjects select one of the nine cell/groups excluding the middle one, the selected cell/group will be enlarged to replace the level-1 matrix. The enlarged matrix is called level-2 matrix, whose symbols are kept topological consistency in term of spatial distribution. The middle of the level-2 matrix will place a "←" symbol with back to level-1 function (see text for details).

The level-1 matrix is consisted of totally sixty four symbols/commands, in order to keep the DOCF same as the control paradigm — 8×8 RC paradigm (see Figure 3.3). The sixty four symbols are divided into eight cells/groups. The middle cell is filled with a "←" symbol, with deleting function. The level-1 matrix is initially presented to users. In order to select one symbol (or trigger one command) from the sixty four candidates, users should firstly select the group which contains the desired one. Then, the eight symbols contained in the selected group will be used to build the level-2 matrix, wherein the eight symbols are placed topologically consistency with their original placement. In the middle of the level-2 matrix, a "←" symbol is placed and functioned as back-to-level-1, in case a wrong group selected in level-1 matrix. Then, the desired symbol can be directly selected in the level-2 matrix. Note that, deletion can only be done at the level-1.

For the design of level-2 matrix, we enlarged symbol size to ensure that the numbers of white pixels in each grid/group were comparable with level-1. This was expected to minimise the possible differences between VEPs (N1/N200) evoked by each level, preventing unbalanced VEPs contribution to system accuracy between levels. Moreover, larger VEPs are expected with larger stimulus symbols as we discussed in spatial manipulation.

3.4.2 Hypothesis I: 3×3 Matrix Paradigm outperforms 8×8 One

The underlying of the two-level paradigm is the basic 3×3 matrix paradigm. Sellers *et al.* [53] compared a 3×3 RC paradigm with a 6×6 RC paradigm, and they found the P300-speller with a smaller-size matrix had better accuracy; additionally, large-size matrix is difficult for some ALS patients[145].

Moreover, we further enlarge both inter-symbol distance and symbol size for the 3×3 matrix paradigm. Based on the spatial optimization discussed in section 3.3, we propose the first hypothesis that the spatially optimized 3×3 RC paradigm (see

section 5.2.2 of chapter 5 for experimental setup) outperforms the 8×8 one (see section 5.2.2 of chapter 5 for experimental setup), in terms of ERP amplitude, ERP classification accuracy, and information transfer rate.

3.4.3 Single Cell (SC) Pattern versus Row/Column (RC) Pattern

Stimulus pattern has direct relationship with oddball configuration. Two kinds of stimulus patterns are investigated for matrix paradigm in this study. The traditional row/column (RC) stimulus pattern and single cell (SC) pattern. The SC pattern means stimuli in a matrix is presented one by one (instead of by row/column) in a random approach. Guan *et al.* [144] tested the RC and SC patterns on a 6×6 matrix paradigm. Their results showed the 6×6 SC outperformed the 6×6 RC in terms of accuracy. Guger *et al.* [38] conducted a similar comparison. However, they achieved conflicting results against Guan's work.

3.4.4 Hypothesis II: SC Pattern outperforms RC One

The SC and RC stimulus patterns which perform better for a 3×3 matrix paradigm is still unknown. However, SC stimulus pattern has a larger average target-to-target temporal interval (TTI) compared with the conventional RC stimulus pattern when keeping stimulus sequence durations same. Hence, we expect the 3×3 SC to generate larger P300 ERPs than the 3×3 RC. Therefore, we propose the second hypothesis that the 3×3 SC (see section 5.2.2 of chapter 5 for experimental setup) outperforms the 3×3 RC one, in terms of ERP amplitude, ERP classification accuracy, and information transfer rate.

3.4.5 Predictive Step: to Encode More Information Bits

Text typing based on letter-wise selection from a large letter-pool is naive and inefficient, especially for communication based on noisy EEG channels. The large letter-pool could be optimized by word prediction. Therefore, ERP-based BCIs may greatly benefit from predictive technologies. However, very few studies have been reported on incorporating predictive programs into ERP-basd BCIs. Ryan *et al.* [160] firstly combined a commercial predictive program with their checkerboard paradigm, see Figure 3.5. A window is presented alongside the matrix paradigm. During spelling, the window is filled with seven most likely English text phrases based on typed text history. Each of seven predicted text phrases is denoted by a numeric ID. If a desired text phrase contains in the predicted text list. Users can select the desired one by selecting its corresponding numeric ID in the matrix paradigm. Their attempt greatly reduced text spelling time; however, it also significantly reduced spelling accuracy. They attributed the accuracy reduction to potentially increased mental workload and dual-task interference. They thought that users should pay attention to both main matrix and predictive window for each selection. These double tasks will interfere with each other. Moreover, users need to evaluate the predictive text list to find the desired one in a short spelling interval. This may increase cognitive workload, then it may lead to deteriorate ERPs. Kaufmann *et al.* [161] directly integrated several predicted text phrases into the classical 6×6 RC paradigm, see Figure 3.6. According to their results, they reduced spelling time, while they did not hamper spelling accuracy.

However, providing users with the possibility to select predicted words/phrases in an ERP-based BCI is a double-edged sword. It may provide powerful correct selections with time saving. Unfortunately, it may also lead to "powerful" errors with time losses. Moreover, it may introduce high mental workload [160]. Thus, we integrate a conservative letter-by-letter predictive system into the two-level 3×3

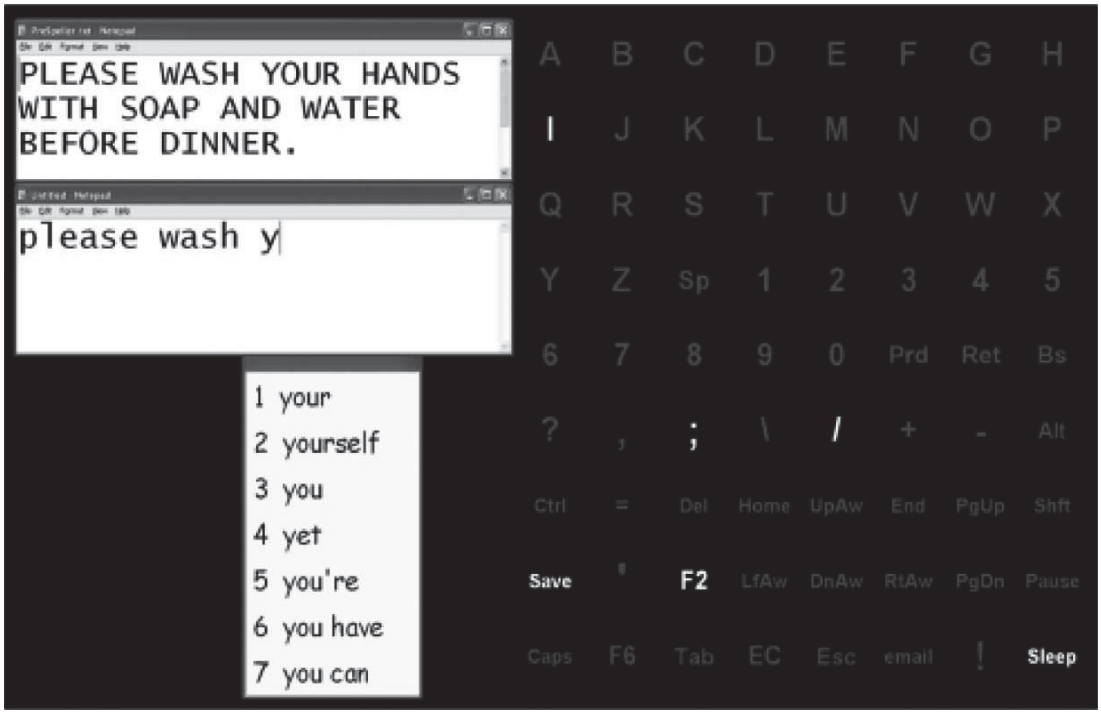


Figure 3.5: Predictive speller proposed by Ryan *et al.* [160]. A window placed at the left bottom of a so-called checkerboard matrix paradigm. The window is filled with seven predicted English text phrases based on typed text history (original figure from [160]).

matrix paradigm. The final paradigm is named two-level predictive (TLP) paradigm, and is depicted in Figure 3.7.

In the original two-level paradigm, each symbol selection needs two steps: selecting a group containing the specific target symbol, and then picking the target symbol from the group. By considering statistical information of a specific language, number of selectable candidates should reduce with increasing typed text history. Hence, desired symbols may be directly selected from a dynamically generated matrix filled with several most likely candidates instead of a redundant two-step selection. In this study, a variant of prediction by partial matching (PPM), called PPMD5 [162], is employed to predict the most likely symbols based on typed text. The top eight

Figure 3.6: Predictive text entry paradigm proposed by Kaufmann *et al.* [161]. A predictive program is intergraded into a 6×6 matrix paradigm. The left column is filled with six German words, which is predicted based on typed text history (original figure from [161]).

symbols are then used to generate a predicted 3×3 matrix (see right sub-figure of Figure 3.7). The selection procedure can be described as follows.

1. *Step-1*:

Initial State (see left sub-figure of Figure 3.7). Each entry contains an 8-element group of symbols, the middle entry "←" is a deletion command. Group selection leads to *Step-2*.

2. *Step-2*:

Subgroup matrix appears (see middle sub-figure of Figure 3.7), with "←" in the middle as a return function to *Step-1*. Symbol selection enqueues the symbol to the

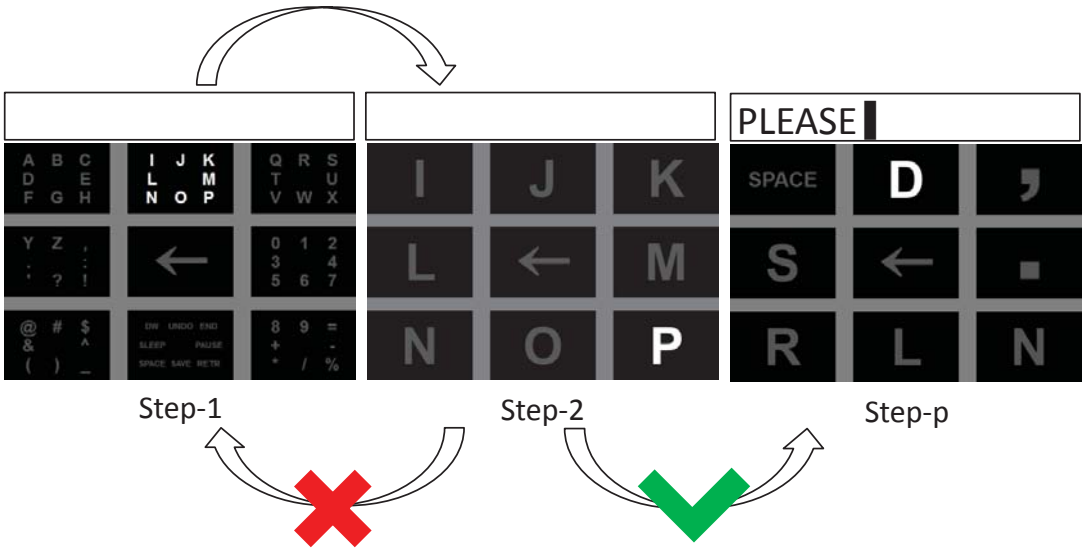


Figure 3.7: An example for describing predictive step (*Step-p*) in the TLP paradigm. The proposed TLP paradigm is built by integrating a predictive step (*Step-p*) into the two-level 3×3 matrix paradigm. We rename level-1 as *Step-1*, and level-2 as *Step-2*. In the original two-level setup, finishing selection in *Step-2* will automatically lead back to *Step-1* for next spelling. In the TLP paradigm, *Step-p* will be presented instead of going back to *Step-1*, in order to reduce two-step selection redundancy (see text for details).

typed history queue. Then, given the typed history, PPMD5 predicts symbols with the top eight conditional probabilities. If the sum of the eight probabilities exceeds a predefined threshold, the prediction process proceeds to *Step-p*, otherwise a return to *Step-1* occurs. (Note that PPMD5 in this study used a limited 33-symbol set: 26 English letters, 6 punctuation marks (“,”, “.”, “?”, “!”, “”’, “:”), and a space symbol. “Illegal” symbol (symbols in group-6 (row-wise), -7, -8, 9 of *Step-1*) selection will directly lead back to *Step-1*).

3. *Step-p*:

Predicted matrix (see right sub-figure of Figure 3.7) is filled with the eight symbols with probability descending from left to right and top to bottom (*e.g.*, in Figure 3.7, given “PLEASE” as typed history, space is most likely to be selected, “D” is second, “,” is third, and so on). The symbol “←” in the middle is a return function to *Step-1*

in case a desired symbol does not exist (prediction failure). Selection logic is the same as *Step-2*.

The threshold used to determine whether leading to *Step-p* or going back *Step-1* is set to zero, *i.e.*, invalid. Spelling simulation (see Figure 3.8) with 100% accuracy based on seven sample texts by cross-validation (some texts as spelling text and remaining as PPMD5 training data) showed that lowest average selection step (< 1.3) was achieved when the threshold was set to zero. This also confirms eight is a proper size for the predicted matrix. The seven sample texts are chosen from various of themes (including technology, economy, politics, sports, and novels), in order to reduce the possible training sample bias. Note that simulation on larger-size training samples (*e.g.*, Emma (86.4 KB) and The Moonstone (1.02 MB)) achieved smaller average step (near 1.1 steps), which means that the prediction ability of the PPMD5 model can be further improved by larger training samples.

Because we are majorally focusing on stimulus paradigm itself, the zero-threshold logic for proceeding to *Step-p* is naive. However, better logic control optimized by deep theories from other disciplines such as natural language processing (NLP) may push the TLP paradigm onto a higher-level stage. For example, the threshold may be optimized by real experiment data instead of simulation used here. Perhaps, it is impractical to beforehand collect large amount of real BCI-spelling data to specifically optimize the threshold, because BCI-spelling is quite time-consuming. Potential users may reject a BCI system which needs long time for preparation. However, we could apply a dynamic strategy that silently collects real spelling data during users' daily using of the BCI system. Then, the threshold could be dynamically optimized based on currently collected data. The BCI system can be dynamically customized to its user.

Multi-threshold strategy may be better than the single-threshold one. Besides checking the summation of top-8 candidates' probabilities, we can further evaluate the "significance" of the top-1 candidate's probabilities. That could be ratio of the

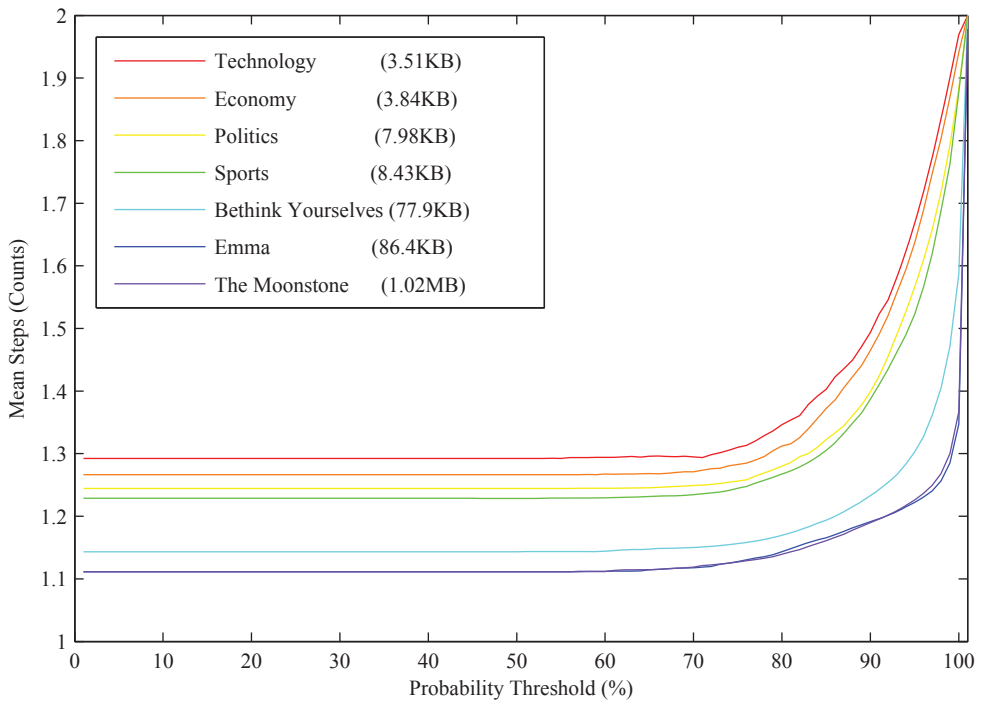


Figure 3.8: Spelling simulation with 100% accuracy based on seven sample texts (with various themes including technology, economy, politics, sports, and novels) by cross-validation (some as spelling text and remaining as PPMD5 training data) showed that lowest average selection steps (< 1.3) was achieved when the threshold was set to zero. This also confirms eight is a proper size for the predicted matrix. Note that simulation on larger-size sample texts achieved smaller average steps, which means the prediction ability of the PPMD5 model can be further improved by more training samples.

top-1's probability to the total sum or other measures. In case the first sum-threshold is way "soft", the second ratio-threshold may be guard "hard".

We can also monitor the "uncertainty" during BCI-spelling. If consecutive high sum of top-8 candidates' probabilities are obtained, the user may be spelling an English paragraph. Suddenly, a quite low sum obtained, the user may switch to type Japanese. Or maybe the user typed something wrong. The uncertainty detection, perhaps, can improve the usage of the BCI system. Moreover, similarly techniques used in conventional augmentative and alternative communication (AAC) devices can be utilized here.

3.4.6 Hypothesis III: No Significant Difference between *Step-1*, *-2*, and *-p*

Identifying a desired symbol from matrices of *Step-1* and *Step-2* are very simple, because the layout of both matrices of *Step-1* and *Step-2* are constant. However, same task may be cognitively harder for *Step-p*.

Word prediction is a double-edged sword for augmentative and alternative communication (AAC) systems, especially for BCI systems. On the one hand, it may provide powerful correct selections with time saving; on the other hand, it may also increase user's cognitive demands leads to lower communication accuracy [163, 164]. In this study, however, we applied a conservative letter-by-letter prediction model. More importantly, the eight predicted symbols are placed in a meaningful order — probability descending from left to right and top to bottom. We suppose that identifying a symbol from an eight-size symbol group with prior ordering information is not a difficult task for human subjects, even though the identification task should be finished in a short spelling interval (generally 3–4 seconds for ERP-based BCIs).

Therefore, we propose the third hypothesis in this study that there is no significant difference between ERPs morphologies and ERPs classification accuracies of *Step-1*, *Step-2*, and *Step-p* in the TLP paradigm, since tasks for each steps are with similar cognitive difficulties.

3.4.7 Hypothesis IV: TLP Paradigm outperforms Classical RC paradigm

Compared with the control paradigm in this study — the classical 8×8 RC paradigm, the proposed 3×3 TLP paradigm is optimized by spatial manipulation to elicit more discriminable ERPs, and further optimized by integrating a statistical language model to reduce encoding redundancy. Hence, we make our fourth hypothesis that the 3×3

TLP paradigm outperforms the classical 8×8 Farwell-Donchin paradigm, in terms of selecting accuracy, information transfer rate, and speed.

3.5 Summary

Based on the neurophysiological background of ERPs and human perception, we optimized the classical Farwell-Donchin paradigm by spatial manipulation on the inter-symbol distance and symbol size. The devised 3×3 matrix paradigm is hypothesised to elicit more discriminable ERPs compared with the control paradigm — classical 8×8 Farwell-Donchin paradigm. Furthermore, the 3×3 TLP paradigm, which combines a two-level 3×3 matrix paradigm with a statistical language model, was proposed based on information theory. The TLP paradigm is supposed to accelerate BCI communication speed. More importantly, although prediction technology is used in the TLP paradigm, we hypothesis that there is no significant performance deficit although heavier cognitive demanding may be introduced. All these hypotheses will be validated through designated experiments in chapter 5.

Chapter 4

THEME II: Trial-based ERP Classification with Natural Language Processing

This chapter discusses THEME II of this dissertation, which is about the proposed trial-based ERP classification algorithm. Main contributions of the chapter are: 1. we propose a more reasonable ERP model; 2. Ensemble of epochs within-trial strategy is proposed and implemented in maximum a posterior (MAP) approach.

First, we state straightforward trial-based approach for ERP-decoding. However, this method has never been successfully applied in any practical system. This should be attribute to "curse of dimension" issue of multi-class classification. Then, we elaborate current state-of-the-art epoch-based approach including their theoretical models. However, this approach suffers from low signal-to-noise ratio (SNR) of EEG signals and an "overlap" effect. Learnt their strengths and avoided their weaknesses, we propose a trial-based ERP classification method. The method is built on a more reasonable multi-class ERP model. It integrates all ERP-epochs within-trial and fuses natural language information to achieve a better decoding performance.

4.1 Motivation: to Exploit Temporal Information in ERP Trials and Fuse Language Knowledge to Improve ERP Classification

Since 1988 when Farwell and Donchin proposed the first P300-speller, ERP decoding based on binary ERP-epoch classification has been dominating the field for more than 20 years, as we stated in section 2.2 of chapter 2. However, the ERP decoding is virtually a multi-class ERP-trial classification. At present, a whole ERP-trial is divided into several ERP-epochs, and then the ERP decoding is cast to multiple binary ERP-epoch classifications that determine the presence or absence of special ERP components (*e.g.*, P300, N1/N200, etc.) in each ERP-epoch of the ERP-trial.

Unfortunately, such epoch-based strategy achieves poor information transfer rate as described in section 2.4 of chapter 2. This is majorally attributed to low binary discriminability of single-trial ERP-epochs. Current countermeasure is ensemble of multi-trial ERP-epochs to enhance signal-to-noise ratio of ERPs, however it hampers speed. Therefore, improving ERP decoding accuracy based on single-trial ERP-epochs is particularly important for practicability of the ERP-based BCIs.

Almost all kinds of modern binary classification algorithms have been exploited to classify ERP-epochs; however, no particular classification algorithm is "the best" for all problems. Essentially, the discriminative information contained in the single-trial ERP-epoch is not reliable. A pragmatic approach proposed in this chapter is to integrate all ERP-epochs in an ERP-trial, utilizing temporal relationships among these ERP-epochs. The integration within ERP-trial is expected to enhance signal-to-noise ratio of ERPs. Furthermore, prior language knowledge is also fused to improve text spelling performance.

4.2 Features for ERP Classification

In machine learning and statistics, classification is the problem of identifying to which of a set of categories a new observation belongs. The individual observations should be quantified as features. Theoretically, the features should be extracted in order to maximize statistically separability of individual observations. In the current epoch-based ERP classification domain, the individual observations are the ERP epochs elicited by stimuli (target and non-target stimuli). The ERP classification is just to determine whether any given ERP epoch is elicited by target stimulus or non-target stimulus. Based on the chapter 2, ERP epochs elicited by target or non-target stimulus contain different ERP components. Target ERP-epoch contains P300 and/or VEP (N1/N200) components, whereas ideal non-target ERP-epoch does not. ERP components are characterized by their temporal evolution and the corresponding spatial distributions. Therefore, the mainstream of ERP classification are based on the spatio-temporal feature.

4.2.1 Spatio-temporal Feature

As denoted in the chapter 2, raw ERP epochs are given in a form of two dimensional matrix $\mathbf{X} \in \mathbb{R}^{M \times T}$, with M being the number of EEG channels, T being the number of sampled time points in an epoch. For the formalism in classification, features are considered to be column vectors. The matrix $\mathbf{X}$ can be concatenated to form a column vector $\mathbf{x} \in \mathbb{R}^{M \cdot T}$. The vector $\mathbf{x}$ can be represented by:

$$\mathbf{x} = [x_{c_1}(t_1), \ldots, x_{c_i}(t_j), \ldots, x_{c_M}(t_T)] \tag{4.1}$$

where $x_{c_i}(t_j)$ denotes the signal point sampled at time t_j over channel c_i. In order to cover all potential ERP components, the dimension $(M \cdot T)$ of raw ERP column vector $\mathbf{x}$ is relatively very high compared to the amount of available training samples. This makes very difficult for supervised learning. Therefore, how to select feature

points from spatio-temporal domain is not trivial. There are generally two kinds of approaches: manually selecting based on separability measures and automatically selection based on statistical procedures. Berlin BCI group manually selected feature point based on correlation coefficient (r-value) [26]. While, researchers from Wadsworth Center, New York State Department of Health preferred an automatic feature selection method – stepwisefit [25]. The stepwisefit will be described in next section.

4.2.2 Other Features

Besides the spatio-temporal (first-order) feature, some researchers have tried to exploit other higher-order features. Mak *et al.* [165] found that theta band ($4.5-8Hz$) power has positive correlation with P300-based BCI performance in ALS patients. However, there is still no successful ERP-based BCI driven by theta power. The spatio-temporal feature still dominates the field. Perhaps, theta power feature may be combined with the spatio-temporal feature to further improve ERP-BCI performance.

4.3 Straightforward Trial-based ERP Classification

From now on, we focus on ERP decoding for our proposed 3×3 two-level predictive (TLP) paradigm. A complete stimulus-trial of the 3×3 TLP is that each of the nine (3×3) stimuli occurs one time and exactly one time. During a stimulus-trial, a user should only attend to one of the nine stimuli, while ignoring others. Meanwhile, an ERP-trial is recorded from the user. In theory, we could know which stimulus the user focus on by analysing the ERP-trial.

Theoretically, each of the nine stimuli could be the target which the user focus on. For each of the nine conditions, there are 362,880 (9! permutations) kinds of

stimulus-trials. Therefore, there could be totally 3,265,920 ($9 \times 9!$) kinds of ERP-trials. We could build 3,265,920 ERP-models in advance. Then, we could decode a new coming ERP-trial by fitting it with these millions of ERP-models. However, this approach is impractical, since it will cost several months for collecting only one sample for each of these millions of ERP-trials.

In practice, the 3,265,920 kinds of ERP-trials could be reduced to only 9 classes. That means an ERP-trial is only characterized by temporal location of the ERP components (P300, N1/N200, etc.). There are totally nine possible temporal locations for the evoked ERP components (see Figure 4.1). This practical simplification assumes any target-stimulus elicits statistically same target ERP components, while any nontarget-stimulus elicits statistically same nontarget ERP components (ideally, zero potentials). As shown in Figure 4.1, the ERP decoding is cast to a 9-class classification problem.

In machine learning, there are generally two popular approaches to construct k-class classifiers based on binary classifiers. One is so-called "One vs. One" (OvO) approach. This approach involves the construction of $k(k-1)/2$ binary classifiers where each one is trained on data from two classes. The final result is determined by voting from all these binary ones. The other is "One vs. Rest" (OvR) approach. Here, k binary classifiers are constructed: class 1 (positive) versus rest classes (negative), class 2 versus rest classes, ..., class k versus rest classes. The combined k binary classifiers predict the class of a test sample that corresponds to the maximal value of k binary ones.

Support vector machine (SVM) is selected to construct the 9-class ERP classifiers. No matter OvO or OvR approach, multiple binary ERP classifiers should be trained. The decision function of the k-th classifier, $f_k(\mathbf{x}) = \mathbf{w}_k^T\mathbf{x} + b_k$, where $\mathbf{w}_k$ is the k-th

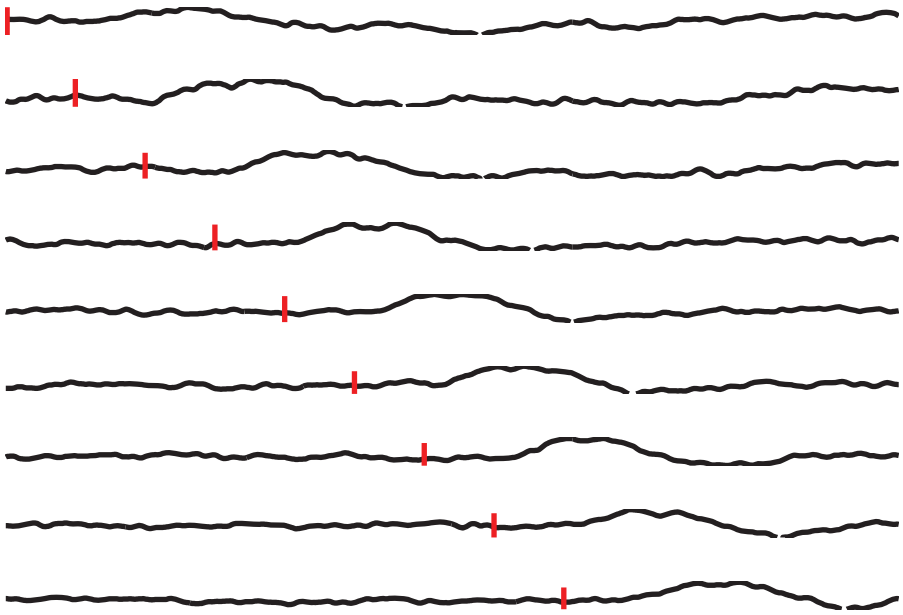


Figure 4.1: A typical example of 9-class ERP-trials (single channel) for the 3×3 TLP. The 9 red lines indicate the 9-class target-stimulus occurrences, respectively. A typical P300 peak is presented about 300 milliseconds after each of 9-class target-stimulus onsets. Note that each of the nine ERP-trials is averaged over hundreds of trials.

weight vector, can be obtained by solving an optimization problem:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\mathbf{w}_k\|^2 + C\Sigma_{i=1}^{n}\xi_i^k, \\ \text{subject to} \quad & y_i^k(\mathbf{w}_k^T\mathbf{x}_i + b_k) \geq 1 - \xi_i^k, \\ & \xi_i^k \geq 0, i = 1, \ldots, n. \end{aligned} \tag{4.2}$$

where y_i^k is the class label for the k-th classifier, and $\mathbf{x}$ denotes ERP-trial.

4.3.1 Drawbacks of Trial-based Approach

The 9-class classification on raw ERP-trial $\mathbf{x}$ is definitely a viable option. However, the dimensionality problem is expected to become more serious when dealing with

multi-class problem, since these are intrinsically more difficult than the binary ones, because the multi-class classification algorithm has to learn to construct a greater number of separation boundaries or relations [166]. Therefore, feature dimension reduction should be helpful. A technique named multi-class SVM with recursive feature elimination (MSVM-RFE) is proposed for multi-class gene selection on DNA microarray data [167]. The MSVM-RFE is described as below:

Algorithm 1 Multi-class SVM-RFE

Input:
$\mathbf{x}_i \in \mathbb{R}^{M \cdot T}, i = 1, \ldots, n$; // training sample vectors
$y_i \in \{1, \ldots, 9\}, i = 1, \ldots, n$; // corresponding class labels
Output:
$S_{\mathbf{w}}$; // selected features set
1: Initialize: $S_{\mathbf{w}} := \emptyset, P_{max} := 0$, $S :=$ full feature set, $f :=$ size of S;
2: **while** $f \neq 0$ **do**
3: Train a multi-class SVM with the features in S;
4: Calculate cross-validation accuracy P;
5: **if** $P > P_{max}$ **then**
6: $P_{max} := P; S_{\mathbf{w}} := S$;
7: **end if**
8: Obtain trained weight vectors $\mathbf{w}_k = [w_{k1}, \ldots, w_{kf}]^T$; // the number of k is determined by implementation (One vs. One/One vs. Rest)
9: Calculate the ranking criteria for each feature in S: $c_i = \sum_k w_{ki}^2, i = 1, \ldots, f$;
10: $\hat{i} := \arg\min_i c_i, \forall i \in S$;
11: $S := S \setminus \hat{i}$; // backward elimination
12: $f :=$ size of S;
13: **end while**
14: **return** $S_{\mathbf{w}}$

The MSVM-RFE selects features in a sequential backward elimination manner, which starts with full feature set. It recursively eliminates one feature at a time. Each time, the least informative feature is discarded. In the feature elimination procedure, the feature set corresponding to the maximum cross-validation accuracy is selected as optimal feature set.

The straightforward approach is expected to suffer from curse of dimension, since the dimension of the ERP-trial is relatively high compared to limited amount of

training samples. Although feature dimension reduction techniques may alleviate the problem to some extent, the straightforward approach has never been applied by any ERP-based BCIs. For comparison, both 9-class SVM on raw ERP-trial and ERP-trial optimized by MSVM-RFE are used as control algorithms in this study.

4.4 State-of-the-Art Epoch-based ERP Classification

As explained in the last section, the straightforward trial-based multi-class classification is intrinsically more difficult. Currently, the dominating ERP decoding for the ERP-based BCIs relies on binary classification on ERP-epochs. For the proposed 3×3 TLP paradigm, the epoch-based approach may divide whole ERP-trial $\mathbf{x}$ into nine ERP-epochs $\mathbf{x}_i$, $i = 1, \ldots, 9$. Each of the ERP-epochs covers ERP components elicited by each stimulus in a stimulus-trial. ERP-epochs corresponding to target stimuli are named "Target" epochs, while others are "Non-target" epochs. Target and non-target ERP-epochs are assumed to be statistically different and could be classified. Binary classification on an ERP epoch $\mathbf{x}_i$ is formalized as below:

$$f(\mathbf{x}_i) = \mathbf{w}^T K(\mathbf{x}_i) + b = \begin{cases} > 0, \mathbf{x}_i : \textit{Target epoch} \\ < 0, \mathbf{x}_i : \textit{Non-target epoch} \end{cases} \tag{4.3}$$

where $K(\cdot)$ denotes a kernel function. Final decoding is determined by finding the maximum decision output: $\hat{i} = \arg\max_i f(\mathbf{x}_i)$, $i = 1, \ldots, 9$.

The epoch-based approach has been dominating since 1988 when Farwell and Donchin proposed the first P300-speller [24]. Because it casts the multi-class modeling problem to an easier binary one. Moreover, it reduces feature dimension by dividing long ERP-trial into short ERP-epochs. Almost all modern supervised binary classification algorithms have been exploited to classify ERP-epochs. A

comprehensive review will be conducted in this section, and a summary is presented in Table 4.1.

Table 4.1: Supervised learning methods exploited in ERP-based BCIs

Linear	Linear Discriminant Analysis (LDA) [29, 168–170] Linear Support Vector Machine (LSVM) [31, 171] Bayesian LDA (BLDA) [172] Regularized/shrinkage LDA (rLDA) [26, 173] Logistic Regression (LR) [174] Pearson's Correlation Method (PCM) [25] Stepwise LDA (SWLDA) [25, 29, 168, 175]
Non-linear	Kernel SVM (kSVM) [29, 168] Artificial Neural Network (ANN) [176]

4.4.1 Conventional Binary-class ERP Model

Before we starting review of binary classification on ERP-epochs, a conventional ERP-epoch model should be stated firstly. Nunez and Srinivasan [177] presented a linear model of EEG generation. As shown in Figure 4.2, the simplified model assumes the brain tissue to be resistive medium and hence only considers effects of volume conduction.

Given a time t, scalp electrode EEG signals $\mathbf{x}(t)$ are assumed to be linear superposition of current sources $\mathbf{s}(t)$ inside brain tissues:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{4.4}$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^T$, with M being the number of scalp electrodes. $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_N(t)]^T$, with N being the number of current sources inside brain tissues. $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the constant propagation matrix.

The instantaneous model can be easily extended for an ERP temporal epoch $\mathbf{X} = \begin{bmatrix} \mathbf{x}(t_1) & \mathbf{x}(t_2) & \ldots & \mathbf{x}(t_T) \end{bmatrix} \in \mathbb{R}^{M \times T}$, with T being the number of sampled time

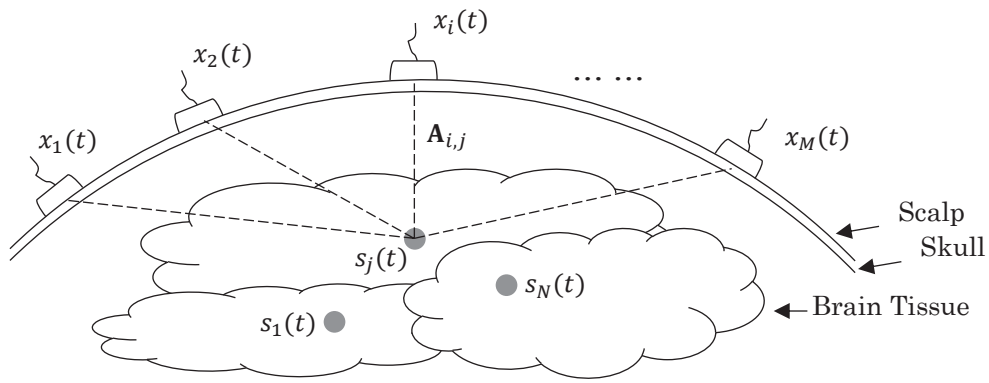


Figure 4.2: Basic macroscopic model of EEG generation [177]. The simplified model assumes the brain tissue to be resistive medium and hence only considers effects of volume conduction. At a given time t, a single current source $s_j(t)$ contributes linearly to the scalp potential $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^T$, with M being the number of scalp electrodes.

points:

$$\begin{bmatrix} \mathbf{x}(t_1) \\ \vdots \\ \mathbf{x}(t_T) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{s}(t_1) \\ \vdots \\ \mathbf{s}(t_T) \end{bmatrix} \tag{4.5}$$

where ERP-epoch $\mathbf{X}$ is represented as column vector. Apparently, this approximately linear mixing process can (in principle) be inverted by a linear transformation of the spatio-temporal EEG sensor measurements. An important consequence is that only linear methods are necessary in the pre-processing and classification. This may be the essential reason that linear methods have been dominating the ERP classification field.

Blankertz *et al.* [26] further evolved the model to a well-fitted Gaussian model:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{4.6}$$

For a given t, ERPs measured from scalp electrodes $\mathbf{x}(t)$ are the superposition of pure ERP pattern (time-locked) $\mathbf{As}(t)$ and background noise $\mathbf{n}(\mathrm{t})$, which is assumed to be identically independent distributed (iid) according to a Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma_n})$. Here, neural background activities and various artifacts are modeled as a Gaussian distribution with a zero expectation.

Again, we extend instantaneous model for an ERP-epoch with sampling time $t_1, \ldots, t_T$, as below:

$$\begin{bmatrix} \mathbf{x}(t_1) \\ \vdots \\ \mathbf{x}(t_T) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{s}(t_1) \\ \vdots \\ \mathbf{s}(t_T) \end{bmatrix} + \begin{bmatrix} \mathbf{n}(t_1) \\ \vdots \\ \mathbf{n}(t_T) \end{bmatrix} \tag{4.7}$$

With this extension, an ERP-epoch recorded from M channels with T sampling points can be regarded as linear superposition of spatio-temporal ERP components and Gaussian noise. Consequently, the target and non-target ERP-epochs fulfill Gaussian distributions $\mathcal{N}(\mathbb{A}\mathbf{s}, \mathbf{\Sigma}_{\mathbb{n}})$ and $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbb{n}})$, respectively:

$$\begin{aligned} &\textit{Target epoch}: && \mathbf{x}_+ = \boldsymbol{\mu}_+ + \mathbb{n} = \mathbb{A}\mathbf{s} + \mathbb{n} && \sim\mathcal{N}(\mathbb{A}\mathbf{s}, \mathbf{\Sigma}_{\mathbb{n}}) \\ &\textit{Non-target epoch}: && \mathbf{x}_- = \boldsymbol{\mu}_- + \mathbb{n} = \mathbf{0} + \mathbb{n} && \sim\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbb{n}}) \end{aligned} \tag{4.8}$$

where, $\mathbf{x} = \begin{bmatrix} \mathbf{x}(t_1) \\ \vdots \\ \mathbf{x}(t_T) \end{bmatrix} \in \mathbb{R}^{M \cdot T}$ denotes column vector form of ERP-epoch.

$\mathbb{A} = \begin{bmatrix} \mathbf{A} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{A} \end{bmatrix} \in \mathbb{R}^{(M \cdot T) \times (N \cdot T)}$ denotes linear spatio-temporal mixing matrix.

$\mathbb{n} = \begin{bmatrix} \mathbf{n}(t_1) \\ \vdots \\ \mathbf{n}(t_T) \end{bmatrix} \in \mathbb{R}^{M \cdot T} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbb{n}})$ denotes the background noise.

This means that target and non-target ERP-epochs are Gaussian distributed (visualized in Figure 4.3) with same covariance being the covariance of ongoing background EEG, and with mean being the spatio-temporal locked ERP components and zero-potential respectively.

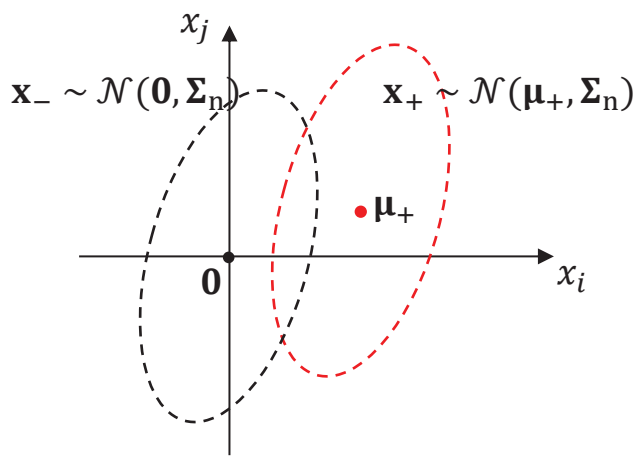


Figure 4.3: Binary-class Gaussian distributions of ERP-epochs. Means of target (red) and non-target (black) distributions are marked by dots. Classwise covariances are visualized as ellipses.

Although the spatio-temporal pattern of ERP components (P300 and N1/N200) $\boldsymbol{\mu}_+$ is actually variable since various factors may affect them (see section 2.1 of chapter 2), the variance of ERP components is relatively small compared to background noise variance. Therefore, the binary-class Gaussian model of ERP-epochs is practically rational. Based on the model, various linear and non-linear binary classifiers have been exploited to classify ERP-epochs.

4.4.2 Linear Classifiers

A linear Classifier makes a classification decision based on the value of a linear combination of the feature vector. Different linear classifiers differ in how to obtain linear coefficients.

Normalized Cross Correlation (NCC)

Strictly speaking, normalized cross correlation (NCC) is not a classifier. It is a measure of similarity of two individual observations. It was originally introduced for "classifying" the P300 in [24]. A target-epoch template $\mathbf{x}_T$ should be firstly prepared. Then, the likelihood of any new coming ERP-epoch $\mathbf{x}_i$ to be target-epoch can be calculated as below:

$$NCC(\mathbf{x}_i, \mathbf{x}_T) = \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T(\mathbf{x}_T - \bar{\mathbf{x}}_T)}{\|\mathbf{x_i} - \bar{\mathbf{x}}_\mathbf{i}\|\|\mathbf{x_T} - \bar{\mathbf{x}}_\mathbf{T}\|} \tag{4.9}$$

where $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_T$ are the mean vectors of $\mathbf{x}_i$ and $\mathbf{x}_T$, respectively. $\|\cdot\|$ is the L^2 norm. Therefore, the target is determined as $\hat{i} = \arg\max_i NCC(\mathbf{x}_i, \mathbf{x}_T)$, given a trial of ERP-epoch $\mathbf{x}_i$, $i = 1, \ldots, 9$.

The NCC is very simple and easy to be implemented; however, its performance is generally low since covariance information has never been exploited.

Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) is the benchmark method for determining the optimal separating hyperplane between two classes. For binary classification between two Gaussian distributed classes with equal covariance, LDA, fisher's linear discriminant (FLD), and the ordinary least-squares regression (LSR) solution are equivalent, with the estimated feature weights $\mathbf{w}$ (in equation 4.3) given as:

$$\mathbf{w} = \mathbf{\Sigma}_{\mathbb{n}}^{-1}(\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-) \tag{4.10}$$

where $\mathbf{\Sigma}_{\mathbb{n}}$ is covariance matrix of ERP-epoch samples, $\boldsymbol{\mu}_+$ and $\boldsymbol{\mu}_-$ are the means of the target (positive) and non-target (negative) ERP-epoch samples.

In practice, the covariance matrix $\mathbf{\Sigma}_{\mathbb{n}}$ is estimated based on labeled training data. However, the number of training data/samples is very limited compared with the dimension of the ERP-epochs. Because training data are collected from a so-called

calibration session which generally lasts less than 30 minutes by considering system practicability and possible user resistance. For instance, in this study, we recorded about 10-minute 3×3 TLP calibration data including 450 target (+) samples and 3600 non-target (−) samples. The dimension of the ERP-epoch is 144 ($9(channel) \times 0.8(seconds) \times 20(Hz)$). The number of unknown parameters in $\mathbf{\Sigma}_{\mathbb{n}}$ that have to be estimated is quadratic to the number of dimensions. Imprecise estimated covariance matrix may be not invertible. Several methods have been proposed to fix LDA in this case, such as regularized LDA (rLDA) [26, 173], stepwise LDA (SWLDA) [25, 29, 168, 175], etc.

Regularized LDA

Regularized LDA (rLDA) is simply to fix $\mathbf{\Sigma}_{\mathbb{n}}$ to make it invertible by adding a fixed constant "ridge" to its diagonal entries, *i.e.*, $\mathbf{\Sigma}_{\lambda} = \mathbf{\Sigma}_{\mathbb{n}} + \lambda\mathbf{I}$, where λ is the ridge magnitude. Berlin BCI group introduced a shrinkage technique to regularise LDA [26], and they showed the shrinkage-LDA performed no worse than SWLDA.

Stepwise LDA

Stepwise LDA (SWLDA) uses a stepwisefit procedure to reduce feature dimension. The stepwisefit procedure performs feature space reduction by selecting suitable features in a forward and backward stepwise analysis. The input features are weighted using ordinary least squares regression to predict the target class labels. Starting with no initial features, the most statistically significant input feature for predicting the target label is added. After each new entry, a backward stepwise analysis is performed to remove the least significant input feature. This process is repeated until a predetermined number of features selected, or until no additional features satisfy the entry/removal criteria. The algorithm is described as below:

The SWLDA is the state of the art method for ERP-epoch classification. In part this popularity is based on the comparison of different ERP-epoch classification

Algorithm 2 Stepwisefit for feature selection

Input:

$\mathbf{x}_i \in \mathbb{R}^{M \cdot T}, i = 1, \ldots, n$; // training sample vectors

$y_i \in \{-1, +1\}, i = 1, \ldots, n$; // corresponding labels (-1: non-target, +1: target)

F_{max}; // max amount of features to be selected

p_{in}, p_{out}; // p-value thresholds for selecting features in and out

Output:

$S_{\mathbf{w}}$; // selected features set

1: Initialize: $\mathbf{w} := \mathbf{0}, b := 0, S_{\mathbf{w}} := \emptyset, P_{\mathbf{w}} := \mathbf{0}$; // $\mathbf{w}$ and b to fit $y = \mathbf{w}^T\mathbf{x} + b$, p-values for $\mathbf{w}$

2: **while** size of $S_{\mathbf{w}} < F_{max}$ **do**

3: conduct the below Least Square Regression:

4: $\varepsilon = \sum_{i=1}^{n}[y_i - (\mathbf{w}(k_1)\mathbf{x}_i(k_1) + \ldots + \mathbf{w}(k_s)\mathbf{x}_i(k_s) + b)]^2, S_{\mathbf{w}} = \{k_1, \ldots, k_s\}$

5: record regression significance p-value to $P_{\mathbf{w}}(k), k \in S_{\mathbf{w}}$;

6: **for each** $l \notin S_{\mathbf{w}}$ **do**

7: conduct the below Least Square Regression:

8: $\varepsilon = \sum_{i=1}^{n}[y_i - (\mathbf{w}(k_1)\mathbf{x}_i(k_1) + \ldots + \mathbf{w}(l)\mathbf{x}_i(l) + b)]^2$

9: record regression significance p-value to $P_{\mathbf{w}}(l)$;

10: **end for**

11: $\hat{l} := \arg\min_l P_{\mathbf{w}}(l), \forall l \notin S_{\mathbf{w}}$

12: **if** $P_{\mathbf{w}}(\hat{l}) < p_{in}$ **then**

13: $S_{\mathbf{w}} := S_{\mathbf{w}} \cup \{\hat{l}\}$; // forward selecting feature

14: **else**

15: $\hat{k} := \arg\max_k P_{\mathbf{w}}(k), \forall k \in S_{\mathbf{w}}$

16: **if** $P_{\mathbf{w}}(\hat{k}) > p_{out}$ **then**

17: $S_{\mathbf{w}} := S_{\mathbf{w}} \setminus \hat{k}$; // backward dropping feature

18: **else**

19: goto 23

20: **end if**

21: **end if**

22: **end while**

23: **return** $S_{\mathbf{w}}$

methods by Krusienski and colleagues from Wadsworth Center [29] which concluded that SWLDA had the greatest potential. Moreover, SWLDA is widely spread with the open-source BCI2000 [178] (used by about 600 laboratories, as to writing the dissertation) provided by Wadsworth Center of the New York State Department of Healthy, in collaboration with the University of Tübingen in Germany. Therefore, the SWLDA is used as standard classifier for comparing different stimulus paradigms,

and is used as one of the control methods for comparing with our proposed ERP classification method.

4.4.3 Non-linear Classifiers

Several comparison studies have indicated that non-linear classifiers (*e.g.*, artificial neural network (ANN) [176], kernel-based support vector machine [29, 168]) performed no better (or even worse) than linear ones for ERP-epoch classification [29, 169, 171]. This could be attribute to the tendency of non-linear classifiers to overfit the training data, leading to an inferior generalization performance.

In this study, we choose SVM with Gaussian kernel as one of the control method for comparing with our proposed ERP classification method, because Gaussian kernel has universal approximation properties. The Gaussian kernel is given as follows:

$$K(\mathbf{u}, \mathbf{v}) = e^{-\frac{\|\mathbf{u}-\mathbf{v}\|^2}{2\sigma^2}} \tag{4.11}$$

The kernel function ($K(\cdot)$ in equation 4.3) is used to map features to a higher-dimensional feature space. Then, SVM is used to determine the hyper-plane that maximizes the separating margin between two classes of a binary classification.

4.4.4 Drawbacks of Epoch-based Approach

In contrast to the straightforward trial-based approach, the epoch-based one predicts whether a stimulus is target or not by analysing the ERP-epoch elicited by the stimulus. Indeed, the feature dimension of the ERP-epoch is decreased compared to the whole ERP-trial; however, part of meaningful information is also lost. That means binary classification on single-trial ERP-epochs is not reliable. Moreover, it may also suffer from a so-called "overlap" effect, which was described in chapter 2 and will be reviewed again in next section.

4.5 The Proposed Trial-based ERP Classification

Learnt strengths and weaknesses of conventional methods, we propose a trial-based ERP classification fusing statistical language information. Similar with state-of-the-art epoch-based method, we divide a whole ERP-trial into several ERP-epochs. Under a conditional independence assumption, all these ERP-epochs are integrated in a naive Gaussian way. Moreover, prior language knowledge is further fused in the maximum a posteriori (MAP) approach.

4.5.1 Proposed Multi-class ERP Model

In the conventional binary-class ERP-epoch model, target-epochs and nontarget-epochs fulfill two Gaussian distributions with same covariance, respectively. However, the mean for target-epochs is the spatio-temporal pattern of the P300 and N200 components, while nontarget-epochs are with a zero-potential mean.

Unfortunately, single-trial ERP-epochs do not fit the binary-class model very well. As shown in Figure 4.4, ERP-epochs are overlapped with adjacent ones. Because the SOA (temporal interval between adjacent stimuli) is generally shorter than the length of ERP components. For example, the P300 component commonly lasts for 600–800 ms, whereas the SOA is usually set to 100–200 ms. Due to the "overlap" effect, nontarget ERP-epochs temporally adjacent to the target may contain part of the P300/N200 components. Moreover, the overlap extent is increased as nontarget ERP-epochs moving toward to the target. Clearly, nontarget-epochs close to the target are prone to be misclassified as the target-epochs by the conventional binary model. This is the second drawback of the epoch-based approach, besides the low SNR issue stated in subsection 4.4.4.

In order to overcome the problem, we propose a k-order multi-class ERP-epoch model. In the model, epochs of target, 1- to k-th nontarget preceding and following target, and other non-targets are modeled with Gaussian distributions with same

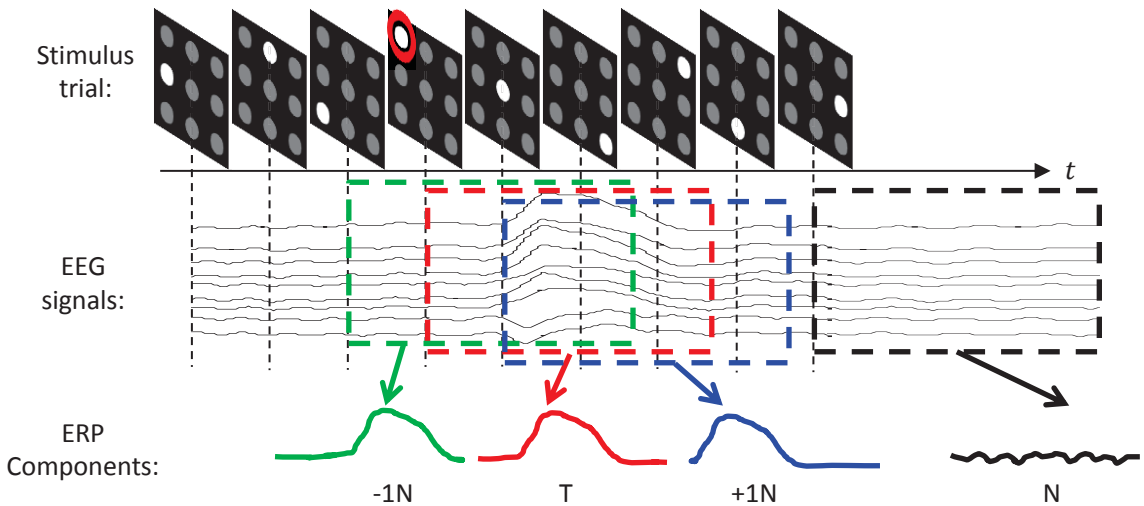


Figure 4.4: An example demonstrates ERP component contamination ("overlap" effect) of non-target ERP-epochs. A stimulus-trial of the 3×3 TLP is shown at top. A red circle denotes the target stimulus which users focus on. Meanwhile, ongoing multi-channel EEG signals (averaged over multi-trials for easy explanation) are displayed at middle. Dashed rectangles indicate each ERP-epochs (red: target ERP-epoch, green: nontarget-epoch just preceding the target, blue: nontarget-epoch just following the target, and black: other non-target epochs). Clearly, nontarget-epochs temporally more close to the target contain more part of P300 and N200 components.

covariance; however, they have distinct means. It is formalized as below:

$$
\begin{aligned}
&\textit{Target (T)}: && \mathbf{x}_T = \boldsymbol{\mu}_T + \mathbb{n} \sim && \mathcal{N}(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_{\mathbb{n}}) \\
&\textit{Nontarget (-1N)}: && \mathbf{x}_{-1N} = \boldsymbol{\mu}_{-1N} + \mathbb{n} \sim && \mathcal{N}(\boldsymbol{\mu}_{-1N}, \boldsymbol{\Sigma}_{\mathbb{n}}) \\
&\textit{Nontarget (+1N)}: && \mathbf{x}_{+1N} = \boldsymbol{\mu}_{+1N} + \mathbb{n} \sim && \mathcal{N}(\boldsymbol{\mu}_{+1N}, \boldsymbol{\Sigma}_{\mathbb{n}}) \\
&\ldots \\
&\textit{Nontarget (-kN)}: && \mathbf{x}_{-kN} = \boldsymbol{\mu}_{-kN} + \mathbb{n} \sim && \mathcal{N}(\boldsymbol{\mu}_{-kN}, \boldsymbol{\Sigma}_{\mathbb{n}}) \\
&\textit{Nontarget (+kN)}: && \mathbf{x}_{+kN} = \boldsymbol{\mu}_{+kN} + \mathbb{n} \sim && \mathcal{N}(\boldsymbol{\mu}_{+kN}, \boldsymbol{\Sigma}_{\mathbb{n}}) \\
&\textit{Nontarget (N)}: && \mathbf{x}_N = \mathbf{0} + \mathbb{n} \sim && \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbb{n}})
\end{aligned}
\tag{4.12}
$$

where $\mathbf{x}_T$ is the target-epoch, $\boldsymbol{\mu}_T$ denotes the target ERP-pattern, and noise is expressed by $\mathbb{n}$. The k-th nontarget epoch preceding or following target is denoted

by $\mathbf{x}_{-kN}$ or $\mathbf{x}_{+kN}$. Different means ($\boldsymbol{\mu}_{-1N}$ and $\boldsymbol{\mu}_{+1N}$) are caused by different temporal offsets to target. Last, other nontarget epochs are modeled as the ideal zero-mean distribution. In this study, models with order $k = 0, \ldots, 4$ are studied. Note that the 0-order model is degenerated to that conventional binary model.

4.5.2 Feature-dimension Reduction by Linear Projection

Multivariate Gaussian modeling is very difficult, since there are generally very limited training samples compared to the feature dimension, in practice. However, if we can reduce the feature dimension, the modeling may be easier.

Supervised feature dimension is to find a feature space where discrimination between classes is optimal. In this study, we apply a LDA extension to project ERP-epochs to a one-dimensional space. Although it is sub-optimal, the LDA extension is simple and easy to implement. The projection vector $\tilde{\mathbf{w}}$ is found by solving an optimization problem:

$$\tilde{\mathbf{w}} = \arg\max_{\mathbf{w}} \left| \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \right| \tag{4.13}$$

where total between-class variance $\mathbf{S}_b = \sum_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$, total within-class variance $\mathbf{S}_w = \sum_i \boldsymbol{\Sigma}_{\mathbb{m}}$, and $\boldsymbol{\mu}$ is the mean of $\boldsymbol{\mu}_i$. Here, subscript i denotes T, $\pm kN$, and N. Consequently, ERP-epoch $\mathbf{x}_i$ can be transformed into a scaler y_i by:

$$y_i = \tilde{\mathbf{w}}^T \mathbf{x}_i \tag{4.14}$$

The procedure is visualized in Figure 4.5. Note that ERP-epoch projection y_i should still fulfill Gaussian distribution. Because linear transform of a Gaussian distribution is still a Gaussian distribution in the transformed space. Actually, it can also be simply explained by the law of large numbers.

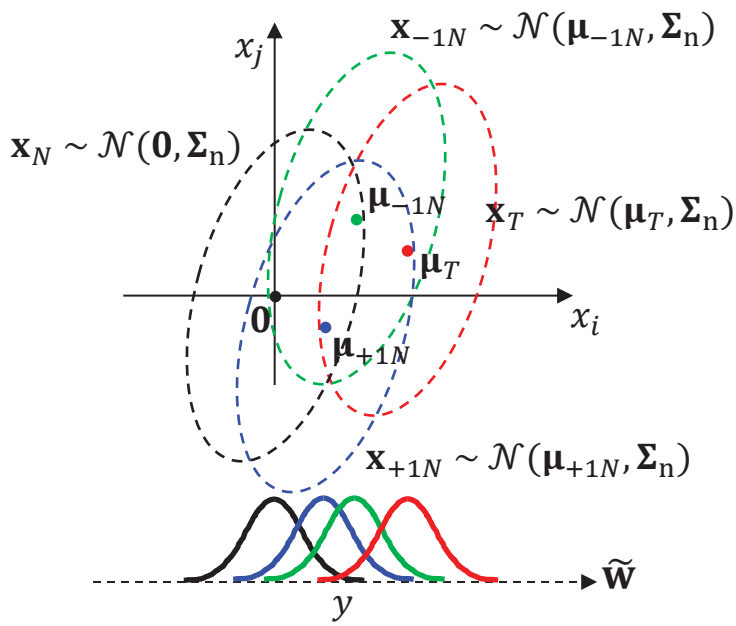


Figure 4.5: Classwise ERP-epochs are projected into a scalar space by the LDA extension, see text for details. Note that ERP-epochs still fulfill Gaussian distribution in the scalar space.

Now, probability density function of classwise ERP-epoch projection y is expressed as below:

$$
\begin{aligned}
&\textit{Target (T)}: && P_T^k(y) = \frac{1}{\sqrt{2\pi}\sigma_T} e^{-\frac{(y-\mu_T)^2}{2\sigma_T^2}} \\
&\textit{Nontarget (-1N)}: && P_{-1N}^k(y) = \frac{1}{\sqrt{2\pi}\sigma_{-1N}} e^{-\frac{(y-\mu_{-1N})^2}{2\sigma_{-1N}^2}} \\
&\textit{Nontarget (+1N)}: && P_{+1N}^k(y) = \frac{1}{\sqrt{2\pi}\sigma_{+1N}} e^{-\frac{(y-\mu_{+1N})^2}{2\sigma_{+1N}^2}} \\
&\ldots && \\
&\textit{Nontarget (-kN)}: && P_{-kN}^k(y) = \frac{1}{\sqrt{2\pi}\sigma_{-kN}} e^{-\frac{(y-\mu_{-kN})^2}{2\sigma_{-kN}^2}} \\
&\textit{Nontarget (+kN)}: && P_{+kN}^k(y) = \frac{1}{\sqrt{2\pi}\sigma_{+kN}} e^{-\frac{(y-\mu_{+kN})^2}{2\sigma_{+kN}^2}} \\
&\textit{Nontarget (N)}: && P_N^k(y) = \frac{1}{\sqrt{2\pi}\sigma_N} e^{-\frac{(y-\mu_N)^2}{2\sigma_N^2}}
\end{aligned}
\tag{4.15}
$$

where mean μ_i and standard deviation σ_i can be estimated from labeled training samples. Here, subscript i denotes T, $\pm kN$, and N. Note that superscript k in equation 4.15 denotes the model order.

4.5.3 ERP-decoding by Maximum A Posteriori (MAP) Estimation

Based on the proposed k-order ERP model and an n-gram statistic language model (see next subsection), a novel ERP-decoding algorithm is raised in this section.

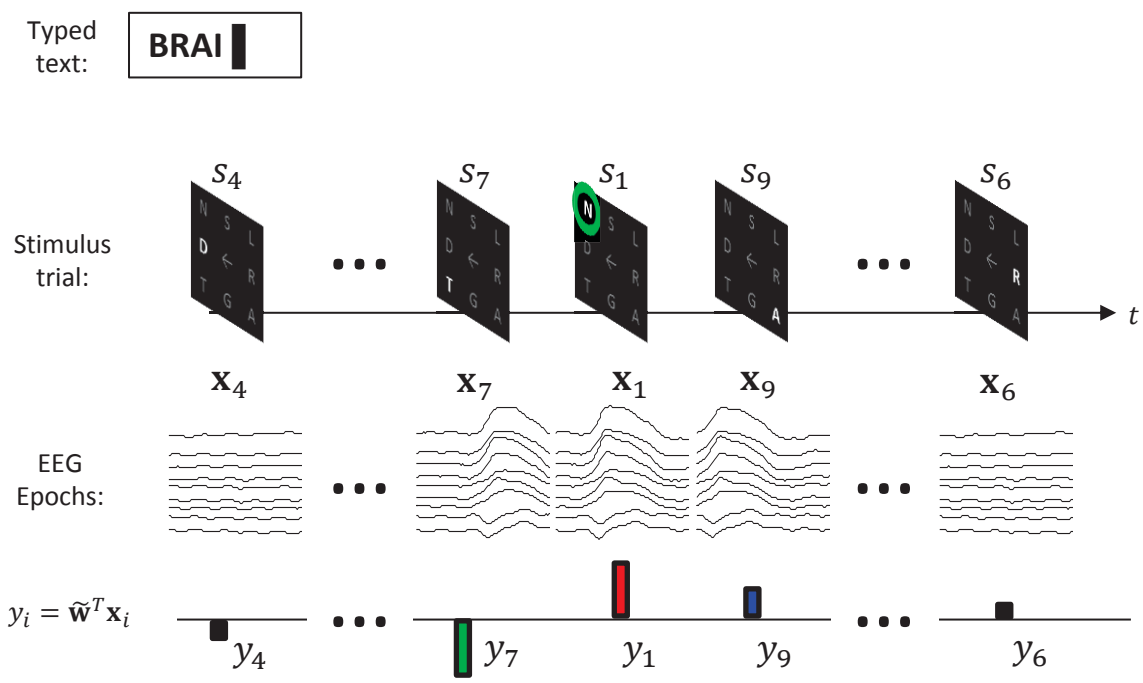


Figure 4.6: A text spelling scene using the 3×3 TLP by a subject. Assume the subject has already typed "BRAI". Based on the typed history, matrix of *Step-p* is generated. Next symbol (desired one or return) should be selected from the matrix. A 3×3 SC stimulation trial is shown here. All nine stimuli are presented once and their corresponding ERP-epochs are listed below each stimulus. The ERP-epochs are further reduced to scalar space.

Let's assume a spelling scene using the 3×3 TLP paradigm, as shown in Figure 4.6. A user has already typed some text $T_{History}$ (*e.g.*, "BRAI" as typed history). Then, the user tries to type next symbol. Given current stimulus matrix with symbol/stimuli set $S = \{s_1, s_2, \ldots, s_9\}$ (*e.g.*, "N", "S", "L", "D", "←", "R", "T", "G", "A" shown in in Figure 4.6), the user attends her/his attention to the desired symbol. After completing a trial of stimulation, single-trial ERP-epochs $\mathbf{x}_i \in \mathbb{R}^{M \cdot T}, i = 1, \ldots, 9$ (with M being the number of channels and T being the number of sampled time points) can be acquired through EEG devices. Then, the ERP-epochs $\mathbf{x}_i$ are transformed into $y_i \in \mathbb{R}, i = 1, \ldots, 9$ by equation 4.14.

Following the proposed multi-class ERP model, the ERP-epoch projection y_i should fulfill different Gaussian distributions regarding to the target stimulus. Obviously, all the nine y_i may contribute to the target prediction, because there are temporal relationships between them. Based on the findings, probability of any stimulus s_i being the target based on the trial of y_i $(i = 1, \ldots, 9)$ and typed text history $T_{History}$ is represented in a posteriori probability form $P(s_i|y_1, \ldots, y_9, T_{History})$. By using the formula of Bayes' theorem, we have:

$$P(s_i|y_1, \ldots, y_9, T_{History}) = \frac{P(y_1, \ldots, y_9|T_{History}, s_i)P(s_i|T_{History})}{P(y_1, \ldots, y_9|T_{History})} \tag{4.16}$$

Clearly, the typed text history $T_{History}$ is independent from current trial of ERP-epochs $y_1, \ldots, y_9$. Hence, $P(y_1, \ldots, y_9|T_{History}, s_i)$ can be simplified as $P(y_1, \ldots, y_9|s_i)$:

$$P(s_i|y_1, \ldots, y_9, T_{History}) = \frac{P(y_1, \ldots, y_9|s_i)P(s_i|T_{History})}{P(y_1, \ldots, y_9|T_{History})} \tag{4.17}$$

Note that the denominator $P(y_1, \ldots, y_9|T_{History})$ is independent from i, thus the probability $P(s_i|y_1, \ldots, y_9, T_{History})$ is proportional to the product of $P(y_1, \ldots, y_9|s_i)$ and $P(s_i|T_{History})$:

$$P(s_i|y_1, \ldots, y_9, T_{History}) \propto P(y_1, \ldots, y_9|s_i)P(s_i|T_{History}) \tag{4.18}$$

Then, we make a conditional independence assumption that $y_1, \ldots, y_9$ are independent from each other given s_i being the target. Hence, the term $P(y_1, \ldots, y_9|s_i)$ can be converted into product form $\prod_{j=1}^{9} P(y_j|s_i)$. Substitute it into equation 4.18, we have:

$$P(s_i|y_1, \ldots, y_9, T_{History}) \propto P(s_i|T_{History}) \prod_{j=1}^{9} P(y_j|s_i) \tag{4.19}$$

The conditional probability $P(y_j|s_i)$ can be estimated based on the proposed k-order multi-class ERP model:

$$P(y_j|s_i) = \begin{cases} P_T^k(y_j), & |q| = 0 \\ P_{qN}^k(y_j), & 1 \leq |q| \leq k \\ P_N^k(y_j), & |q| > k \end{cases} \tag{4.20}$$

where $P_T^k(\cdot)$, $P_{qN}^k(\cdot)$, and $P_N^k(\cdot)$ are defined in equation 4.15. The parameter q denotes the temporal offset of stimulus s_j to the assumed target stimulus s_i. Negative q means the stimulus s_j is $|q|$-th preceding the stimulus s_i. Similarly, positive q indicates the stimulus s_j is $|q|$-th following the stimulus s_i. Given the stimulation order in a trial, the parameter q can be easily deduced. Note that $P(y_j|s_i)$ should be updated to $P^k(y_j|s_i)$, since the proposed model is with an order-k. Consequently, the posteriori probability $P(s_i|y_1, \ldots, y_9, T_{History})$ also gains an order-k placed as superscript:

$$P^k(s_i|y_1, \ldots, y_9, T_{History}) \propto P(s_i|T_{History}) \prod_{j=1}^{9} P^k(y_j|s_i) \tag{4.21}$$

Finally, the equation 4.21 is normalized as below:

$$P^k(s_i|y_1, \ldots, y_9, T_{History}) = \frac{P(s_i|T_{History}) \prod_{j=1}^{9} P^k(y_j|s_i)}{\sum_{i=1}^{9} \{P(s_i|T_{History}) \prod_{j=1}^{9} P^k(y_j|s_i)\}} \tag{4.22}$$

In next subsection, the term $P(s_i|T_{History})$ in equation 4.22 may be learnt from an n-gram statistical language model.

4.5.4 N-gram Language Model

Prediction techniques are generally utilized by various human-computer interfaces (HCIs) to improve communication efficiency. Naturally, the ERP-based BCIs based on noisy EEG channels may benefit from such prediction techniques. It is well known that single-trial ERP classification is a very difficult task [26], since the single-trial ERP data is very noisy. Fusing prior language knowledge with the low-quality EEG data may facilitate the ERP classification. The term $P(s_i|T_{History})$ in equation 4.22 devotes to fuse prior language knowledge into the probabilistic model for ERP-based BCI spelling.

An n-gram statistical language model can be used to estimate the probability of any symbol s to be selected as next symbol based on most recently typed $n-1$ symbols:

$$P_{n-gram}(s^i|s^{i-(n-1)}, s^{i-(n-2)}, \ldots, s^{i-1}) = \frac{N(s^{i-(n-1)}, \ldots, s^{i-1}, s^i)}{\sum_s N(s^{i-(n-1)}, \ldots, s^{i-1}, s^i)} \tag{4.23}$$

where $N(\cdot)$ is the counting function. $N(s^{i-(n-1)}, \ldots, s^{i-1}, s^i)$ means the number of text $s^{i-(n-1)} \ldots s^{i-1}s^i$ existing in a training text database, superscript i denotes the occurring order of symbol s. In case of 0-gram, equal probability is assigned to any symbol s, *i.e.*, no statistical language information is exploited.

Therefore, the term $P(s_i|T_{History})$ in our proposed ERP-decoding method can be estimated based on an n-gram language model, and it gains a parameter gram-n:

$$P^n(s_i|T_{History}) = P_{n-gram}(s^i|s^{i-n+1}, s^{i-n+2}, \ldots, s^{i-1}) \tag{4.24}$$

In *Step-1* of the 3×3 TLP, each stimulus is an 8-symbol group rather than a single symbol. Hence, equal probability (1/9) is assigned to each of the nine $Step-1$-stimuli. For the *Step-2* and *Step-p*, prior probability of each symbol s_i (except the symbol '$\leftarrow$') is obtained by the equation 4.24. The center symbol '$\leftarrow$' functioned

as "return back to *Step-1*" means other choices. Therefore, the residual probability $1 - \sum_{i=1}^{8} P^n(s_i|T_{History})$ is assigned to the center symbol.

Finally, the equation 4.22 should be updated to add another superscript n indicating the gram-n:

$$P^{k,n}(s_i|y_1, \ldots, y_9, T_{History}) = \frac{P^n(s_i|T_{History}) \prod_{j=1}^{9} P^k(y_j|s_i)}{\sum_{i=1}^{9} \{P^n(s_i|T_{History}) \prod_{j=1}^{9} P^k(y_j|s_i)\}} \tag{4.25}$$

4.5.5 From Single-trial to Multi-trial

Now, the target stimulus/symbol can be determined by maximum a posterior (MAP) estimation:

$$\hat{i} = \arg\max_i P^{k,n}(s_i|y_1, \ldots, y_9, T_{History}) \tag{4.26}$$

Currently, the target prediction is based on single-trial ERP signals $(y_1, \ldots, y_9)$ and it is also the main purpose of the dissertation. In addition, this method can be naturally extended for multi-trial cases (T trials) by updating the equation 4.25 as below:

$$P^{k,n}(s_i|\overbrace{y_1^1, \ldots, y_9^1}^{\text{Trial 1}}, \ldots, \overbrace{y_1^t, \ldots, y_9^t}^{\text{Trial t}}, \ldots, \overbrace{y_1^T, \ldots, y_9^T}^{\text{Trial T}}, T_{History}) = \frac{P^n(s_i|T_{History}) \prod_{t=1}^{T} \prod_{j=1}^{9} P^k(y_j^t|s_i)}{\sum_{i=1}^{9} \{P^n(s_i|T_{History}) \prod_{t=1}^{T} \prod_{j=1}^{9} P^k(y_j^t|s_i)\}} \tag{4.27}$$

where superscript t denotes the trial order and T is the total number of trials.

The multi-trial extension is elegant; however, the calculation of term $P^k(y_j^t|s_i)$ is a bit different from the single-trial case. Note that there is only one stimulus s_i under single-trial condition, whereas the stimulus s_i occurs T times for T-trial case. Therefore, for evaluating any $P^k(y_j^t|s_i)$, the parameter q defined in equation 4.20 is determined as the temporal offset of stimulus s_j to the temporally closest target stimulus s_i. This is because the ERP-epoch projection y_j^t is majorally influenced by the temporally closest target stimulus. Given multi-trial ERP-epochs, the target

stimulus/symbol is finally determined by the MAP estimation:

$$\hat{i} = \arg\max_{i} P^{k,n}(s_i | \overbrace{y_1^1, \ldots, y_9^1}^{\text{Trial 1}}, \ldots, \overbrace{y_1^t, \ldots, y_9^t}^{\text{Trial t}}, \ldots, \overbrace{y_1^T, \ldots, y_9^T}^{\text{Trial T}}, T_{History}) \tag{4.28}$$

4.5.6 Hypothesis V: Conditional Independence of Epoch-Projections

In subsection 4.5.3, we made a conditional independence assumption. The assumption is that each (y_i) of the nine ERP-epoch projections in a single-trial is independent from each other, given the target stimulus s_i. Although the nine ERP-epochs are overlapped with adjacent ones, large variance caused by heavy background noise may make their dependencies very weak. Therefore, we make the fifth hypothesis in this study that the conditional independence assumption holds in practice.

4.5.7 Hypothesis VI: Proposed Method outperforms Conventional Methods

In contrast to single ERP-epoch utilization in state-of-the-art epoch-based methods, ensemble of ERP-epochs within-trial and statistical language information fusion are expected to suppress prediction variance. Hence, the proposed method is hypothesized to have better single-trial ERP-decoding accuracy.

Unlike straightforward trial-based methods working on high-dimensional ERP-trials, we divide the ERP-trial into nine separate ERP-epochs and further transform them into one-dimensional space. Therefore, the proposed method is hypothesized to have better performance even with limited training samples.

4.6 Summary

We proposed a trial-based ERP-decoding method for ERP-based BCI spelling in this chapter. The proposed one was built from learning strengths and overcoming shortcomings of conventional methods. Conventional binary ERP model was replaced with a more rational multi-class ERP model which aims to model the overlap effect. Based on the model, a LDA extension is used to reduce feature dimension. This strategy is critically important since there are generally very limited training samples for practical BCI using. Moreover, statistical language knowledge is fused to enhance low signal-to-noise ratio (SNR) of EEG signals. Finally, the proposed method is hypothesized to outperform conventional methods. The hypothesis will be validated through designated experiments in chapter 5.

Chapter 5

Experiments, Results, and Discussion

In this chapter, the hypotheses proposed in chapter 3 and chapter 4 will be validated by designated experiments. Statistical tests including student t-test, analysis of variance (ANOVA), Kolmogorov-Smirnov (KS) test, and Pearson's χ^2 independence test are used for significance testing.

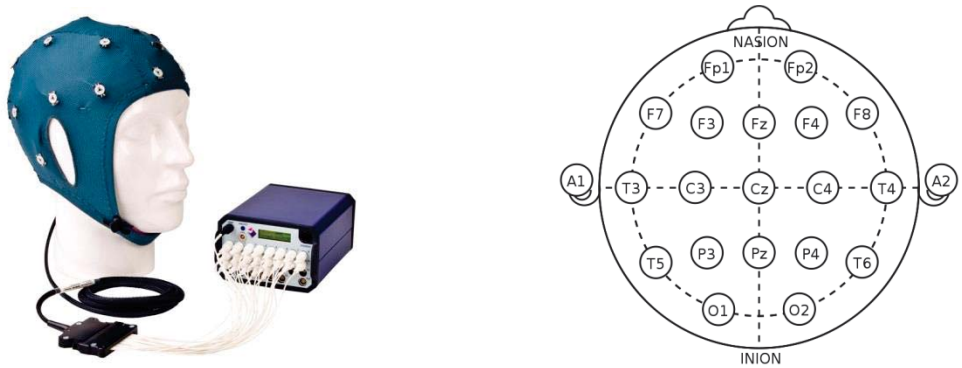
5.1 Subjects and Apparatus

5.1.1 Subjects

We recruited totally 48 healthy subjects (four females, age: 26.02 ± 1.96). All except one were naive to BCIs. The study was approved by Research Ethics Committee, Department of Engineering, the University of Tokyo. The study was in accordance with the Declaration of Helsinki and all subjects gave written consent. Note that subjects were assigned to take part in the following six experiments. Some of them participated in all the experiments, whereas others just took part in some experiments.

5.1.2 Hardwares

A 21-channel electrode cap (NeXus EEG cap, Mind Media, Netherlands, see Figure 5.1a) was used to record EEG signals. The electrodes embedded in the cap are placed according to 10-20 international system (see Figure 5.1b), and the electrodes are standardly named. The EEG signals were amplified and sampled by a NeXus-23 System (Mind Media, Netherlands, see Figure 5.1a). EEG signals were stored and processed on a laptop computer (Dell Latitude E6500). Stimulus paradigm was presented on a 21.3″ TFT screen (refresh rate: 60 Hz, EIZO FlexScan L985EX, Japan). Experiments were conducted in a quiet meeting room.



(a) NeXus-32 Device **(b)** 10-20 International System

Figure 5.1: EEG Apparatus and electrode placement.

5.1.3 Softwares

Stimulus paradigm, data acquisition, and signal processing were controlled by BCI2000 platform [178] and run on the laptop (Dell Latitude E6500). Offline analysis used BCI2000 offline tools and our own C++ programs.

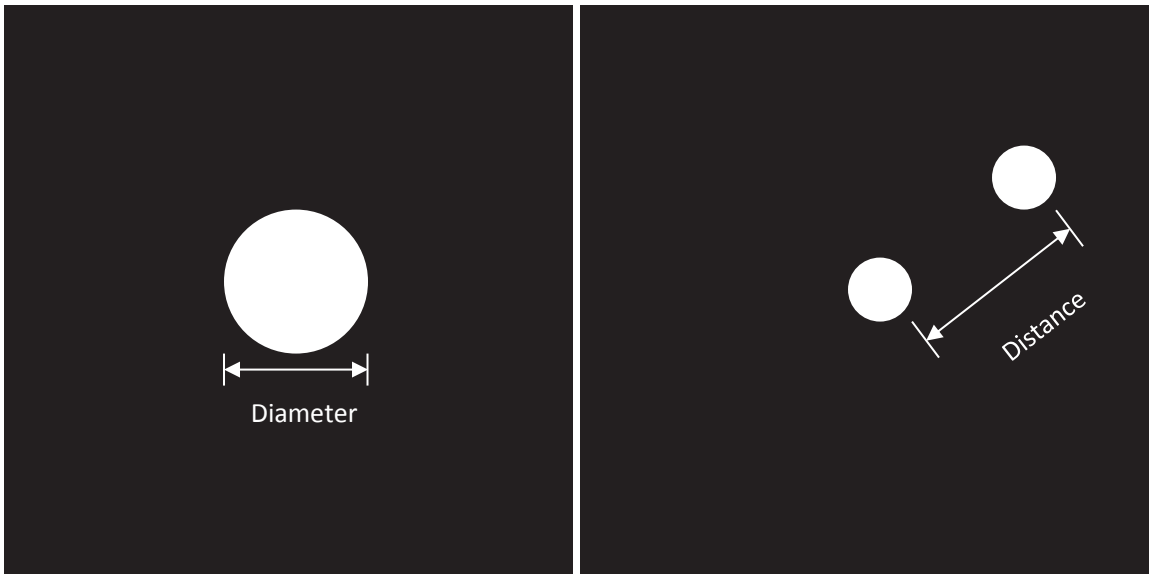
5.2 Study I: Initial Investigation of Stimulus Size and Inter-Stimulus Distance Effects

This study was designed to initially investigate the stimulus size (SS) and inter-stimulus distance (ISD) effects on ERPs. The SS and ISD was quantitatively manipulated to elicit ERPs, respectively.

5.2.1 Stimulus Setup

For investigating stimulus size (SS) effect on ERPs, a filled circle stimulus (see Figure 5.2a) is placed in the center of human visual view. The circle is colored "white" (RGB (255, 255, 255)) for stimulus "on", while it is filled with "gray" (RGB(128, 128, 128)) for stimulus "off". Stimulus luminances amount to 250 (white) and 83 (gray), given in candela/m^2. Their Michelson contrast is 50.15%. The stimulation constitutes an oddball paradigm, with one target stimulus and eight standard (non-target) stimuli. The target stimulus is mapped to the center circle stimulus, while the standard stimulus is map to time-reserved null-stimulus.

Similarly, for investigating inter-stimulus distance (ISD) effect on ERPs, a filled circle (target) stimulus with 1° diameter is placed in the center of human visual view. Meanwhile, a distractor stimulus is placed with a given distance from the target stimulus, while the direction to the target stimulus is randomized, see Figure 5.2b. The luminance configuration is same as the setup for SS-experiment. The stimulation also constitutes an oddball paradigm with one target and eight standard stimuli. The target stimulus is mapped to the center circle stimulus, while seven of the eight standard stimuli is mapped to time-reserved null-stimulus, and the rest one is mapped to the distractor.



(a) Stimulus size (SS) manipulation **(b)** Inter-stimulus distance (ISD) manipulation

Figure 5.2: (a) Stimulus setup for investigating stimulus size (SS) effect on ERPs. A circle stimulus is placed at the center of a monitor with black background. The diameter of the circle stimulus is manipulated to 0.5°, 1°, 2°, 3°, 4°, 5°, 6°, 7°, and 8°, respectively. (b) Stimulus setup for investigating inter-stimulus distance (ISD) effect on ERPs. A circle (target) stimulus with 1° diameter is placed at the center of a monitor with black background. A distractor stimulus with 1° diameter is placed with a given distance from the target stimulus, while the direction to the target stimulus is randomized. The distance is set to 1°, 2°, 3°, 4°, 5°, 6°, 7°, and 8°, respectively.

5.2.2 Experimental Procedure

Subjects sat comfortably at one meter from a 21.3″ TFT screen (refresh rate: 60 Hz) that displayed the stimulus paradigm depicted in Figure 5.2a or Figure 5.2b. Fifteen subjects participated in this study. For stimulus size (SS) investigation, each subject completed an experimental trial with stimulus diameter set to 0.5°, 1°, 2°, 3°, 4°, 5°, 6°, 7°, and 8° counterbalanced against order effect. Similarly, for inter-stimulus distance (ISD) investigation, each subject completed an experimental trial with ISD set to 1°, 2°, 3°, 4°, 5°, 6°, 7°, and 8° counterbalanced against order effect.

For each experimental trial, subjects were instructed to attend to the center circle stimulus and silently count the number of times the symbol flashed. The

symbols intensified for 62.5 milliseconds followed with a 62.5-milliseconds inter-stimulus interval, *i.e.*, the SOA of the oddball paradigm was set to 125 milliseconds. For each attending trial, constant 45-stimulus sequences were presented. One stimulus sequence included 9 (1 target and 8 standard) stimuli and lasted for 1.125 seconds.

5.2.3 Data Acquisition and Processing

EEG was recorded with a 21-channel electrode cap according to 10-20 system. All channels were referenced to the linked mastoid using a forehead ground, and impedances were kept below 10 kΩ. The EEG signals were amplified and sampled at 512 Hz by NeXus-32 System (Mind Media, Netherlands). Next, signals were bandpass filtered (0.5 Hz–30 Hz) and sectioned into overlapping epochs ranging from −100 ms prestimulus to 800 ms poststimulus. The prestimulus interval was used for baseline correction. Epochs were rejected if their peak-to-peak voltage difference in any channel exceeded $100\mu V$. Only data from channel Fz, Cz, Pz, O1, and O2 (see Figure 5.1b) were used for further analysis.

Maximum positive amplitudes in epochs averaged over the central electrodes Fz, Cz, and Pz ranging from 0 to 500 milliseconds were extracted as P300 peak amplitudes, while maximum negative amplitudes in epochs averaged over occipital electrodes O1 and O2 ranging from 0 to 300 milliseconds were extracted as N200 peak amplitudes. Note that P300 peak amplitudes for target stimulus were averaged over all target epochs, whereas distractor P300 peak amplitudes were determined as the largest P300 peak amplitude extracted from all distractor epochs.

5.2.4 Event-Related Potentials

The P300 and N200 peak amplitudes against stimulus size (SS) are depicted in Figure 5.3a and Figure 5.3b, respectively. In order to investigate the *SS* effect on P300 peak amplitudes, we conducted an one-way repeated ANOVA with factor *SS*. However, no significant difference was found ($F = 0.31, p = 0.96$). Same test was also conducted

to investigate the *SS* effect on N200 peak amplitude. Significant *SS* effect was found ($F = 7.99, p < 0.001$). Bonferroni *post hoc* tests found that N200 peak (absolute) amplitudes were significantly increased from 0.5° to 2∘, $p < 0.05$. However, there was no any statistical difference among the rest cases.

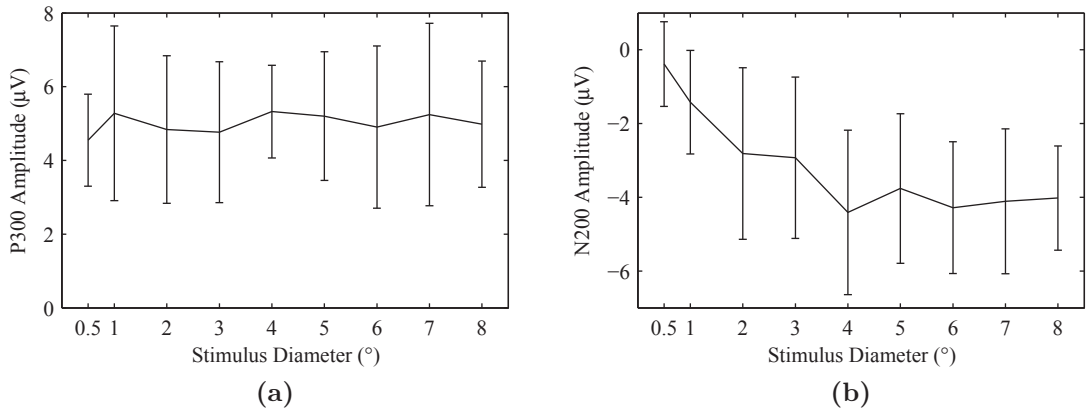


Figure 5.3: ERP results obtained for stimulus size (SS) investigation. (a) Mean and standard deviation of P300 peak amplitudes against the SS (circle diameter in visual angle). (b) Mean and standard deviation of N200 peak amplitudes against the SS.

The target and distractor P300 peak amplitudes against inter-stimulus distance (ISD) are depicted in Figure 5.4a and Figure 5.4b, respectively. One-way repeated ANOVA yielded significant *ISD* effect on target P300 peak amplitude ($F = 9.56, p < 0.001$). Pairwise *post hoc* tests revealed that there was no significant difference among ISDs from 1° to 4°. Same results were found for ISDs from 5° to 8°. More importantly, target P300 peak amplitudes achieved by ISDs from 5° to 8° were significantly larger than the amplitudes achieved by ISDs from 1° to 4°, $p < 0.05$, except for 3°-ISD.

In order to investigate *ISD* effect on distractor P300 peak amplitude, we subjected the amplitudes of all 15 subjects into an one-way repeated ANOVA. Significant *ISD* effect was revealed ($F = 5.80, p < 0.001$). Bonferroni *post hoc* tests found that P300 amplitudes accidently elicited by distractors with 7°- and 8°- ISDs were significantly lower than those elicited by distractors with 1°- and 2°- ISDs, $p < 0.05$. The amplitude for 7°-ISD was significantly smaller than the amplitude for 5°-ISD,

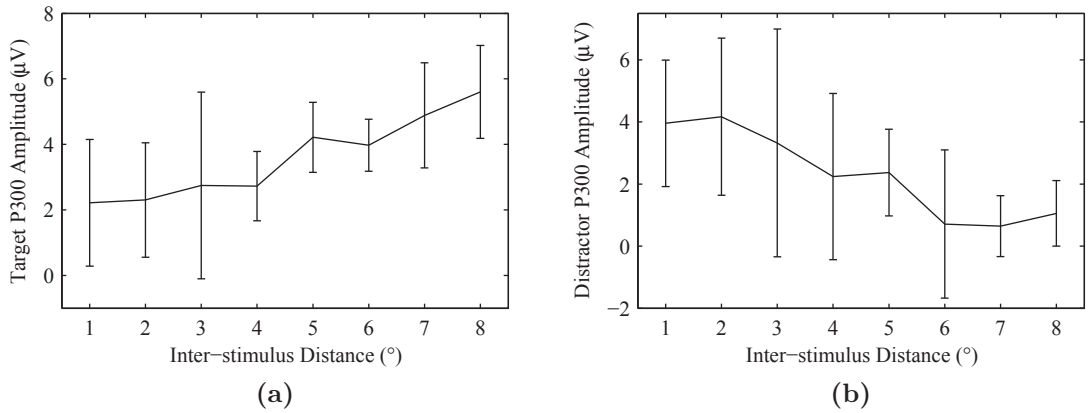


Figure 5.4: ERP results obtained for inter-stimulus distance (ISD) investigation. (a) Mean and standard deviation of target P300 peak amplitudes against the ISD (distance between target and distractor stimuli in visual angle). (b) Mean and standard deviation of distractor P300 peak amplitudes against the ISD.

$p < 0.05$. Moreover, the amplitude for 6°-ISD was also significantly smaller than the amplitude for 1°-ISD, $p < 0.05$.

5.2.5 Discussion

The P300 peak amplitude is not significantly affected by the stimulus size (SS). This may be attributed to the fact that P300 is an endogenous component, since it is caused by internal cognitive processes rather than by some automated preattentive processing of incoming stimuli. However, the N200 is an exogenous component, which is evoked by external visual stimulation. Larger stimulus may evoke larger N200 potential, because stronger visual stimulation is received. In study, N200 amplitudes for SS from 0.5° to 2° exactly met the above speculation. However, it seems reached the "ceiling" from 4°-SS.

Target P300 amplitude was strongly affected by surrounding distractor within 4°-ISD. This may be attributed to the fact that attention resource of subject is occupied by the distractor to some extent. This may deteriorate P300 amplitude according to the discussion in chapter 2. Meanwhile, distractor stimulus can attract subject's

attention to elicit "fake" P300 potentials. The results showed that distractor spatially close to the target stimulus may worsen the distracting effect.

Based on the results of initial SS- and ISD-investigations, we suggest that stimuli with $\geq 2°$-SS and over $\geq 6°$-ISD are favoured for further ERP-based BCI design.

5.3 Study II: Effects of Spatial Manipulation and Stimulus Pattern on ERPs and Offline Performance

This study was designated to validate Hypothesis I and Hypothesis II. Moreover, data collected in this study were also served as calibration (training) data for next study.

5.3.1 Experimental Procedure

Subjects sat comfortably at about one meter from a 21.3″ TFT screen (refresh rate: 60 Hz) that displayed the 8×8 or 3×3 matrix (see Figure 5.5 for setup details). Ten of 20 subjects participated in this study. For simplicity, we named them S10. Each subject completed an 8×8 RC, a 3×3 RC, and a 3×3 SC calibration experiment counterbalanced against order effect.

Subjects were instructed to attend to a symbol and silently count the number of times the symbol flashed. A total of 30 symbols (10 runs×3 symbols/run) were provided for selection (see Table 5.1). For the 8×8 RC calibration, 30 symbols were allocated to evenly cover the 8×8 matrix; whereas for the 3×3 RC/SC calibration, each of 9 symbols was chosen as the target for 3 or 4 times to get 30 targets in total. Each set of symbols intensified for 62.5 milliseconds followed with a 62.5-milliseconds inter-stimulus interval (*i.e.*, SOA: 125 milliseconds) for both 8×8 RC and 3×3 SC, whereas the 3×3 RC applied a 187.5-milliseconds SOA that consisted of 62.5-milliseconds intensification duration and 125-milliseconds inter-stimulus interval in

(a) 8×8 RC **(b)** 3×3 RC **(c)** 3×3 SC

Figure 5.5: (a) 8×8 row/column (RC) paradigm. (b) 3×3 RC paradigm. (c) 3×3 single cell (SC) paradigm. In this experimental setup, horizontal and vertical adjacent distances of the 8×8 and 3×3 matrices are $3.05\,^\circ$ and $2.31\,^\circ$, $8.14\,^\circ$ and $6.15\,^\circ$, respectively. Stimulus luminances amount to 250 (white) and 83 (gray), given in candela/m^2. Their Michelson contrast is 50.15%. The numbers of pixels for single symbol/group are adjusted to 1.82 ± 0.84 10^3px (8×8 RC) and 9.90 ± 2.67 10^3px (3×3 RC/SC), with each pixel occupying 0.27×0.27 mm^2.

order to keep complete stimulus sequence durations of the 3×3 SC and 3×3 RC the same. For each of the 30 symbol selections, constant 15-stimulus sequences were presented. After each selection, a 3.5-seconds pause ensued before the next selection. One 8×8 RC sequence included 16 (2 target and 14 non-target) flashes and lasted 2 seconds, while one 3×3 SC sequence and one 3×3 RC sequence included 9 (1 target and 8 non-target) flashes and 6 (2 target and 4 non-target) flashes, respectively, and lasted 1.125 seconds. Thus, for each subject, 16 minutes, 41.5 seconds of calibration data were collected for the 8×8 RC, while 10 minutes, 7.75 seconds were collected for the 3×3 RC/SC.

At the end of each run, subjects were asked to report their counting number. After each of the three calibrations (8×8 RC, 3×3 RC, and 3×3 SC), subjects were asked to rate a visual fatigue score and a distraction score for the stimulus pattern by visual analogue scales (VAS), ranging from 0 (not at all tiring/distracting) to 10 (very tiring/distracting). The questionnaire is attached in Appendix B (see Figure B.1).

Table 5.1: 30 symbols (10 runs $\times$ 3 symbols/run) for calibration experiments for 8×8 and 3×3 matrices, respectively.

Run	8×8	3×3
#1	ASX	AEC
#2	JBQ	IHA
#3	_U5	DFG
#4	ZI@	BDE
#5	4M#	CAF
#6	(H$	EBI
#7	%0)	GCH
#8	ENR	FIB
#9	LO3	HGD
#10	V=D	AFG

5.3.2 Data Acquisition, Pre-processing, and Classification

EEG was recorded with a 21-channel electrode cap according to 10-20 system. All channels were referenced to the linked mastoid using a forehead ground, and impedances were kept below 10 kΩ. The EEG signals were amplified and sampled at 512 Hz by NeXus-32 System (Mind Media, Netherlands). Next, signals were bandpass filtered (0.5 Hz–30 Hz) and sectioned into overlapping epochs ranging from -100 ms prestimulus to 800 ms poststimulus. The prestimulus interval was only used for offline baseline correction. Epochs from the calibration data were rejected if their peak-to-peak voltage difference in any channel exceeded $100\mu V$. Before final classification, the epochs were moving averaged and downsampled to 20 Hz.

Preprocessed 800-ms epochs were used to train SWLDA classifiers for the 8×8 RC (900 target and 6300 non-target epochs, artifact rejection rate: $< 1\%$), 3×3 RC (900 target and 1800 non-target epochs, ditto), and 3×3 SC (450 target and 3600 non-target epochs, ditto). Max feature size, p-value for entry, and p-value for remove were set to 60, 0.1, and 0.15, respectively. Before online test, symbol written rate (WSR) for each sequence number (1–15) was calculated by 10-fold cross validation for each of the three patterns. Fixed optimal subject-specific sequence numbers of the three patterns were determined by maximising corresponding WSR for each subject.

5.3.3 Behavioral Data

Three repeated ANOVAs were performed separately for the subjective ratings of visual fatigue, distraction, and counting accuracy with factor *Pattern* (8×8 RC, 3×3 RC, 3×3 SC).

Counting accuracies were $93.94\% \pm 5.18\%$ (8×8 RC), $92.14\% \pm 7.72\%$ (3×3 RC), and $99.39\% \pm 0.89\%$ (3×3 SC). Pairwise comparisons found that the counting accuracy for the 3×3 SC was significantly higher than accuracies for the 8×8 and 3×3 RC, $p < 0.05$.

Visual fatigue scores were 5.20 ± 0.84 (8×8 RC), 7.80 ± 0.55 (3×3 RC), and 2.20 ± 0.47 (3×3 SC). Pairwise comparisons found that the 3×3 SC score was significantly lower than the 8×8 and 3×3 RC scores, $p < 0.01$.

Distraction ratings for the 8×8 RC (8.10 ± 0.62) was significantly higher than for the 3×3 RC (2.70 ± 0.63) and 3×3 SC (0.70 ± 0.30), $p < 0.0005$. However, other pairwise differences were not significant.

5.3.4 Event-Related Potentials

Grand average ERP responses at electrodes Fz, Cz, and Pz for the three patterns are depicted in Figure 5.6, while grand mean ERPs at electrodes O1 and O2 are depicted in Figure 5.7. Please refer to Appendix Figure A.1 for grand mean ERP responses at all the nine electrodes. To test whether ERP amplitude and latency differed across patterns, we subjected the calibration data (survived previous artifact rejection) of the S10 to a two-way repeated ANOVA with factors *Pattern* (8×8 RC, 3×3 RC, 3×3 SC) and *Electrode* (Fz, Cz, Pz or O1, O2).

The analysis of P300 peak amplitude revealed a significant effect of *Pattern* ($F = 13.00, p < 0.001$) but not *Electrode* ($p = 0.121$). *Pattern*$\times$*Electrode* interaction was not significant ($p = 0.495$). Bonferroni *post hoc* tests found that P300 peak amplitude was significantly higher for the 3×3 SC than for the other two patterns. However, there was no significant difference between the latter two. A significant *Pattern* effect

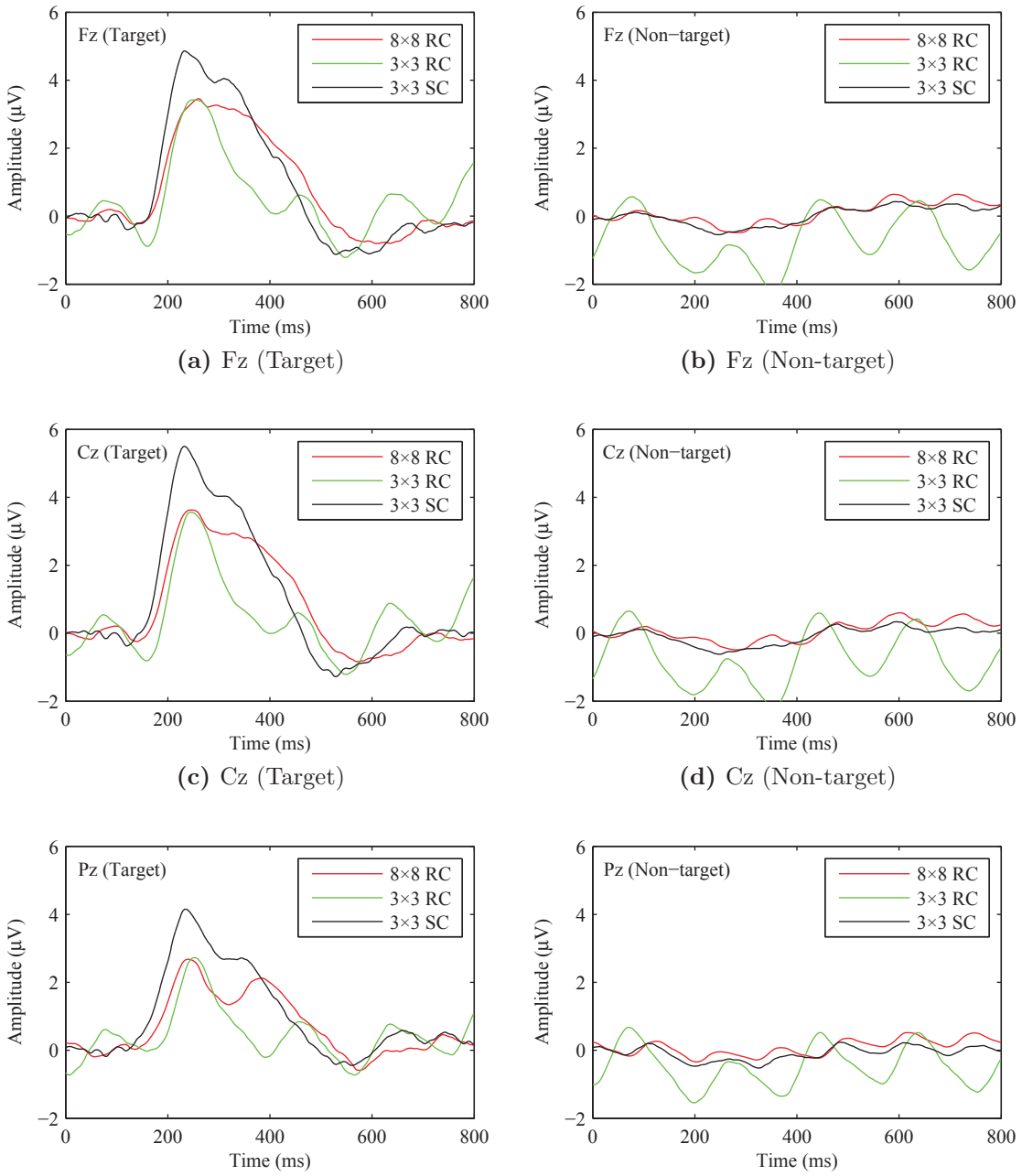
on P300 peak latency was found ($F = 4.74, p < 0.05$), but not *Electrode* ($p = 0.306$) and *Pattern*×*Electrode* interaction ($p = 0.532$). Bonferroni *post hoc* tests showed that the latencies (listed in Table 5.2) were significantly larger for the 8×8 RC than for the other two patterns. Difference between the latter two was not significant.

Table 5.2: P300 peak latencies of 8×8 RC, 3×3 RC, and 3×3 SC over electrodes Fz, Cz, and Pz for each of S10

Subject	**Fz (ms)**			**Cz (ms)**			**Pz (ms)**		
	RC8	**RC3**	**SC3**	**RC8**	**RC3**	**SC3**	**RC8**	**RC3**	**SC3**
#1	236.33	261.72	232.42	234.38	253.91	234.38	238.28	257.81	246.09
#2	296.88	281.25	257.81	287.11	259.77	250.00	294.92	261.72	261.72
#4	410.16	279.30	359.38	408.20	279.30	267.58	406.25	265.63	394.53
#5	441.41	212.89	212.89	441.41	207.03	212.89	419.92	189.45	187.50
#6	271.48	269.53	253.91	277.34	265.63	248.05	363.28	265.63	347.66
#9	296.88	255.86	427.73	257.81	257.81	224.61	257.81	261.72	224.61
#10	414.06	511.72	339.84	398.44	238.28	337.89	367.19	251.95	335.94
#11	265.63	242.19	234.38	261.72	242.19	234.38	257.81	244.14	234.38
#12	244.14	238.28	230.47	244.14	238.28	228.52	242.19	244.14	228.52
#13	300.78	234.38	285.16	337.89	232.42	347.66	363.28	234.38	351.56
M	317.77	278.71	283.40	314.84	247.46	258.59	321.09	247.66	281.25
SD	75.40	84.63	70.00	76.00	20.28	46.91	70.30	22.98	69.74
SE	23.84	26.76	22.13	24.03	6.41	14.84	22.23	7.27	22.05

Mean N200 peak amplitudes averaged across electrodes O1 and O2 were $-1.88\mu V$ (8×8 RC), $-1.65\mu V$ (3×3 RC), and $-2.25\mu V$ (3×3 SC). Although no significant *Pattern* effect was found ($p = 0.124$), the largest absolute N200 peak amplitude was achieved by the 3×3 SC. No any significant effect was found on N200 peak latencies (listed in Table 5.3).

We also compared the absolute maximum peak-to-peak values of the non-target responses between the three patterns, see Table 5.4. Significant larger peak-to-peak amplitudes were achieved by the 3×3 RC compared with the other two patterns for electrodes Fz, Cz, and Pz. However, other pairwise differences were not significant.

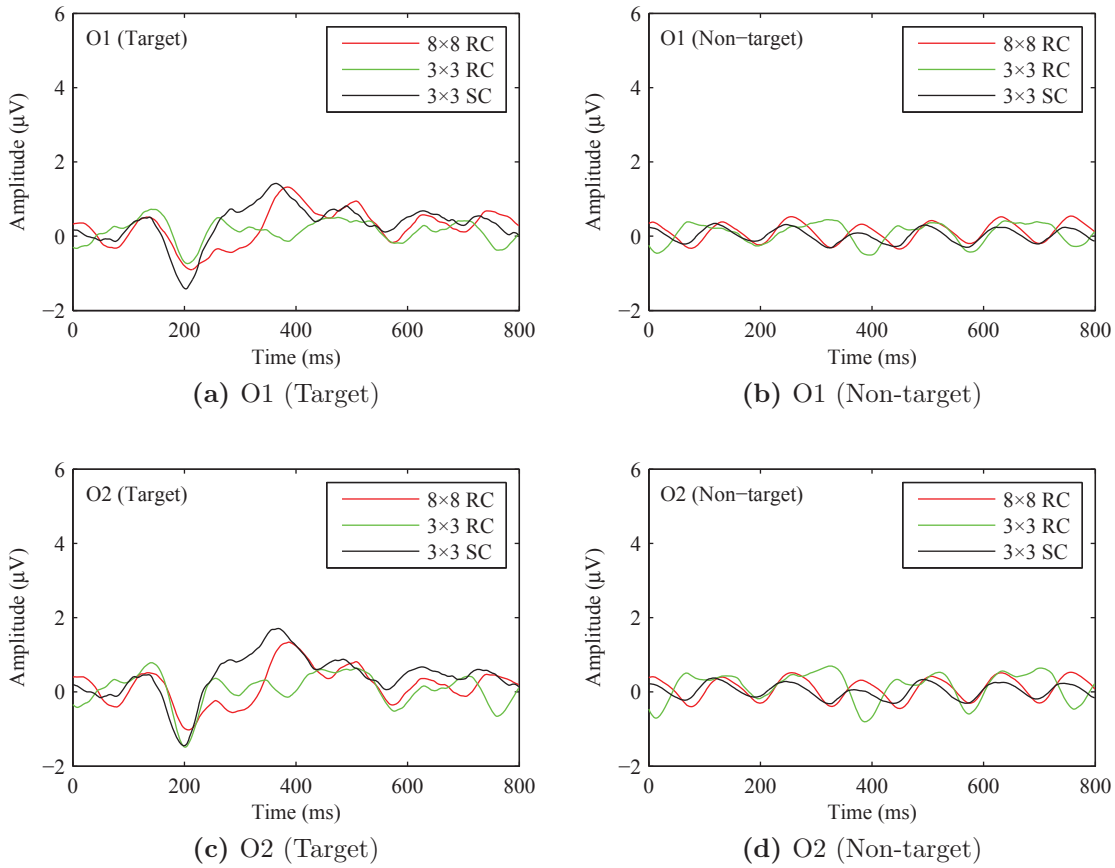


(a) Fz (Target) **(b)** Fz (Non-target)

(c) Cz (Target) **(d)** Cz (Non-target)

(e) Pz (Target) **(f)** Pz (Non-target)

Figure 5.6: Grand mean waveforms of Target (left) and Non-target (right) responses for the 8×8 RC (red), 3×3 RC (green), and 3×3 SC (black) at middle-line electrodes Fz, Cz, and Pz.



(a) O1 (Target) **(b)** O1 (Non-target)

(c) O2 (Target) **(d)** O2 (Non-target)

Figure 5.7: Grand mean waveforms of Target (left) and Non-target (right) responses for the 8×8 RC (red), 3×3 RC (green), and 3×3 SC (black) at occipital electrodes O1 and O2.

5.3.5 Single-trial Binary Classification Accuracy

Single-trial class-wise normalized binary classification accuracy was computed for each of S10 and each of the three patterns, see Figure 5.8. Binary classification of target and non-target epochs (survived previous artifact rejection) was performed using SWLDA and was estimated by a 10-fold cross validation based on balanced training samples between the three patterns. Over all the S10, the accuracies for the 3×3 SC ($79.05\% \pm 2.22\%$) were significantly higher than for the 8×8 RC ($73.20\% \pm 2.03\%$, $p < 0.05$) and 3×3 RC ($72.56\% \pm 2.14\%$, $p < 0.01$). However, there was no significant difference between the latter two.

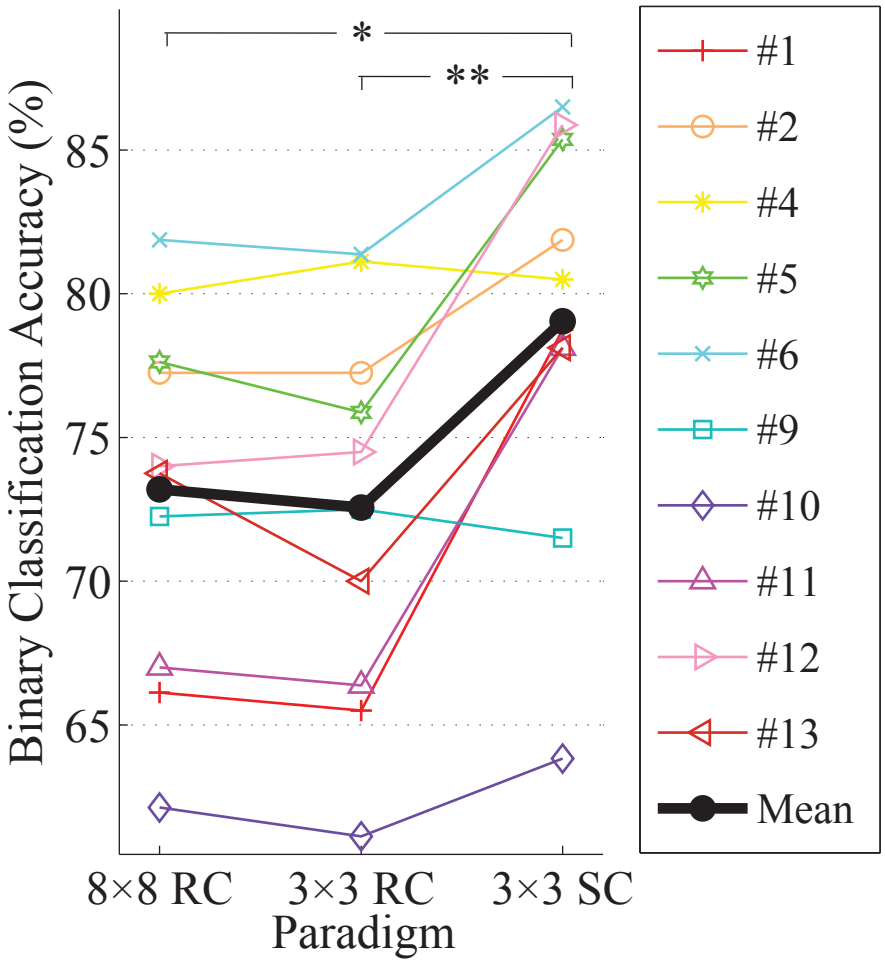


Figure 5.8: Single-trial binary classification accuracy for each of the S10 and mean for each of the three patterns. Single and double asterisks indicate that the accuracy for 3×3 SC is significantly higher than for 8×8 RC ($p < 0.05$) and for 3×3 RC ($p < 0.01$).

Table 5.3: N200 peak latencies of 8×8 RC, 3×3 RC, and 3×3 SC over electrodes O1 and O2 for each of S10

Subject	**O1 (ms)**			**O2 (ms)**		
	RC8	**RC3**	**SC3**	**RC8**	**RC3**	**SC3**
#1	199.22	39.06	185.55	199.22	21.48	183.59
#2	220.70	208.98	210.94	216.80	210.94	207.03
#3	283.20	287.11	220.70	289.06	289.06	214.84
#4	226.56	226.56	216.80	250.00	224.60	210.94
#5	199.22	195.31	187.50	208.98	193.36	187.50
#6	210.94	197.27	207.03	212.89	193.36	173.83
#7	169.92	1.95	199.22	195.31	201.17	199.22
#8	58.59	11.72	300.78	60.55	119.14	27.34
#9	78.13	214.84	167.97	199.22	201.17	187.50
#10	201.17	191.41	205.08	201.17	201.17	203.13
M	184.77	157.42	210.16	203.32	185.55	179.49
SD	68.00	100.63	35.63	58.04	70.76	55.05
SE	21.50	31.82	11.27	18.36	22.38	17.41

Table 5.4: Peak-to-peak amplitudes of non-target responses of 8×8 RC, 3×3 RC, and 3×3 SC over electrodes Fz, Cz, and Pz for each of S10

Subject	**Fz** (μV)			**Cz** (μV)			**Pz** (μV)		
	RC8	**RC3**	**SC3**	**RC8**	**RC3**	**SC3**	**RC8**	**RC3**	**SC3**
#1	2.42	3.34	2.03	2.68	4.07	2.21	1.70	3.34	1.94
#2	1.26	5.66	1.89	1.15	4.17	1.64	1.40	3.45	1.46
#4	1.58	4.16	1.11	1.92	4.37	1.39	1.69	2.05	1.00
#5	2.28	3.20	2.06	2.03	2.98	1.76	1.75	1.75	1.50
#6	1.88	3.78	1.89	2.33	4.05	1.91	2.66	3.49	1.71
#9	1.26	1.91	1.14	1.13	2.15	1.14	1.20	2.09	0.90
#10	1.39	1.80	1.23	1.04	2.28	0.79	1.21	2.16	0.79
#11	0.92	3.82	1.56	0.86	4.14	1.74	1.05	3.10	1.85
#12	1.25	4.38	2.14	1.99	4.62	2.00	2.15	3.95	2.70
#13	1.12	2.30	1.63	1.35	2.22	1.16	1.12	1.51	0.94
M	1.54	3.44	1.67	1.65	3.50	1.57	1.59	2.69	1.48
SD	0.50	1.20	0.39	0.62	0.99	0.45	0.51	0.86	0.60
SE	0.16	0.38	0.12	0.20	0.31	0.14	0.16	0.27	0.19

5.3.6 Offline Accuracy, Information Transfer Rate (ITR), and Written Symbols Rate (WSR)

Offline classification accuracy (Figure 5.9a), information transfer rate (ITR, Figure 5.9b), and written symbol rate (WSR, Figure 5.9c) as a function of sequence number were obtained by performing 10-fold cross validation on the surviving calibration data of the S10. We subjected the results to a two-way repeated ANOVA with factors *Pattern* (8×8 RC, 3×3 RC, 3×3 SC) and *Sequence*.

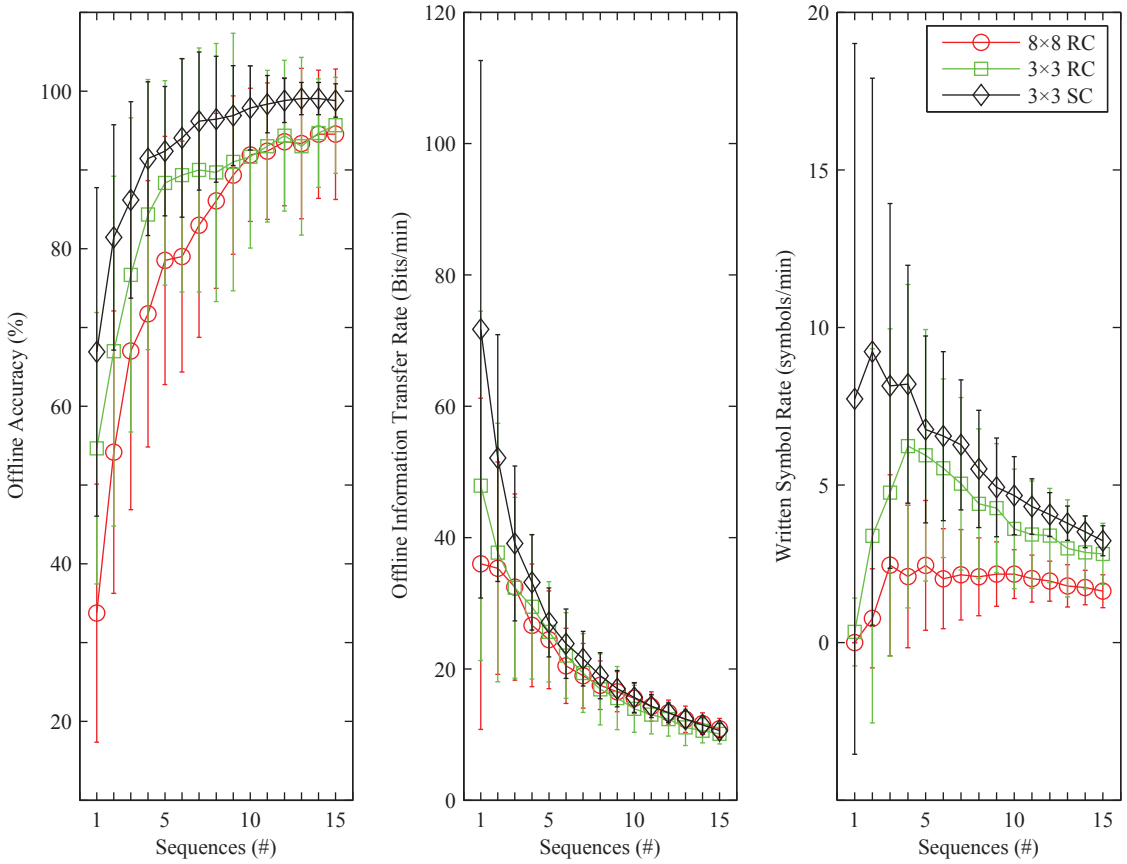
As expected, accuracies increased sharply with the growth of sequence numbers ($F = 64.00, p < 0.001$). A significant *Pattern* effect was also revealed ($F = 10.68, p < 0.01$). Pairwise *post hoc* comparisons found the overall accuracy for the 3×3 SC was significantly higher than for the 8×8 RC ($p < 0.01$) and 3×3 RC ($p < 0.05$). However, there was no significant difference between the latter two, although the 3×3 RC had superior accuracies with few sequence repetitions than the 8×8 RC.

Statistic analysis of ITR revealed a significant *Pattern* effect ($F = 7.97, p < 0.01$). Bonferroni *post hoc* tests showed that the overall ITR was significantly higher for the 3×3 SC than for the 8×8 RC ($p < 0.01$) and 3×3 RC ($p < 0.05$). No significant difference was found between the latter two.

Moreover, each of the S10 achieved smaller optimal number of sequences for the 3×3 SC than for the 3×3 RC. Thus, the SC stimulus pattern was applied in the 3×3 TLP online test for all subjects. A paired t-test of the optimal sequence numbers gave further evidence ($p < 0.05$). More importantly, the corresponding maximum WSR was significantly higher for the 3×3 SC than for the 3×3 RC ($p < 0.01$).

5.3.7 Discussion

We devised a 3×3 matrix paradigm with RC/SC pattern to offline compare with an 8×8 RC matrix paradigm. On the whole, the 3×3 SC is a better candidate for further developing a BCI aiming for middle stage of ALS, because it achieved less visual fatigue, easier gaze deployment, larger ERPs, and better offline performances.



(a) Offline Accuracy **(b)** Offline Information Transfer Rate **(c)** Offline Written Symbol Rate

Figure 5.9: Offline accuracy, information transfer rate (ITR), and written symbol rate (WSR) of the 8×8 RC, 3×3 RC, and 3×3 SC for each of S10.

Larger ERPs might be particularly important for clinical application, because ALS patients usually suffer from ERP attenuations [69].

The ERP analysis results revealed that largest P300 amplitude was evoked by the 3×3 SC. This may be mainly attributed to the TTI effect [43, 155]. Moreover, large inter-stimulus distance (ISD) may also contribute. 3×3 SC superiority over 3×3 RC is consistent with [38, 144]. Previous studies [53, 146] reported significant matrix-size effect; however, there was no significant difference between the 8×8 RC and 3×3 RC in this study. This may be attributed to the larger SOA, symbol-size, and/or inter-symbol distance that was applied to the 3×3 RC. Significant larger P300 latency was achieved by the 8×8 RC over the 3×3 RC/SC. Selection in a more complex and distracting matrix should demand more mental resources. This may cause a delay in P300 [43, 62]. The results was in accordance with that six of the S10 felt very hard to concentrate on specific target in the 8×8 RC. Furthermore, N200 amplitudes revealed no difference between the 8×8 and 3×3 matrices, despite our expectation that the 3×3 matrices should evoke larger N200 potentials since larger stimuli applied in the both 3×3 matrices. In addition, significantly strong background oscillations (about 5 Hz) were observed at all 9 electrodes except O1 and O2 for the 3×3 RC. Perhaps this revealed effect was caused by the 1/3 of the matrix symbols that were intensified for the 3×3 RC in one flash in contrast to 1/8 and 1/9 for the 8×8 RC and 3×3 SC. This might can be used to explain that the S10 all reported more visual fatigue for the 3×3 RC than for the other two.

Single-trial binary classification accuracy was also exploited to reveal how discriminant target/non-target signals modulated by the three patterns are. As expected, the 3×3 SC achieved the best discriminant performance. This should be attributed to larger target ERP amplitudes and relatively smooth background activities generated under the 3×3 SC. Furthermore, the 3×3 SC significantly outperformed the other two in terms of offline classification accuracy, ITR, and WSR.

5.4 Study III: Online Validation of the Proposed TLP Paradigm

This study was designated for validating Hypothesis IV. Moreover, the online data during this study were collected for next study.

5.4.1 Experimental Procedure

Subjects sat comfortably at about one meter from a 21.3″ TFT screen (refresh rate: 60 Hz) that represented the 8×8 RC paradigm (see Figure 5.10) or the 3×3 TLP paradigm (see Figure 5.11). Illumination and other configurations are same as in Figure 5.5. Totally 14 subjects (the S10 in the study II plus additional four subjects) participated in this study. Each subject completed an 8×8 RC and a 3×3 TLP online experiment counterbalanced against order effect.

Classifiers used in this study were trained on calibration data collected from the study II for each subject, respectively. Different from the last calibration experiment, result feedback was provided during the online test phase. Subjects were asked to correctly spell an English sentence with 57 symbols ("PLEASE WASH YOUR HANDS WITH WATER AND SOAP BEFORE DINNER!"). A break-continue procedure occurred when a subject could not correctly type a target symbol in five consecutive tries. In this case, the investigator pauses the experiment, manually corrects all errors, and continues the task after the subject had enough rest. For the 3×3 TLP, the 3×3 SC or 3×3 RC pattern was applied based on calibration result analysis for the S10. The 3×3 SC pattern was used for the remaining 4 subjects. Subjects were informed the fact that symbols in *Step-p* of TLP were arranged row-wisely with prior probability descending. The statistical language model, PPMD5, was trained by an 864 kB novel (Emma, by Jane Austen). Eight (14 %) of 57 target-symbol selections suffered prediction failure in *Step-p*. For both 8×8 RC and 3×3

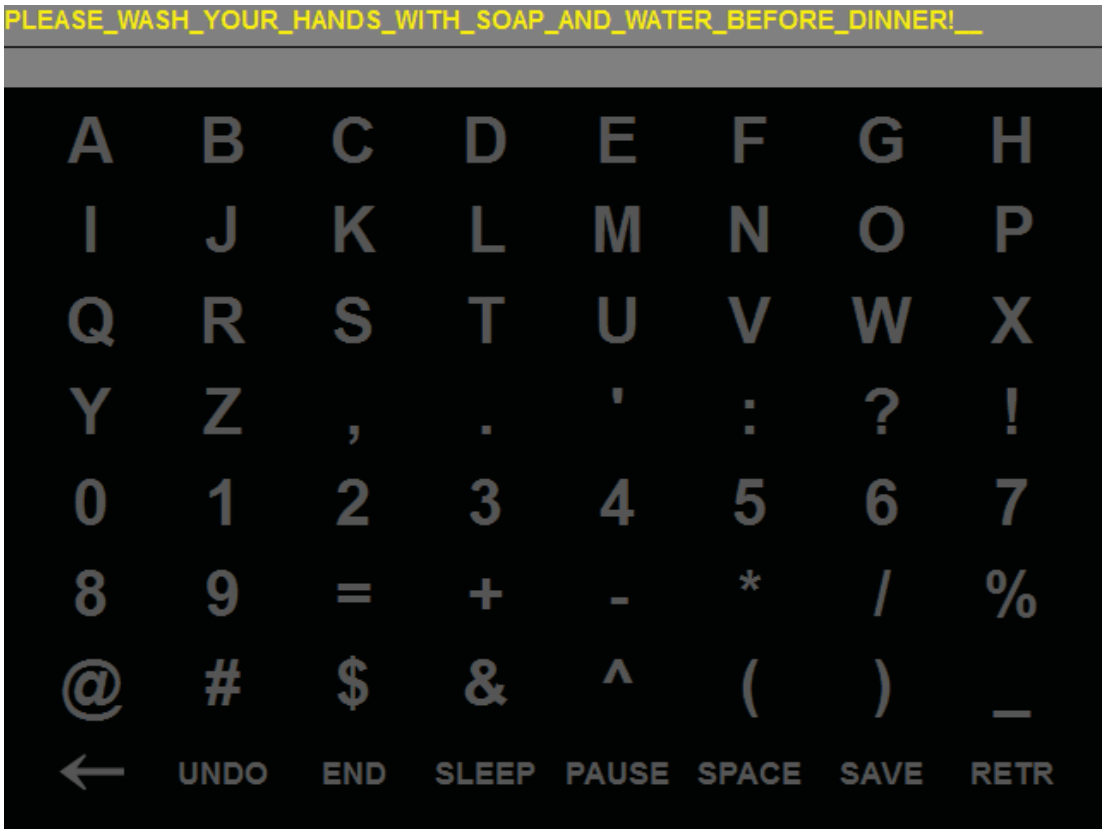


Figure 5.10: During online spelling, target text are shown at the first line, while selection result will be shown at the second line. Bottom is the 8×8 RC matrix paradigm whose configuration is same as the 8×8 RC in Figure 5.5.

TLP, the number of sequence for one selection was changed from 15 to a subject-specific number (see last subsection). The whole procedure was videotaped. After finishing the online spelling task, subjects were asked to recall whether the target was successfully attended before stimulation starting for each selection, with the aid of the video. Finally, subjects were asked to express their preference between the 8×8 RC and 3×3 TLP paradigm.

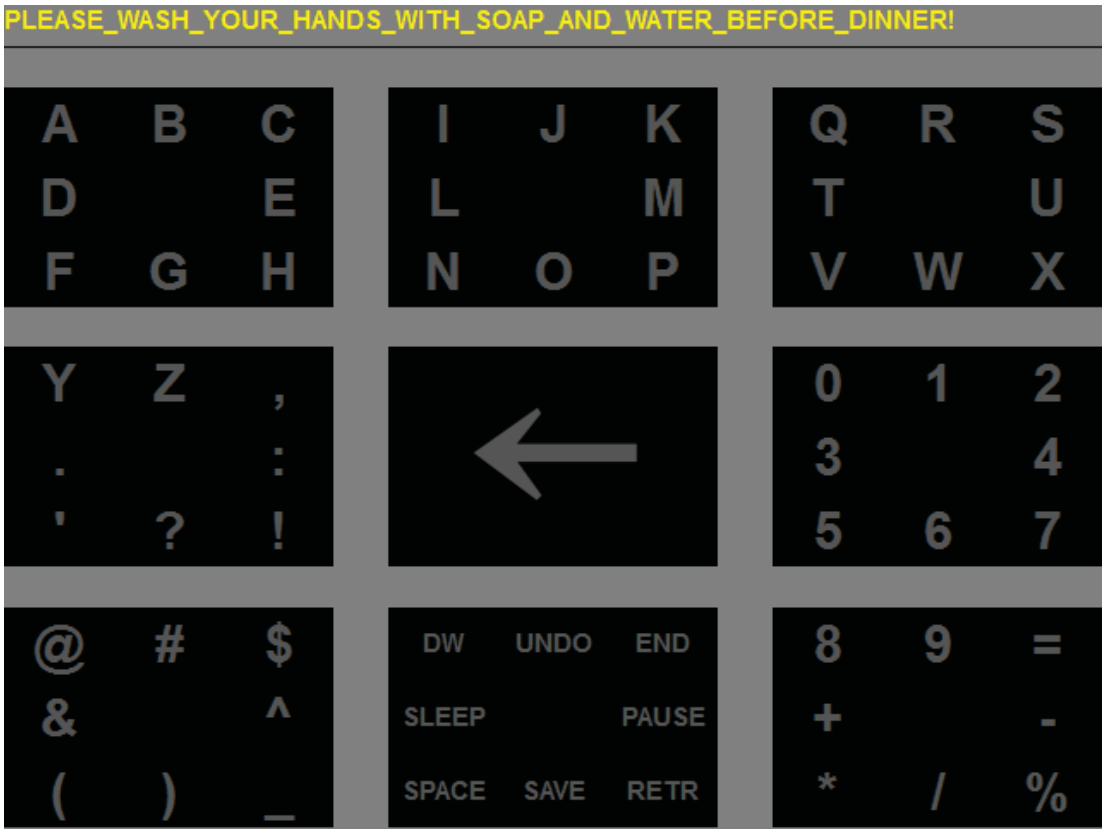


Figure 5.11: During online spelling, target text are shown at the first line, while selection result will be shown at the second line. Bottom is the 3×3 TLP paradigm whose configuration is same as the 3×3 RC/SC in Figure 5.5.

5.4.2 Data Acquisition, Pre-processing, and Classification

EEG was recorded with a 21-channel electrode cap according to 10-20 system. All channels were referenced to the linked mastoid using a forehead ground, and impedances were kept below 10 kΩ. The EEG signals were amplified and sampled at 512 Hz by NeXus-32 System (Mind Media, Netherlands). Next, signals were bandpass filtered (0.5 Hz–30 Hz) and sectioned into overlapping epochs ranging from 0-ms to 800-ms poststimulus. Before final classification, the epochs were moving averaged and downsampled to 20 Hz. Only electrodes Fz, C3, Cz, C4, P3, Pz, P4, O1, and O2 were used for online BCI operation.

Fixed optimal subject-specific sequence numbers of the three patterns were determined by maximising corresponding WSR for each subject by calibration experiment (see last subsection). Moreover, the 3×3 SC or 3×3 RC pattern which achieved smaller optimal sequence number was used in the 3×3 TLP online test for each of the S10. During online test, symbol stimuli were presented for a specified number of times, epochs were averaged for each stimulus, and then the averaged epoch was multiplied by SWLDA coefficients to obtain a SWLDA score for each stimulus. In the RC paradigm, the target symbol was determined as the intersection of the row and column stimuli with the highest scores and was presented to the subjects as feedback. In the SC paradigm, the symbol with the highest score was output as the selection result.

5.4.3 Behavioral Data

For all the 14 subjects participated in this study, only one of them missed the target once for selection in the 8×8 RC. Thirteen of them preferred the 3×3 TLP, and the remaining one expressed no preference.

5.4.4 Online Accuracy, ITR, and Time to Complete the Task

Online classification results are depicted in Figure 5.12. Online accuracy for selecting one symbol out of 64 (chance level 1.56%) was $73.65\% \pm 4.63\%$ for the 8×8 RC and $88.99\% \pm 1.11\%$ for the 3×3 TLP. Since all 14 subjects applied the 3×3 SC pattern in the 3×3 TLP online test, we also obtained online selection accuracy for the 3×3 SC ($88.46\% \pm 1.09\%$), i.e. selection accuracy for each step of the 3×3 TLP (chance level 11.11%). To test whether there was a significant difference between the performances of the three paradigms, we conducted a one-way repeated ANOVA yielding a significant *Paradigm* effect ($F = 11.14, p < 0.01$). Bonferroni *post hoc* tests revealed that accuracies for the 3×3 SC ($p < 0.05$) and 3×3 TLP ($p < 0.05$) were

significantly higher than for the 8×8 RC. However, there was no significant difference between the first two paradigms.

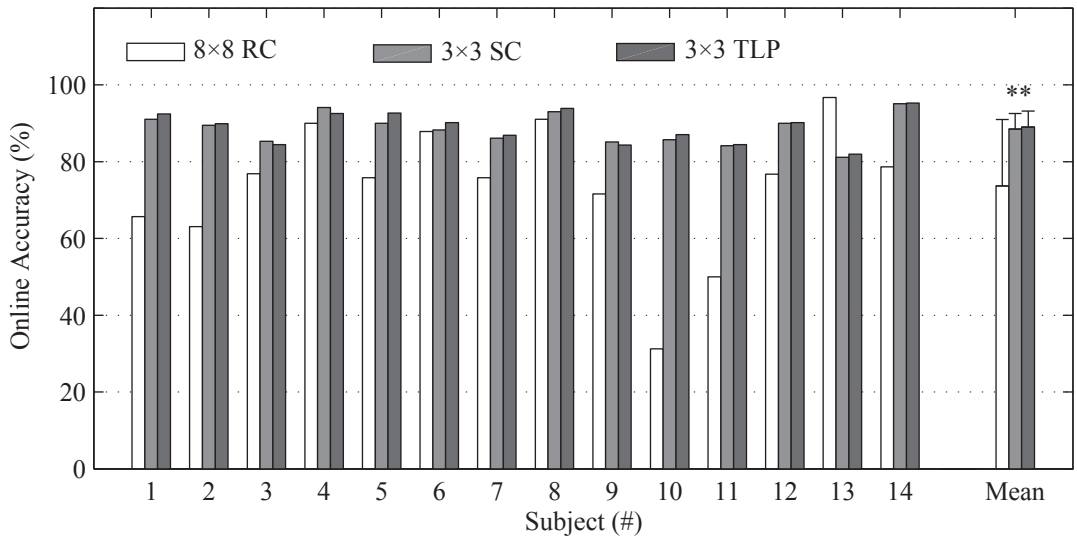


Figure 5.12: Online accuracy of the 8×8 RC (white), 3×3 SC (gray), and 3×3 TLP (dark) for each subject and mean (with standard deviation). Double asterisks denote that a statistically significant difference exists across the three paradigms, $p < 0.01$.

Online ITR was $17.44 bits/min \pm 2.26 bits/min$ for the 8×8 RC, $18.79 bits/min \pm 1.19 bits/min$ for the 3×3 SC, and $23.30 bits/min \pm 1.55 bits/min$ for the 3×3 TLP, as shown in Figure 5.13. A one-way repeated ANOVA revealed a significant *Paradigm* effect ($F = 8.77, p < 0.01$). Bonferroni *post hoc* tests revealed that the ITR for the 3×3 TLP was significantly higher than for the 8×8 RC ($p < 0.001$) and 3×3 SC ($p < 0.05$). No significant difference was found between the latter two paradigms. Note that the ITR calculated for the 3×3 TLP was a conservative one since the number of possible choices (defined in page 19 of [6]) was set to 33 (i.e. PPMD5 symbol set size). However, symbols selected in *Step-2* were actually selected from 64 candidates.

Total time to complete the online test tasks are depicted in Figure 5.14. Subject #10 and #11 gave up on the spelling task under the 8×8 RC paradigm due to their poor performances, whereas they both finished the task using the 3×3 TLP paradigm.

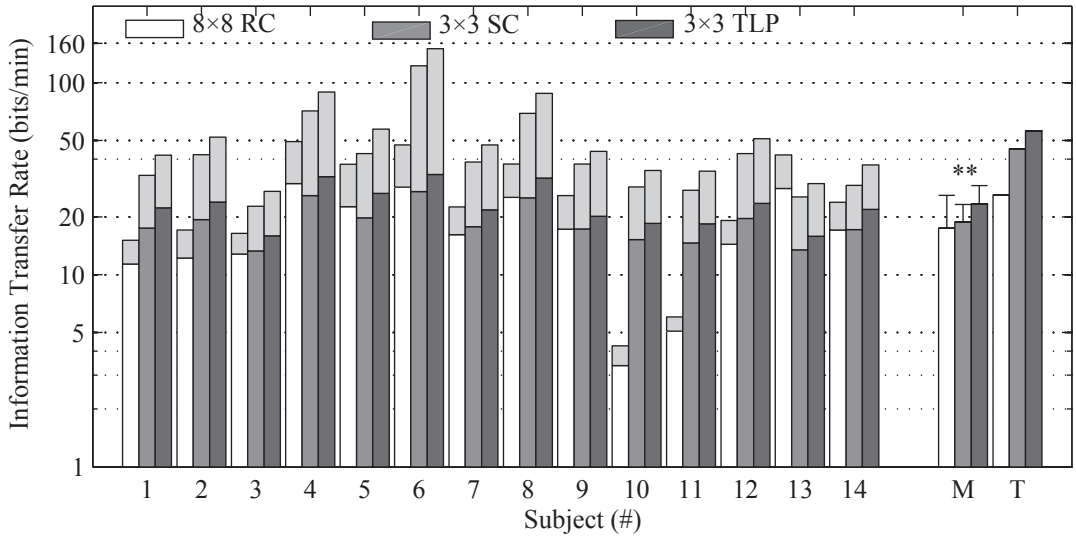


Figure 5.13: Online information transfer rate (ITR) of the 8×8 RC (white), 3×3 SC (gray), and 3×3 TLP (dark) for each subject and mean with standard deviation (denoted by 'M'). Double asterisks indicate that a significant difference is revealed between the three paradigms, $p < 0.01$. Light-gray extended bars show the theoretical increasing of the ITRs by excluding time-intervals between selections. The mean theoretical ITRs are denoted by 'T'.

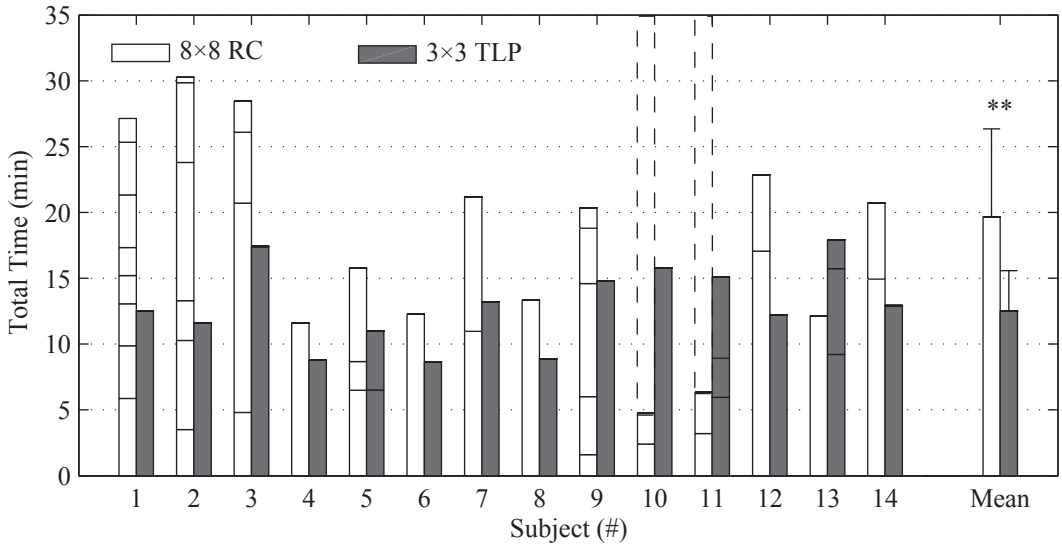


Figure 5.14: Total time to spell (with error correction) an English sentence with 57 characters for each subject and mean (with standard error) using the 8×8 RC (white) and 3×3 TLP (gray). Double asterisks denote that consumed time is significantly decreased by the 3×3 TLP when compared with the 8×8 RC, $p < 0.01$. Stacked bars mean non-continuous spelling (see online test description in subsection 5.4.1) and dashed bars indicate subject #10 and #11 gave up on the tasks after two runs due to their poor performances.

Excluding the two subjects, a paired t-test revealed significantly less consumed time for the 3×3 TLP ($12.50min \pm 0.89min$) than for the 8×8 RC ($19.67min \pm 1.93min$), $p < 0.01$. Furthermore, much more breaks happened for the 8×8 RC than for the 3×3 TLP ($p < 0.05$).

5.4.5 Discussion

The 3×3 SC applying two-level strategy and incorporating a statistical language model (PPMD5) constitutes the 3×3 TLP paradigm with the same degree of communication freedom (DOCF) as the 8×8 RC. In this study, online accuracy and ITR were significantly higher by 20.83% and 33.60% for the 3×3 TLP compared to the 8×8 RC, respectively. Time to complete the spelling task was also significantly decreased by 36.45% for the 3×3 TLP. It is worth noting that two subjects could

even improve their online accuracies from below 50% to more than 80% with the transition from the 8×8 RC to the 3×3 TLP. The results were due to the fact that high single-step accuracy was achieved by the 3×3 SC and redundant steps were partly reduced by applying *Step-p*. Moreover, the 3×3 TLP was much more robust when considering the breaks happened during the online tests. Note that less amount of calibration data were used for training classifiers used in the 3×3 TLP online test compared with the 8×8 RC online test. In other words, the proposed paradigm needs less calibration time, and can achieve better performance.

Direct performance comparison of BCIs is tedious, as reported measures are based on different experimental configurations. However, two predictive BCIs [160, 161] with similar setup configuration (SOA, inter-trial interval, calibration time, etc.) as our TLP are good choices for comparison. Average times to correctly spell an English sentence with 58, 45, and 57 characters were 12.7 minutes [160], 12.4 minutes [161], and 12.5 minutes (TLP), respectively. Their performances are comparable. Although they are still can not compete with traditional AAC devices (e.g. eyetracker with 10 words/min), they all advanced a solid step compared with non-predictive ERP-BCIs. Moreover, comparing with the two comparable BCIs, transition from the 3×3 TLP to AMUSE [103] or PASS2D [56] when patients transfer from middle-stage to late-stage of ALS may be much more smooth.

It is concluded that the proposed 3×3 TLP paradigm outperformed the 8×8 RC paradigm in online accuracy, ITR, speed, and robustness. This can explain why most subjects reported to prefer the TLP paradigm.

5.5 Study IV: Prediction Impact on Online Performance of TLP Paradigm

This study was designed to validate Hypothesis III that there is no significant performance differences between *Step-1*, *Step-2*, and *Step-p* in the proposed TLP paradigm.

5.5.1 Dataset Description

Online spelling data including stepwise EEG signals and selection accuracies were recorded in the last study. The online data from all the 14 subjects are offline analyzed in this study. ERP responses for *Step-1*, *Step-2*, and *Step-p* were averaged for electrodes Fz, C3, Cz, C4, P3, Pz, P4, O1, and O2, respectively, for each subject. Grand average of P300 (Cz) and N200 (O1) responses over the 14 subjects are shown in Figure 5.15.

5.5.2 Comparison between *Step-1*, *-2*, and *-p*

To investigate prediction impact on online performance, we subjected the online stepwise accuracies to a one-way repeated ANOVA with factor *Step* (*Step-1, Step-2, Step-p*). No significant difference of online accuracies between *Step-1* (86.52% $\pm$ 2.56%), *Step-2* (92.05%$\pm$1.58%), and *Step-p* (88.48%$\pm$1.19%) was found ($p = 0.131$). In order to investigate the underlying reason, sample-wise one-way repeated ANOVAs were conducted between the three steps for both target and non-target responses, respectively. The analysis results showed that there was no any significant sample-wise difference for both target and non-target responses at all nine electrodes. Comparison results between steps for electrodes Cz and O1 are visualized in Figure 5.15. Please refer to Appendix Figure A.2 for all results.

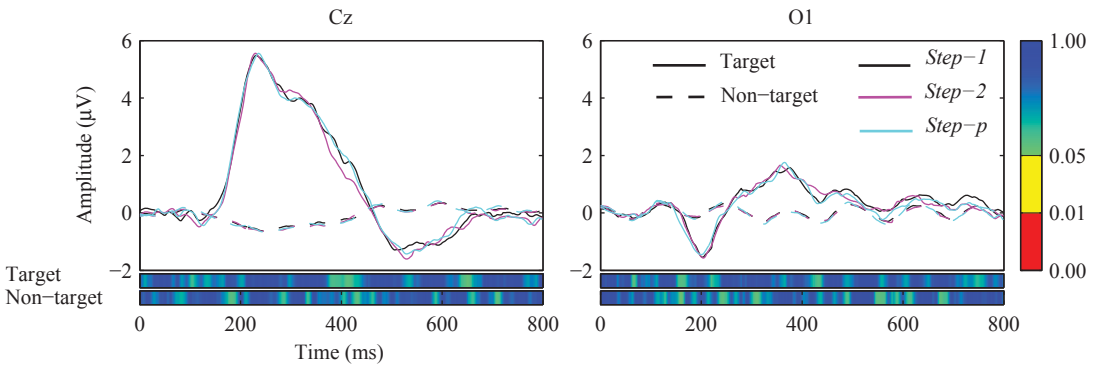


Figure 5.15: Grand mean waveforms for *Step-1* (black), *-2* (pink), and *-p* (cyan) of Target (solid) and Non-target (dashed) at electrodes Cz and O1. Color-bars below each waveform are coded by sample-wised p-values of ANOVAs conducted between *Step-1*, *-2*, and *-p* for Target and Non-target responses, respectively. No significant difference is found between the three steps.

5.5.3 Discussion

Prediction assistance greatly improved the BCI speed in [160], but it also significantly hampered the accuracy. Due to a word/sentence-wise predictive program they applied, subjects had to evaluate a list of word or sentence candidates before each selection in a limited time. Hence, increased cognitive workload was expected. This might can be proved by reduced ERP amplitudes. In our 3×3 TLP paradigm, selection in dynamically generated *Step-p* may be cognitively more difficult than selection in the *Step-1* and *Step-2* with constant symbol arrangements. However, statistical tests did not reveal any significant effect on online accuracies and ERPs between the three steps. This implies letter-wise prediction with 8-letter candidates in the *Step-p* might have introduced high cognitive workloads, but it did not significantly hamper ERPs, which can be attributed to the intuitive control scheme and simplicity of symbol identification from a small set of random labels with prior probability descending. This was also in accordance with that no target-attending failures reported for *Step-p*. However, such results were obtained from young and healthy

subjects. It may be an obstacle for patients with oculomotor impairments or the aged.

Most current gaze-independent BCIs [56, 97, 98, 103] exploit very limited amount of stimuli, because they require subjects to remember the feature coding. They generally also apply two-level strategies. Thus, they may also benefit from the TLP paradigm.

5.6 Study V: Validation of Conditional Independence Assumption

This study was designated to validate Hypothesis V that conditional independence holds in practice between ERP-epoch projections within a trial.

5.6.1 Experimental Procedure

The 3×3 SC paradigm was used in this study. Stimulus illumination and temporal configurations are same as setting in the study II. Totally 20 subjects participated in this study. Same instructions were given to subjects as in the study II. Only difference is the designated stimulation trials. In order to collect same amount of ERP-trial data under the nine conditions (each of nine stimuli being the target in a stimulus trial for each condition), target symbols were carefully selected, and then randomly given to subjects to attend in order to counter the order effect. For each subject, totally 500 trials were conducted under each condition.

5.6.2 Data Acquisition and Pre-processing

EEG was recorded with a 21-channel electrode cap according to 10-20 system. All channels were referenced to the linked mastoid using a forehead ground, and impedances were kept below 10 kΩ. The EEG signals were amplified and sampled at 512 Hz by NeXus-32 System (Mind Media, Netherlands). Next, signals were bandpass

filtered (0.5 Hz–30 Hz) and sectioned into overlapping epochs ranging from 0-ms to 800-ms poststimulus. The epochs were then moving averaged and downsampled to 20 Hz.

Based on the collected epoch-data (9(conditions)×500(trials)×9(epochs)), subject-optimal k-order ERP models were trained by 10-fold cross-validation for each of the 20 subjects, respectively. Then, trial-wise 9 epochs were transformed into trial-wise 9 epoch-projections.

5.6.3 Results of Pearson's χ^2 Independence Test

In order to test whether the conditional independence assumption holds or not in practice, we conducted Pearson's χ^2 independence test between each pair ($C_9^2 = 36$) of the nine epoch-projections within a trial, under each of the nine conditions. The tests were conducted for each of the 20 subjects, respectively. The Pearson's χ^2 independence test was conducted with a null-hypothesis that the input-pair is statistically independent from each other. If p-value of the test reaches the significance level, the null-hypothesis is rejected. The pairwise p-values under nine conditions for subject #1 are visualized in Figure 5.16. Average 2.22 times of rejections (violations) happened over the nine conditions for the subject #1. Independence violation times for all the 20 subjects are stated in Table 5.5. For the total 36 pairs, there are 2.97 ± 1.01 violations over all the 20 subjects.

5.6.4 Discussion

The conditional independence assumption is not perfectly holds. This is expected since the fact that ERP-epochs are overlapped with adjacent ones. According to the results, there are about three violations among the total 36 cases. This is very low, hence the assumption can be considered valid, in practice. This assumption can also be considered reasonable based on ERP neurophysiological background. Because each ERP-epoch is elicited by an independent stimulus. Although there is overlap

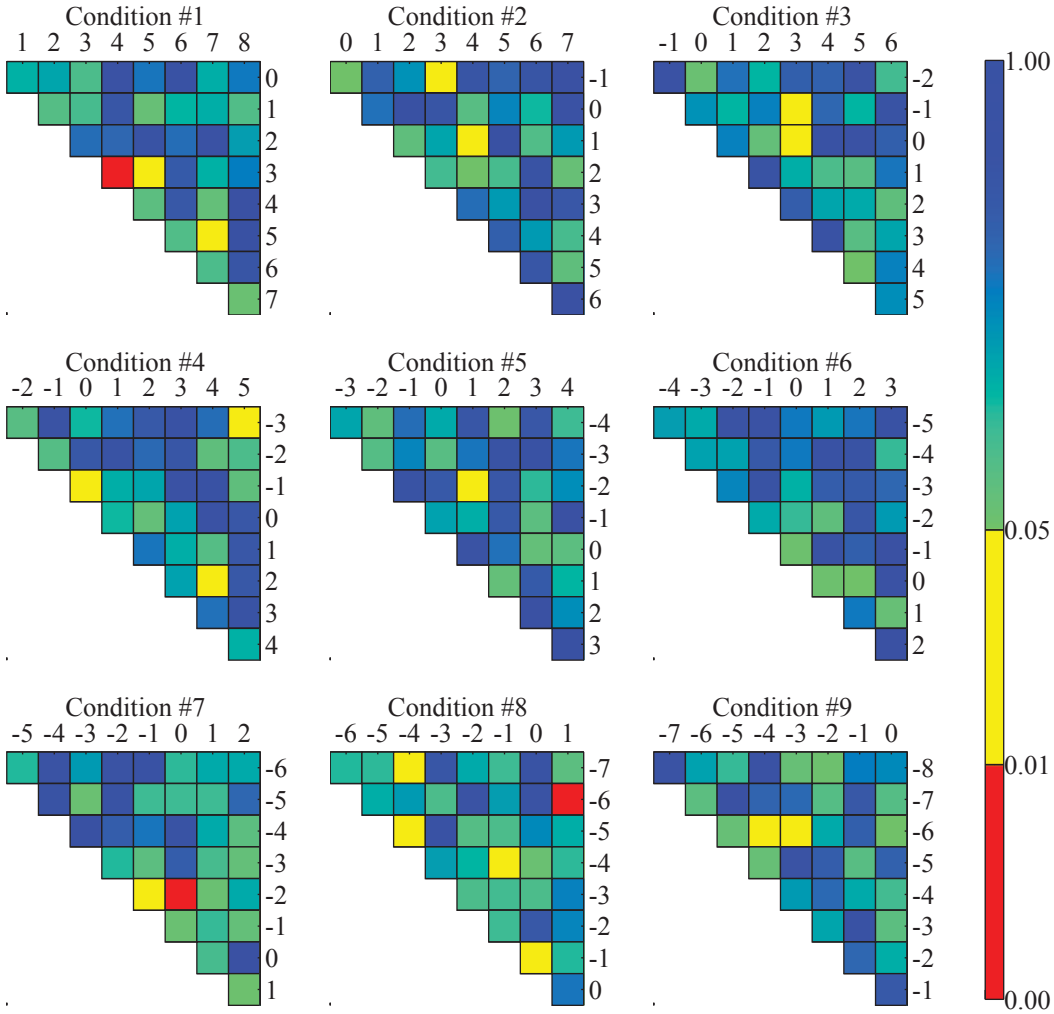


Figure 5.16: Pairwise p-values of Pearson χ^2 independence tests under condition #1 to condition #9 for subject #1 are color-coded. Yellow and Red denote significance level $p < 0.05$ and $p < 0.01$, respectively.

effect, large variance caused by heavy background noise may make the dependence very weak.

Table 5.5: Independence violation times for all the 20 subjects. 'M' and 'SD' denote mean and standard deviation, respectively.

Subject ID	Violation Times	Subject ID	Violation Times
#1	2.22	#11	2.67
#2	5.33	#12	4.11
#3	2.89	#13	1.56
#4	2.11	#14	3.89
#5	2.67	#15	2.89
#6	3.22	#16	4.33
#7	1.22	#17	4.00
#8	3.00	#18	1.78
#9	3.33	#19	2.33
#10	3.11	#20	2.67
M	2.97	**SD**	1.01

5.7 Study VI: Validation of the Proposed Trial-based ERP Classifier

This study was designated to validate the last Hypothesis that the proposed trial-based ERP classification method could outperform conventional methods. The proposed approach is expected to have good learning curve than straightforward trial-based method, since the severe "curse of dimension" issue in the conventional trial-based method. Moreover, the proposed method should be more robust on single-trial prediction, since multiple within-trial epochs and statistical language information are utilized, in contrast to single-epoch based methods.

5.7.1 Dataset Description

Totally 20 subjects were participated in this study. The 3×3 TLP was used as the stimulus paradigm. EEG data from the nine electrodes Fz, C3, Cz, C4, P3, Pz, P4, O1, and O2 were recorded for further analysis. Raw 512-Hz EEG signals were bandpass filtered (0.5 Hz–30 Hz), moving averaged, and downsampled to 20 Hz.

Calibration experimental protocol for collecting training samples was almost same as the study II. Only difference is the stimulus trial is designated to achieve classwise balanced amounts of EEG-trials for later trial-based classifier training. There were totally 30-symbol attending tasks. For each symbol, 15-trials were conducted. Therefore, totally 450 EEG-trials were recorded in the training dataset.

Testing dataset was constructed by collecting EEG data from online testing experiments. That is to let subjects to spell an English sentence with 57 symbols ("PLEASE WASH YOUR HANDS WITH WATER AND SOAP BEFORE DINNER!"). All the EEG data and spelling contexts were recorded for further analysis.

5.7.2 Pre-processing and Classifier Training

For different kinds of classification methods, training dataset was preprocessed in different ways:

1. For straightforward trial-based methods, EEG-trials from first-stimulus onset in a trial to 800 milliseconds extension after last-stimulus in the trial were extracted, totally 1,800 milliseconds length. The EEG-trials were labeled by the temporal order $(1, 2, \ldots, 9)$ of the target.

2. For state-of-the-art epoch-based methods, EEG signals were sectioned into overlapping epochs ranging from 0-ms to 800-ms poststimulus. Each of the epoch was labeled by 0 (nontarget) or 1 (target).

3. For the proposed method, epochs were extracted in same way as case 2. However, the epochs were labeled by temporal offset to the target. That means target-epochs were labeled by 0, whereas nontarget-epochs were labeled by negative/positive numbers regarding to whether they were preceding or following the target.

After preprocessing, totally 17 classifiers were trained for each of the 20 subjects:

1. Four epoch-based classifiers are normalized cross correlation (NCC), stepwise linear discriminant analysis (swLDA), linear support vector machine with recursive feature elimination (rfeLSVM$_2$), and rfeSVM with Gaussian kernel (rfeGSVM$_2$),

respectively. The four are representative: NCC is a naive method; swLDA is the state-of-the-art method; SVM with linear/Gaussian kernel was achieved best performance in BCI competitions [31]. For training the epoch-based classifiers, 450 target-epochs and 3600 nontarget-epochs are provided. Then, target and nontarget samples were balanced by randomly selecting 450 nontarget-epochs from the 3600 ones. The feature dimension is 144 (9(channel)×0.8(seconds)×20(Hz)).

2. Straightforward trial-based classifiers include 9-class SVM on raw EEG-trials (SVM_9), 9-class SVM with recursive feature elimination ($rfeSVM_9$), and 9-class SVM on epoch-projections ($pSVM_9$). Each of them have four variants (linear/Gaussian kernel × One vs. One/One vs. Rest approach). SVM with linear kernel and One vs. One approach is denoted by LOvO, similarly LOvR, GOvO, and GOvR for others. Hence, there are totally 12 distinct classifiers. The group of classifiers were designed to show the possible low learning performance which is attributed to high dimensional features and relatively small training samples. Note that same amount of training samples as for epoch-based training were provided. Because 450 labeled EEG-trials were needed to be divided into 9 parts, only 50 labeled EEG-trials for each of 9 classes were provided for training these 9-class classifiers. The EEG-trial dimension is 324 (9(channel)×1.8(seconds)×20(Hz)).

3. For the proposed method, all 450 trials of labeled epochs were used to train a projection vector by the LDA extension. Classwise one-dimensional Gaussian distributions were trained by 10-fold cross validation for the k-order models ($k = 0, 1, 2, 3, 4$), respectively. Then, subject-optimal parameters k-order and n-gram were selected by maximizing single-trial classification accuracy on the testing dataset.

5.7.3 Single-trial Classification Accuracy of the Proposed Method

In order to investigate the n-gram (language-model parameter) and k-order (ERP-model parameter) effects on classification accuracy, the proposed method with k-order

from 0 to 4 and n-gram from 0 to 9 was performed on testing database. Single-trial classification accuracy maps of all 20 subjects are depicted in Figure 5.18 and Figure 5.19. Meanwhile, the grand mean map over the 20 subjects is depicted in Figure 5.17. As expected, both n-gram and k-order have positive effects on single-trial ERP classification. That because language model with larger n-gram predicts more robust. Similarly, ERP-model with larger k-order defines finer ERP-epoch distributions. Mean single-trial accuracy with n-gram (5) and k-order (2) was about 70.00% (chance level: 11.11), in contrast to about 58.00% accuracy with n-gram (0) and k-order (0). Note that 0-order means the conventional binary ERP model, while 0-gram means no language information is used. The performances with n-gram (> 6) and k-order (> 2) were decreased that may be attributed to under-training (or over-fitting) of the models.

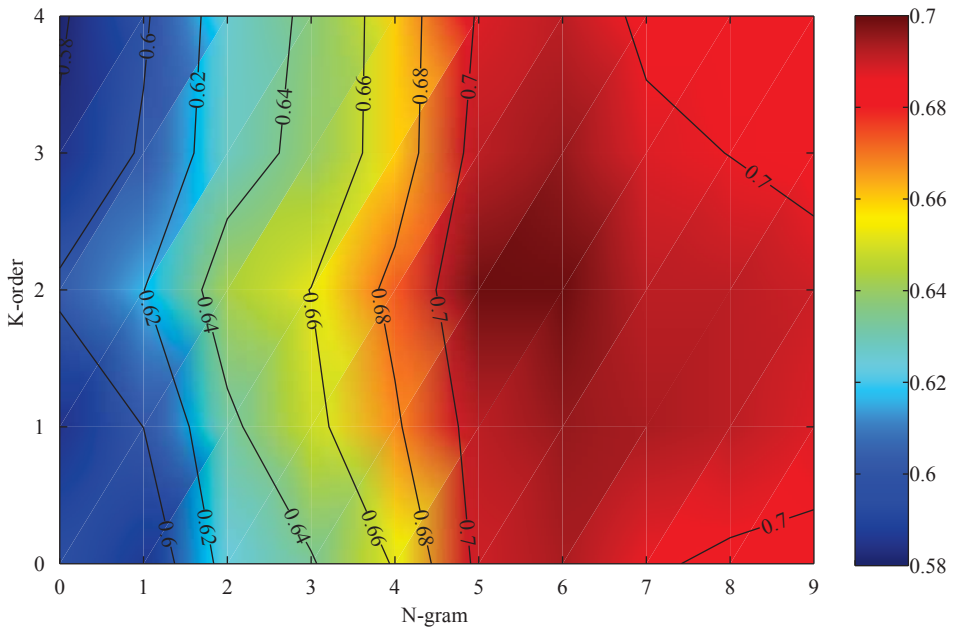


Figure 5.17: Grand mean single-trial classification accuracies against language model n-gram and ERP model k-order.

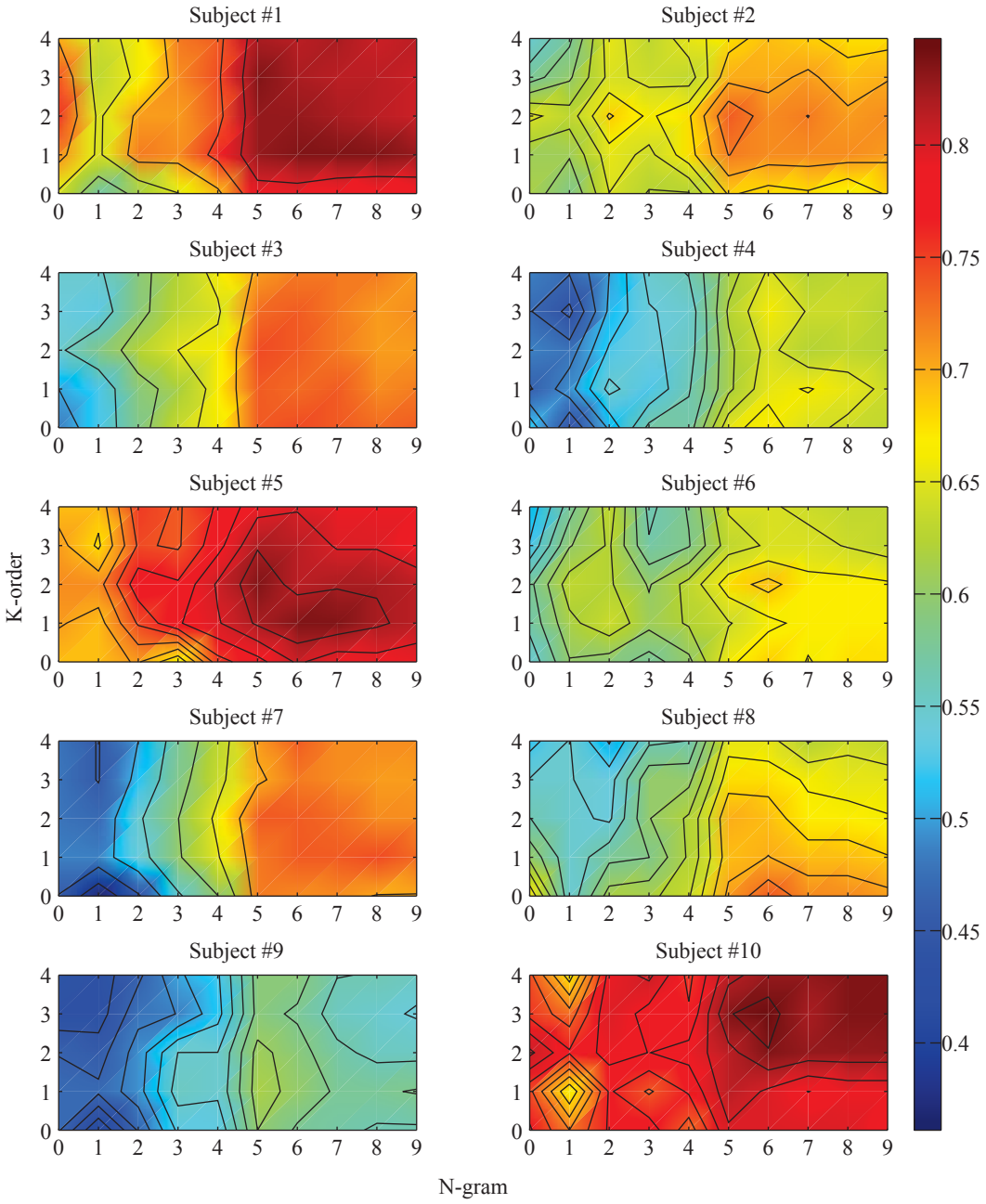


Figure 5.18: Single-trial classification accuracies against language model n-gram and ERP model k-order for subject #1–#10.

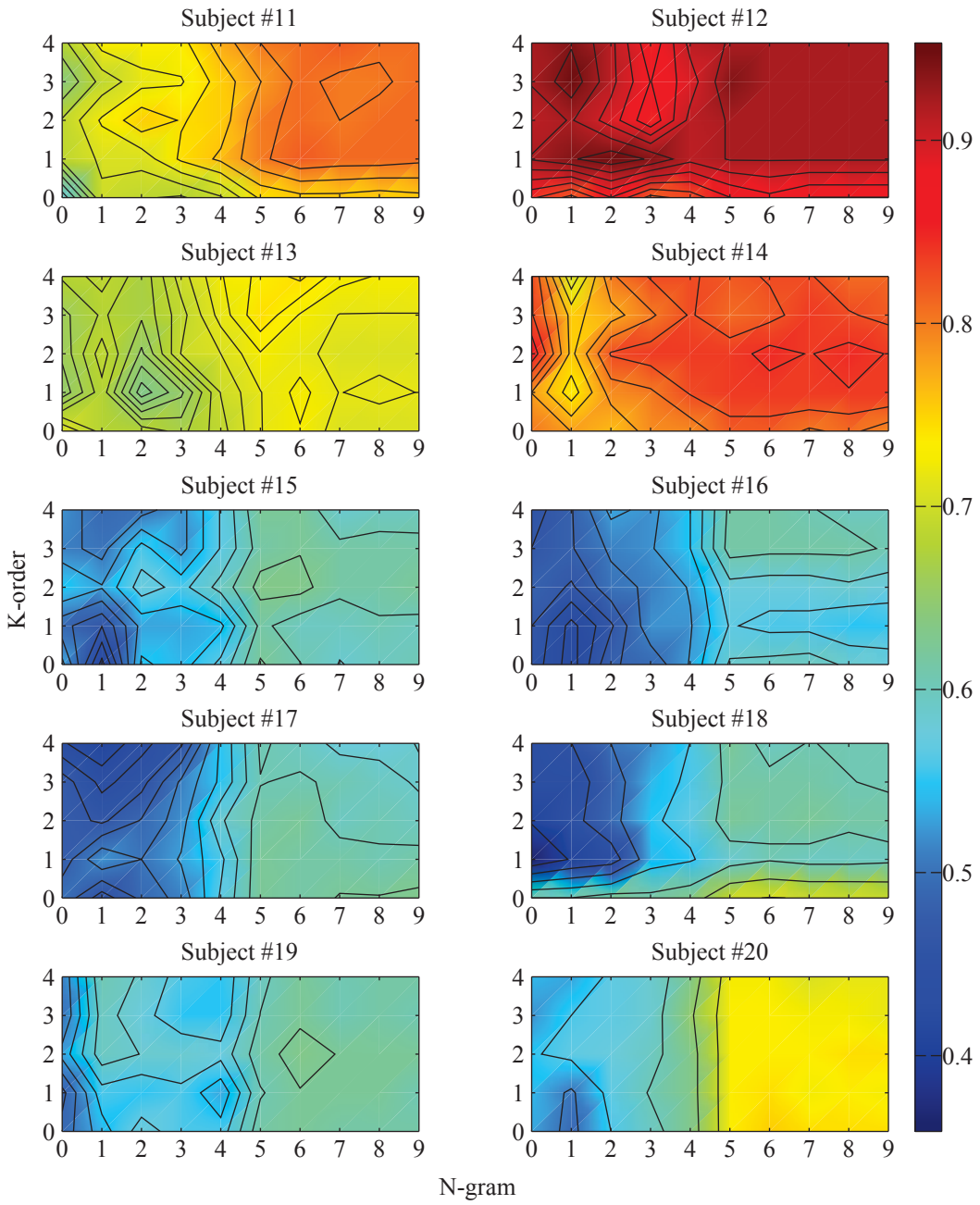


Figure 5.19: Single-trial classification accuracies against language model n-gram and ERP model k-order for subject #11–#20.

5.7.4 Proposed Method versus Conventional Ones

The ERP-decoding accuracies of 12 trial-based methods, 4 epoch-based methods, and proposed method with/without language model were performed on testing dataset for

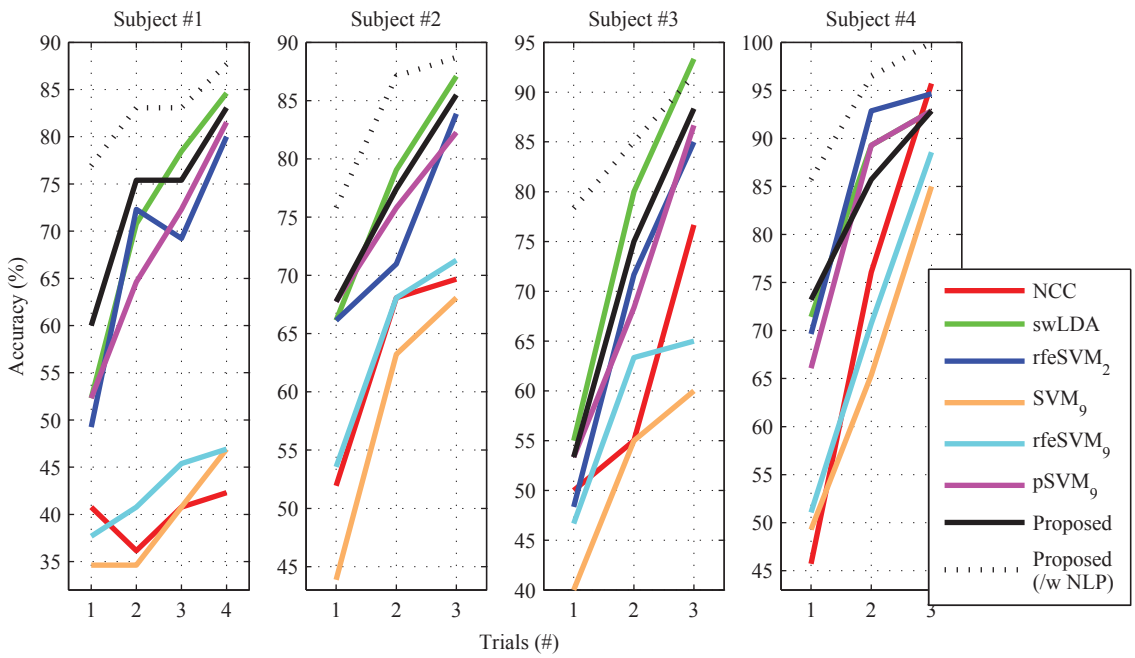


Figure 5.20: Offline accuracies against subject-specific number of trials for subject #1–#4.

each of the 20 subjects, respectively. The decoding accuracies as a function of trial number for each of 20 subjects are depicted in Figure 5.20 – 5.24. Note that only one of the SVM variants with best performance was shown here. Epoch-based rfeSVM_2 with linear kernel was selected for visualization, while trial-based SVM_9, rfeSVM_9, and pSVM_9 with linear-kernel and OvO-approach were selected as representatives.

To verify our hypothesis that the proposed method could improve ERP-decoding accuracy, we conducted an one-way repeated ANOVA with factor *Algorithm* (all 18 algorithms) for the accuracies achieved on single-trial and subject-specific all-trials, respectively.

A significant *Algorithm* effect was found for single-trial accuracies ($F = 37.04, p < 0.001$). Bonferroni *post hoc* tests revealed that single-trial accuracy for the proposed method with natural language processing (w/ NLP) (76.26%±2.01%) was significantly higher than all others, $p < 0.001$. Among the rest methods, the proposed method (w/o NLP) (67.31%±2.61%) achieved significantly better single-trial accuracies than

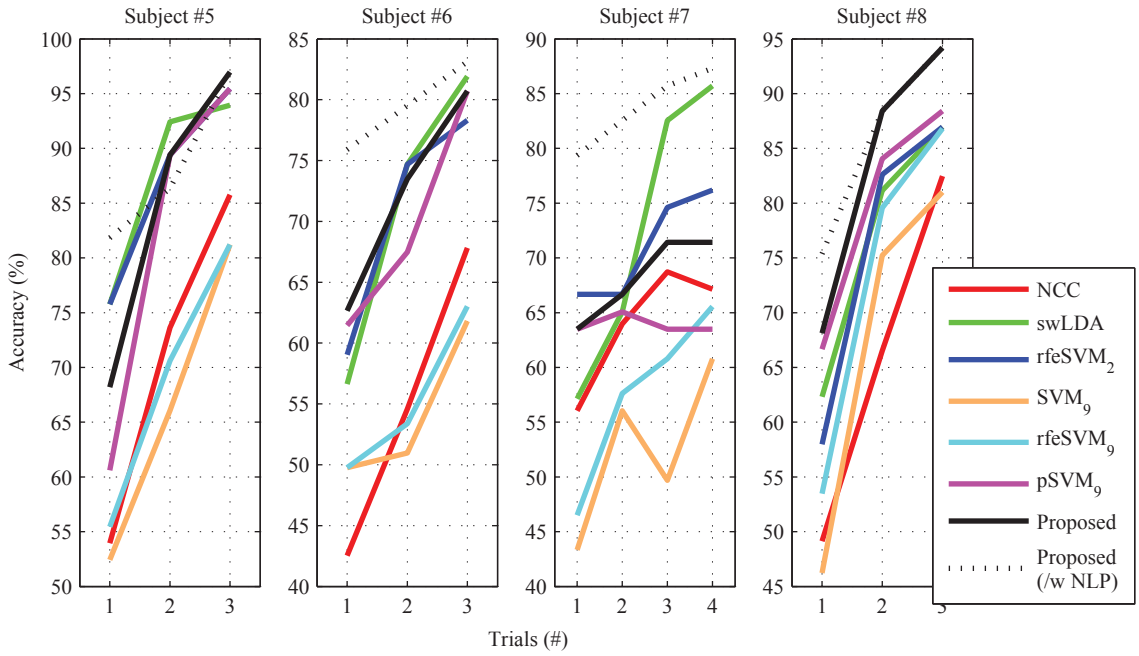


Figure 5.21: Offline accuracies against subject-specific number of trials for subject #5–#8.

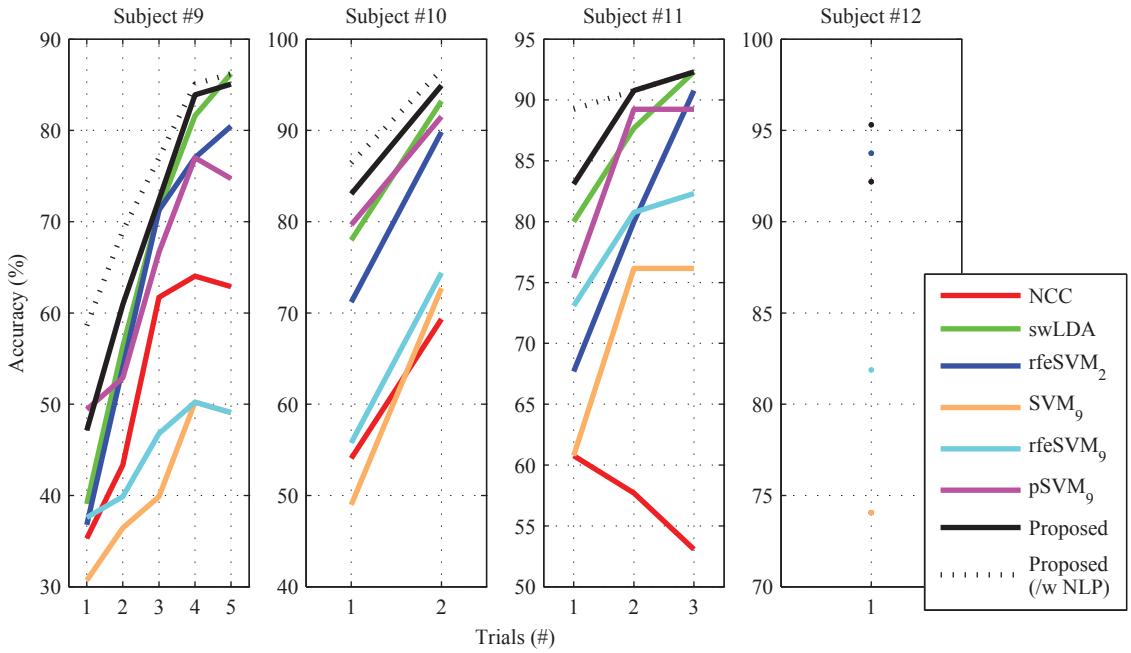


Figure 5.22: Offline accuracies against subject-specific number of trials for subject #9–#12.

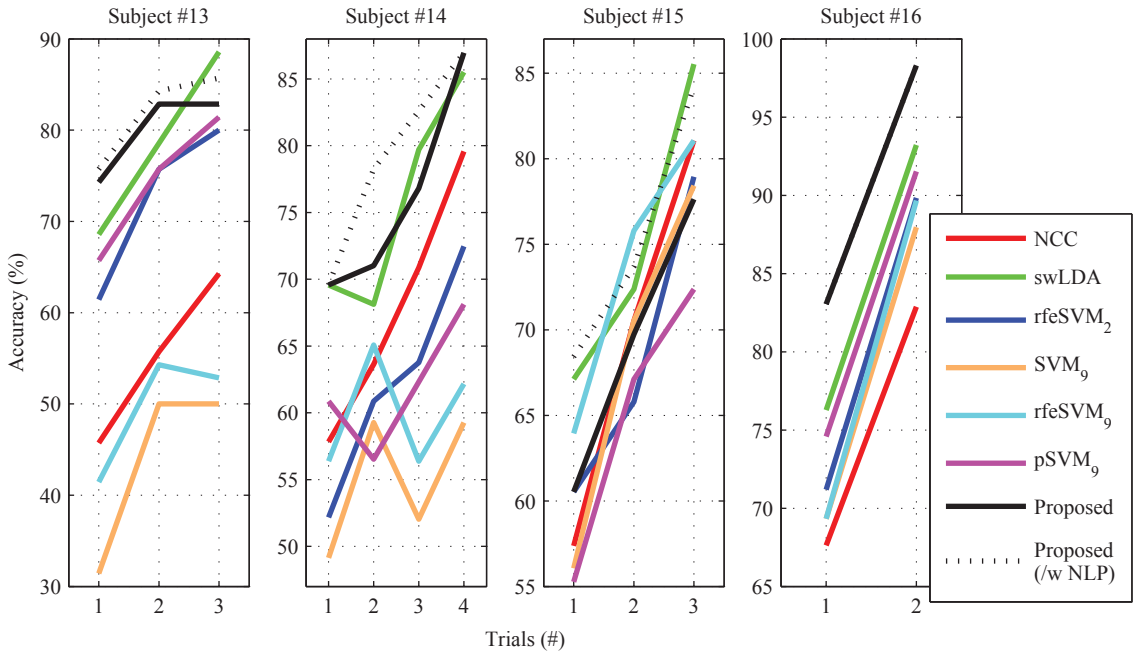


Figure 5.23: Offline accuracies against subject-specific number of trials for subject #13–#16.

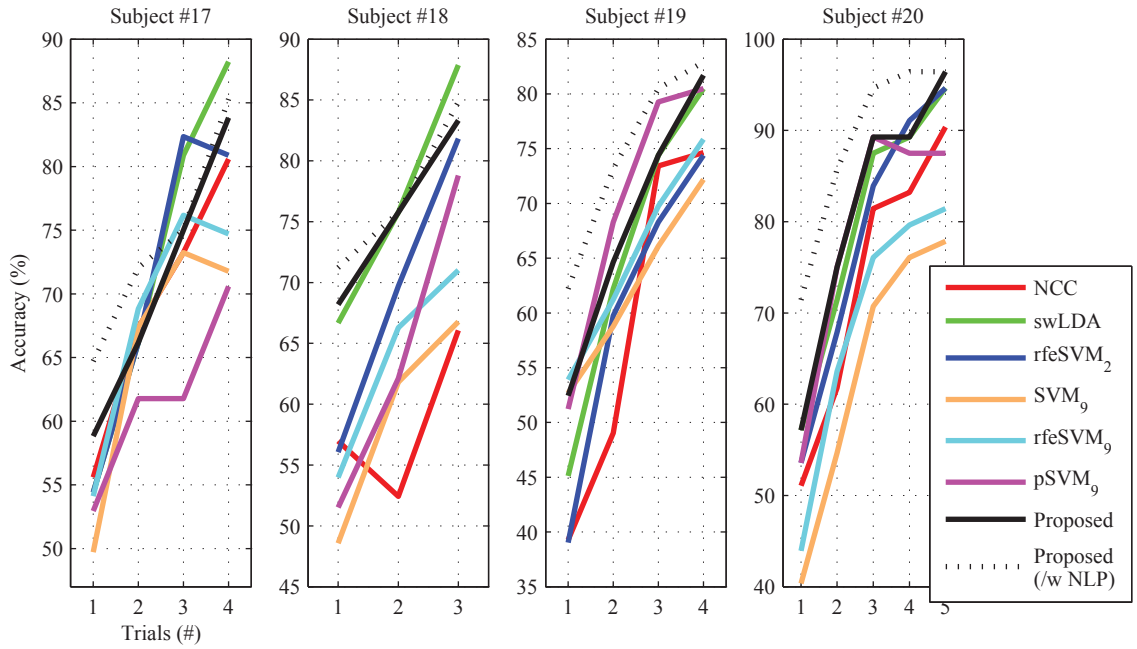


Figure 5.24: Offline accuracies against subject-specific number of trials for subject #17–#20.

all others except swLDA. Epoch-based swLDA (64.62% $\pm$ 2.90%) and rfeSVM$_2$ with linear (60.53% $\pm$ 2.93%)/Gaussian (60.70% $\pm$ 2.92%) kernel achieved no statistical different performances compared to trial-based four pSVM$_9$ variants (LOvO: 61.53%$\pm$ 2.65%, LOvR: 62.70%$\pm$2.51%, GOvO: 59.52%$\pm$3.02%, and GOvR: 58.38%$\pm$3.24%). Moreover, they were significantly better than the rest methods including epoch-based NCC and trial-based four variants of SVM$_9$ and rfeSVM$_9$. Comparisons of the rest methods revealed no significant difference.

Statistical ANOVA with Greenhouse-Geisser correction on all-trial accuracies yielded a significant *Algorithm* effect ($F = 28.13, p < 0.001$). Bonferroni *post hoc* tests revealed that all-trial accuracies for both proposed method (w/ NLP: 90.24%$\pm$1.25%, w/o NLP: 87.43% $\pm$ 1.60%) and swLDA (88.80% $\pm$ 0.96%) were significantly higher than all others, $p < 0.05$. However, there was no significant difference between the proposed methods and swLDA. Among the rest algorithms, similar comparison results were found as the single-trial case. Epoch-based NCC and trial-based algorithms except pSVM$_9$ could not compete with other epoch-based ones.

5.7.5 Error Distributions

In order to investigate the overlap effect on single-trial ERP-decoding, the ratio of nontarget-stimulus incorrectly predicted as the target against the nontarget-stimulus temporal offset to the target was calculated and visualized in Figure 5.25. Note that not all results of the 18 classifiers were shown here. We just selected 8 representative classifiers. Besides the two proposed methods, they also include epoch-based NCC, swLDA, rfeSVM$_2$ with linear/Gaussian kernel, and trial-based rfeSVM$_9$ (LOvO), pSVM$_9$ (LOvO). For a stimulus trial of the 3×3 TLP, the largest offset for nontarget-stimuli is ± 8. However, nontarget-stimuli with offsets $\geq +5$ and ≤ -5 are treated as $+5$ and -5, respectively, for calculating error-ratios in this subsection. That because there are very few nontarget-stimuli with offsets $\geq +5$ or ≤ -5. Clearly, the proposed method (especially with NLP) greatly reduced the overall error ratios. This should be

majorally attributed to the contribution from statistical language model. Moreover, lots of errors happened where nontarget-stimuli with ±1 offset for all the conventional methods, especially for the NCC and rfeSVM$_9$. That means nontarget-stimuli just preceding/following the target is prone to be wrongly selected. However, the overlap effect was reduced in the two proposed methods. This should be attributed to the proposed multi-class ERP model.

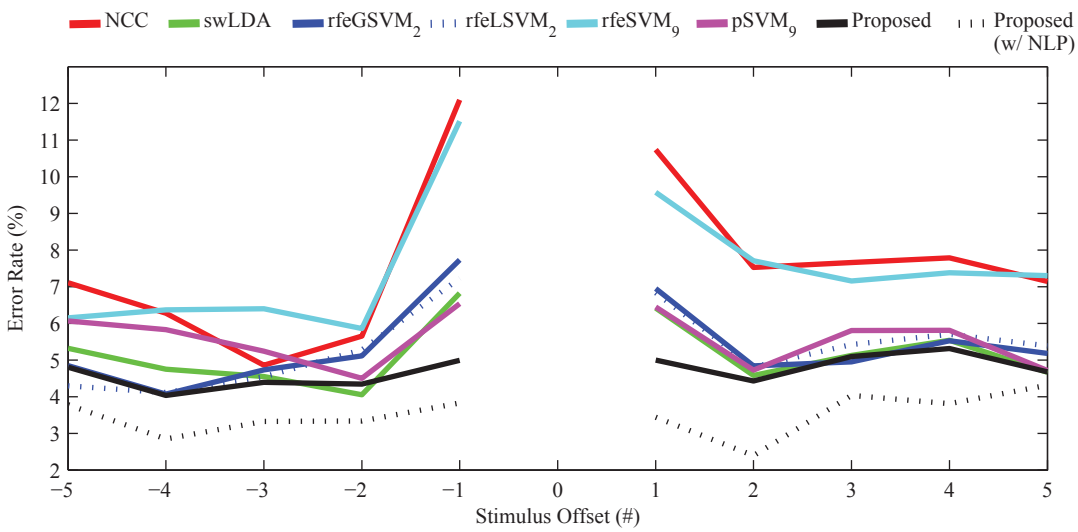


Figure 5.25: ERP-decoding error ratios (averaged over all the 20 subjects) against the stimulus temporal offset to the target. Note that 0 of horizontal axis denotes the target-stimulus. Negative values denote nontarget-stimuli preceding the target, while positive values denote nontarget-stimuli following the target. For a stimulus trial of the 3×3 TLP, the largest offset for nontarget-stimuli is ± 8. However, nontarget-stimuli with offsets $\geq +5$ and ≤ -5 are treated as $+5$ and -5, respectively.

5.7.6 Classifier Outputs

In order to reveal the underlying reasons that why the proposed methods outperformed those conventional ones, we investigated the classifier outputs against the stimulus temporal offset to the target. Note that the stimulus-offset range setting was same as the setting in the above error distribution. Moreover, since heterogeneous classifiers used in this study, their outputs are with distinct range. Therefore, all

the outputs were scaled to same range. Mean and standard deviation of classifier outputs as a function of stimulus offset for subject #1 are depicted in Figure 5.26. As expected, outputs for target-stimuli are with a larger mean; however, standard deviations across stimulus offsets are at same level, for each classifier. Standard deviations averaged over stimulus offsets for each subject and for each classifier were subjected into an one-way repeated ANOVA. The test yielded a significant *Algorithm* effect ($F = 76.54, p < 0.001$). Bonferroni *post hoc* tests found that proposed method (w/ NLP) achieved significantly smaller variance than all the others, $p < 0.001$. Moreover, proposed method (w/o NLP) was also with a significantly smaller deviation than all the conventional methods, $p < 0.001$. However, there was no significant difference between those conventional methods. This should be attributed to our trial-based strategy which is expected to enhance signal-SNR, and n-gram language model utilization.

5.7.7 Learning Curves

Actually, data used in this subsection were collected from separate (complement) experiment. Just because it logically should be placed under the Study VI. Ten subjects were recruited for this complement experiment. The experiment protocol was same as Study VI described in subsection 5.7.1. However, each subject repeated the experiment four times in order to obtain enough data. Totally, 1800 (450×4) ERP-trials were collected for each subject. Four training datasets were constructed by first-450 ERP-trials (10.13 min), first-900 ERP-trials (20.26 min), first-1350 ERP-trials (30.39 min), and the total 1800 ERP-trials (40.52 min).

In order to investigate learning ability against amount of training samples, eight representative classifiers were performed on each of the four datasets. Accuracies were estimated by 10-fold cross validation. The learning curves are depicted in Figure 5.27. The NCC curve was almost a straight line. The ERP templates estimated from the four datasets were almost same, since 10.13-min is lengthy enough for obtaining

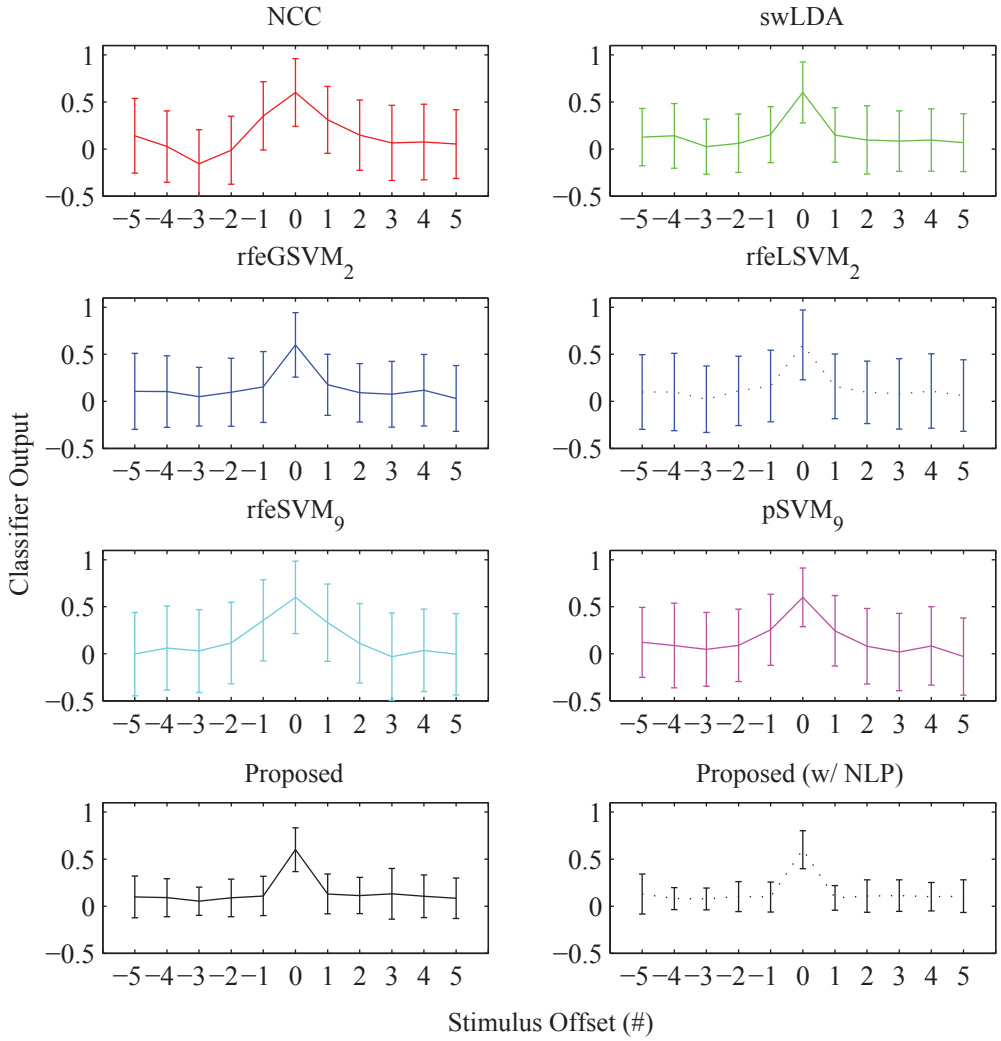


Figure 5.26: ERP-decoding outputs (subject #1) against the stimulus temporal offset to the target. Horizontal axis setting is same as the setting in Figure 5.25. Note that all these classifier outputs were scaled in same range.

a smooth ERP template. Therefore, template-based NCCs achieved almost same performances.

However, trial-based SVM_9 (LOvO) and $rfeSVM_9$ (LOvO) achieved clear growing learning curves. That means these two classifiers need more samples to be fully trained. Comparing with epoch-based $rfeSVM_2$, these two trial-based classifiers

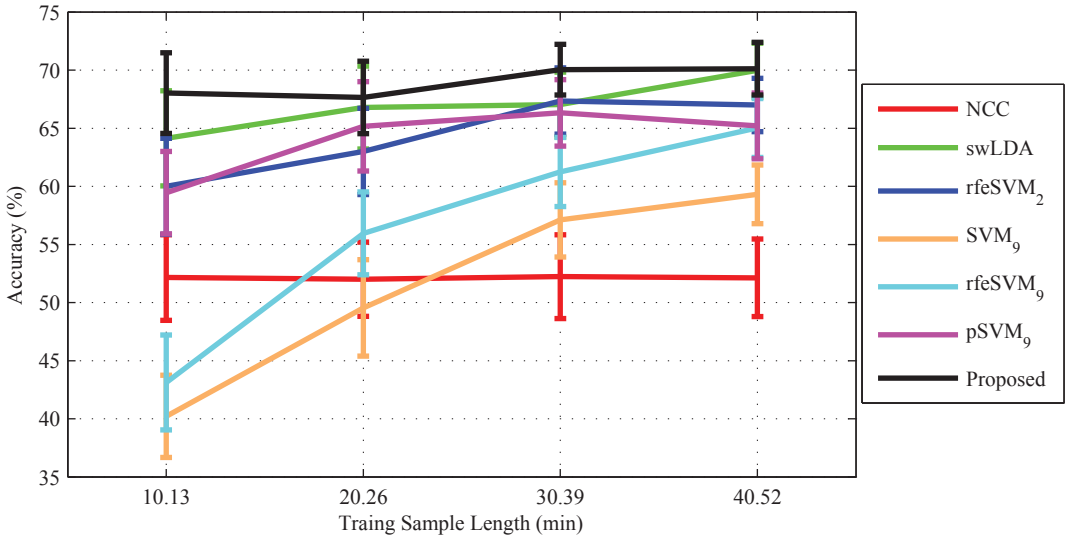


Figure 5.27: Mean and standard error (over 10 subjects) of single-trial ERP classification accuracies against training sample length (in minutes) for each of the 8 methods.

showed very low performances on the 10.13-min dataset. Although SVM is famous for insensitive to feature dimension, multi-class issue was involved here. On the one hand, 9-class task is intrinsically more difficult than binary task, because the 9-class classifier has to learn to construct a greater number of separation boundaries (OvO: $C_9^2 = 36$ and OvR: 9). The dimensionality problem may become serious. On the other hand, training samples will be divided into 9 parts that further reduces training samples.

Learning curves for other classifiers showed very slow growing trend. That means these classifiers have already been well trained by small training dataset.

5.7.8 Discussion

Totally 18 heterogeneous ERP-decoding methods were compared. The proposed method with subject-optimal k-order and n-gram significantly outperformed all other methods in terms of single-trial ERP classification accuracies. It is worth noting that subjects with low accuracies achieved by conventional methods profited more than

subjects with high accuracies from transferring to the proposed method. This might be promising for clinical application, since motor-impaired patients usually suffer from low accuracy [179]. Moreover, current gaze-independent BCIs [56, 57, 97, 98, 103] are still incomparable with the BCIs rely on overt attention in terms of accuracy, or they need more stimulus repetition. The proposed approach may improve their performances.

Why does the proposed method seems to be successfully on single-trial ERP classification? First, it reduced the overlap effect since k-order multi-class ERP model applied. Second, the classifier output based on trial-feature and statistical language information was significantly more robust. Third, the proposed method could be well trained even by limited training samples.

Chapter 6

Conclusions and Perspectives

6.1 Conclusions

Studies over the past few decades have shown that ERP-based BCIs could be a viable communication channel for the paralyzed. However, practical applications of ERP-based BCI technology that meet the needs of people with severe disabilities are significantly impeded primarily by two issue: 1. ERPs as control signals of the BCI system have an unfavorable ratio between signal and noise. 2. Current state-of-the-art ERP classification methods are still inefficient. In this dissertation, I devoted to develop an efficient ERP-based BCI, with two purposes: 1. To elicit better ERP signals by optimizing stimulus paradigm. 2. To improve single-trial ERP classification accuracy.

In THEME I, significantly better ERP signals have been elicited by the proposed 3×3 TLP paradigm. Moreover, the TLP has also been optimized based on information theory, in order to improve information transfer rate.

Firstly, we devised a 3×3 matrix paradigm with SC pattern based on initial results of spatial manipulation. On the whole, the 3×3 SC is a good candidate for further developing a BCI aiming for middle stage of ALS, because it achieved less visual fatigue, easier gaze deployment, larger ERPs, and better offline performances.

Larger ERPs might be particularly important for clinical application, because ALS patients usually suffer from ERP attenuations [69]. Besides larger target ERPs, the 3×3 SC also achieved relatively smooth background signals. These two led to a significantly high single-trial binary classification accuracy for the 3×3 SC paradigm. Up to now, we have achieved the first purpose.

Secondly, the 3×3 SC paradigm was utilized to build practical spelling applications. The 3×3 SC applying two-level strategy and incorporating a predictive language model constitutes the 3×3 TLP paradigm. Vulnerable BCI communication based on noisy EEG channels may benefit from predictive techniques. However, providing users with the possibility to select predicted words/phrases in an ERP-based BCI is a double-edged sword. It may provide powerful correct selections with time saving. Unfortunately, it may also lead to "powerful" errors with time losses. Moreover, it may introduce high mental workload [160]. Thus, we integrated a conservative letter-by-letter predictive system into the 3×3 TLP paradigm. Online validation showed that the proposed TLP paradigm achieved world-level performances [160, 161]. Moreover, most current gaze-independent BCIs [56, 97, 98, 103] exploit very limited amount of stimuli, because they require subjects to remember the feature coding. They generally also apply two-level strategies. Thus, they may also benefit from this study.

In THEME II, significantly high single-trial ERP classification accuracy has been achieved by the proposed trial-based ERP-decoding method. That because the proposed method reduced the overlap effect and with high robustness. Moreover, the method can be well trained even by limited training samples.

Learnt strengths and weaknesses of conventional methods, we proposed a trial-based ERP classification fusing statistical language information. Firstly, we proposed a k-order multi-class ERP model when we realized that an overlap effect existed in the conventional binary ERP model. Then, we reduced the feature dimension by applying a multi-class extension of LDA, in order to avoid the curse of dimension.

Finally, ERPs are decoded by fusing a n-gram language model in the maximum a posteriori (MAP) approach.

Experiment results showed that significantly higher single-trial ERP classification accuracy was achieved by the proposed method, in contrast to sixteen conventional straightforward trial-based or state-of-the-art epoch-based methods. Further investigations found that the three obstacles (low robustness, overlap effect, and under-training) preventing efficient ERP classification have been (partly) solved.

It is worth noting that subjects with low accuracy achieved by conventional methods benefited more than subjects with high accuracy when transferring to the proposed method. This might be promising for clinical application, since motor-impaired patients usually suffer from low accuracy [179]. Moreover, current gaze-independent BCIs [56, 57, 97, 98, 103] are still incomparable with the BCIs rely on overt attention in terms of accuracy, or they need more stimulus repetition. The proposed approach may improve their performances.

Finally, the major contribution of the dissertation is that a simple but reasonable modeling is very robust for the ERP decoding even with very limited calibration data. The knowledge we learnt during the study is "simple is the best". Brain is a fantastic organ and even the neuroscientists do not understand how it works. The brain signals are extremely noisy and non-stationary, therefore a simple and interpretable model is preferred than other complex and "black box" models. That because complex models are prone to overfit the limited observations, whereas every step of the simple model is interpretable and easy to control.

6.2 Perspectives

Although an efficient ERP-based BCI has been proposed and validated in this dissertation, some perspectives for future long-term research are left.

1. In Study I, we manipulated stimulus size (SS) and inter-stimulus distance (ISD), respectively. However, there may be an interaction ($SS \times ISD$) effect on ERPs.

Moreover, the two factors may have other interaction effects with other stimulus paradigm factors such as stimulation modes, etc.

2. As we discussed in subsection , the control logic for the TLP paradigm is still naive. A better control logic optimized by deep theories from other disciplines such as natural language processing (NLP) or stochastic processes may push the TLP paradigm onto a higher-level stage. There is still long way to go.

3. Multi-class ERP-epochs projected into one-dimensional space by a LDA extension stated in subsection 4.5.2 is suboptimal. Because LDA is in general unable to find the lower-dimensional feature space which maximizes the class discrimination [180]. However, it simple and easy to implement. More importantly, it achieved good results. Better results may be obtained by more advanced techniques [180, 181].

4. The most potential users of the ERP-based BCIs are neuromuscular patients. However, all the subjects participated in the experiments are young and healthy students. The great results achieved by non-patients, perhaps, can not directly used for evaluating the practical value of the BCIs. In future, collaborations with hospitals may solve the problem.

Bibliography

Bibliography

[1] Labour Japanese Ministry of Health and Welfare, "The statistics of specific disorders (als) certification by sex and prefecture", 11 2012. 1

[2] Labour Japanese Ministry of Health and Welfare, "The statistics of specific disorders (als) certification by sex and age", 2012. 1

[3] Japan Intractable Diseases Information Center, "The statistics of specific disorders certification by year", 3 2012. 1

[4] Christopher JL Murray, Dean T Jamison, Alan D Lopez, Majid Ezzati, and Colin D Mathers, *Global Burden of Disease and Risk Factors*, Washington, DC: World Bank and Oxford University Press, 2006. 1

[5] H Mitsumoto, H Mitsumoto, and FH Norris, "Classification and clinical features of amyotrophic lateral sclerosis", *Amyotrophic lateral sclerosis: a comprehensive guide to management. New York: Demos Publications*, pp. 1–19, 1994. 2

[6] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, Theresa M Vaughan, et al., "Brain-computer interfaces for communication and control", *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002. 2, 32, 34, 117

[7] Joseph N Mak and Jonathan R Wolpaw, "Clinical applications of brain-computer interfaces: current state and future prospects", *Biomedical Engineering, IEEE Reviews in*, vol. 2, pp. 187–199, 2009. 3

[8] Brendan Allison, *P3 or not P3: Toward a Better P300 BCI*, PhD thesis, University of California, San Diego, 2003. 3, 18, 46

[9] Edwin M Maynard, Craig T Nordhausen, and Richard A Normann, "The utah intracortical electrode array: A recording structure for potential brain-computer interfaces", *Electroencephalography and clinical Neurophysiology*, vol. 102, no. 3, pp. 228–239, 1997. 4

[10] DL Keene, S Whiting, and EC Ventureyra, "Electrocorticography.", *Epileptic disorders: international epilepsy journal with videotape*, vol. 2, no. 1, pp. 57, 2000. 4

[11] Hans Berger, "Über das elektrenkephalogramm des menschen", *European Archives of Psychiatry and Clinical Neuroscience*, vol. 87, no. 1, pp. 527–570, 1929. 5, 14

[12] Richard W Homan, John Herman, and Phillip Purdy, "Cerebral location of international 10–20 system electrode placement", *Electroencephalography and clinical neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987. 5

[13] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil, "Brain computer interfaces, a review", *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012. 6

[14] David Cohen, "Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents", *Science*, vol. 161, no. 3843, pp. 784–786, 1968. 5

[15] Jonathan R Wolpaw, Dennis J McFarland, Gregory W Neat, and Catherine A Forneris, "An eeg-based brain-computer interface for cursor control", *Electroencephalography and clinical neurophysiology*, vol. 78, no. 3, pp. 252–259, 1991. 7

[16] Dennis J McFarland, A Todd Lefkowicz, and Jonathan R Wolpaw, "Design and operation of an eeg-based brain-computer interface with digital signal processing

technology", *Behavior Research Methods, Instruments, & Computers*, vol. 29, no. 3, pp. 337–345, 1997.

[17] Dennis J McFarland, William A Sarnacki, and Jonathan R Wolpaw, "Electroencephalographic (eeg) control of three-dimensional movement", *Journal of Neural Engineering*, vol. 7, no. 3, pp. 036007, 2010. 7, 9

[18] Niels Birbaumer, Andrea Kubler, Nimr Ghanayim, Thilo Hinterberger, Jouri Perelmouter, Jochen Kaiser, Iver Iversen, Boris Kotchoubey, Nicola Neumann, and Herta Flor, "The thought translation device (ttd) for completely paralyzed patients", *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 2, pp. 190–193, 2000. 7, 9

[19] Eleanor A Curran and Maria 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*, vol. 51, no. 3, pp. 326–336, 2003. 7

[20] Mathew Salvaris, Caterina Cinel, Luca Citi, and Riccardo Poli, "Novel protocols for p300-based brain–computer interfaces", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 20, no. 1, pp. 8–17, 2012. 7, 19, 37, 38, 47, 49, 50, 51

[21] Gert Pfurtscheller and Christa Neuper, "Motor imagery and direct brain-computer communication", *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001. 8

[22] E. C. Lalor, S. P. Kelly, C. Finucane, R. Burke, R. Smith, R. B. Reilly, and G. McDarby, "Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment", *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 3156–3164, January 2005. 8

[23] Dennis J McFarland and Jonathan R Wolpaw, "Sensorimotor rhythm-based brain-computer interface (bci): feature selection by regression improves performance", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 13, no. 3, pp. 372–379, 2005. 8

[24] Lawrence Ashley Farwell and Emanuel Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials", *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988. 8, 9, 17, 18, 22, 23, 27, 30, 74, 79

[25] Dean J Krusienski, Eric W Sellers, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw, "Toward enhanced p300 speller performance", *Journal of neuroscience methods*, vol. 167, no. 1, pp. 15, 2008. 8, 70, 75, 80

[26] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Müller, "Single-trial analysis and classification of erp componentsa tutorial", *Neuroimage*, vol. 56, no. 2, pp. 814–825, 2011. 8, 9, 30, 70, 75, 76, 80, 90

[27] Neng Xu, Xiaorong Gao, Bo Hong, Xiaobo Miao, Shangkai Gao, and Fusheng Yang, "Bci competition 2003-data set iib: enhancing p300 wave detection using ica-based subspace projections for bci applications", *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1067–1072, 2004. 8

[28] DJ Krusienski, EW Sellers, and TM Vaughan, "Common spatio-temporal patterns for the p300 speller", in *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on*. IEEE, 2007, pp. 421–424. 8

[29] Dean J Krusienski, Eric W Sellers, François Cabestaing, Sabri Bayoudh, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw, "A comparison of classification techniques for the p300 speller", *Journal of neural engineering*, vol. 3, no. 4, pp. 299, 2006. 8, 9, 75, 80, 81, 82

[30] Shijian Lu, Cuntai Guan, and Haihong Zhang, "Unsupervised brain computer interface based on intersubject information and online adaptation", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 17, no. 2, pp. 135–145, 2009. 8

[31] Alain Rakotomamonjy and Vincent Guigue, "Bci competition iii: dataset ii-ensemble of svms for bci p300 speller", *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 3, pp. 1147–1154, 2008. 9, 75, 128

[32] Robert Leeb, Doron Friedman, Gernot R Müller-Putz, Reinhold Scherer, Mel Slater, and Gert Pfurtscheller, "Self-paced (asynchronous) bci control of a wheelchair in virtual environments: a case study with a tetraplegic", *Computational intelligence and neuroscience*, vol. 2007, 2007. 9

[33] Xiaorong Gao, Dingfeng Xu, Ming Cheng, and Shangkai Gao, "A bci-based environmental controller for the motion-disabled", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, no. 2, pp. 137–140, 2003. 9

[34] Janis J Daly, Jonathan R Wolpaw, et al., "Brain–computer interfaces in neurological rehabilitation", *Lancet neurology*, vol. 7, no. 11, pp. 1032, 2008. 9

[35] Jana I Münßinger, Sebastian Halder, Sonja C Kleih, Adrian Furdea, Valerio Raco, Adi Hösle, and Andrea Kübler, "Brain painting: first evaluation of a new brain–computer interface application with als-patients and healthy volunteers", *Frontiers in neuroscience*, vol. 4, 2010. 9

[36] Andrea Finke, Alexander Lenhardt, and Helge Ritter, "The mindgame: a p300-based brain–computer interface game", *Neural Networks*, vol. 22, no. 9, pp. 1329–1333, 2009. 9

[37] Benjamin Blankertz, Michael Tangermann, Carmen Vidaurre, Siamac Fazli, Claudia Sannelli, Stefan Haufe, Cecilia Maeder, Lenny Ramsey, Irene Sturm,

Gabriel Curio, et al., "The berlin brain–computer interface: non-medical uses of bci technology", *Frontiers in neuroscience*, vol. 4, 2010. 9

[38] Christoph Guger, Shahab Daban, Eric Sellers, Clemens Holzner, Gunther Krausz, Roberta Carabalona, Furio Gramatica, and Guenter Edlinger, "How many people are able to control a p300-based brain? computer interface (bci)?", *Neuroscience letters*, vol. 462, no. 1, pp. 94–98, 2009. 9, 21, 47, 58, 112

[39] Jun'ichi Katayama and John Polich, "Auditory and visual p300 topography from a 3 stimulus paradigm.", *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, vol. 110, no. 3, pp. 463, 1999. 14, 18

[40] Hiroshi Shibasaki and Mark Hallett, "What is the bereitschaftspotential?", *Clinical Neurophysiology*, vol. 117, no. 11, pp. 2341–2356, 2006. 14

[41] Aihide Yoshino, Masayuki Inoue, and Atsuhiro Suzuki, "A topographic electrophysiologic study of mental rotation", *Cognitive Brain Research*, vol. 9, no. 2, pp. 121–124, 2000. 14

[42] Steven J Luck, "An introduction to the event-related potential technique (cognitive neuroscience)", 2005. 14

[43] John Polich, "Updating p300: an integrative theory of p3a and p3b", *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128, 2007. 15, 17, 51, 112

[44] M Fabiani, G Gratton, D Karis, and E Donchin, "Definition, identification, and reliability of measurement of the p300 component of the event-related brain potential", *Advances in psychophysiology*, vol. 2, no. S 1, pp. 78, 1987. 15

[45] Samuel Sutton, Margery Braren, Joseph Zubin, and ER John, "Evoked-potential correlates of stimulus uncertainty", *Science*, vol. 150, no. 3700, pp. 1187–1188, 1965. 15

[46] Emanuel Donchin, WALTER Ritter, W CHEYNE McCallum, et al., "Cognitive psychophysiology: The endogenous components of the erp", *Event-related brain potentials in man*, pp. 349–411, 1978. 15

[47] Walter S Pritchard, "Psychophysiology of p300.", *Psychological bulletin*, vol. 89, no. 3, pp. 506, 1981. 15, 18

[48] Reza Fazel-Rezai, "Human error in p300 speller paradigm for brain-computer interface", in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, 2007, pp. 2516–2519. 16, 38, 48, 55

[49] Matthew A Conroy and John Polich, "Normative variation of p3a and p3b from a large sample", *Journal of Psychophysiology*, vol. 21, no. 1, pp. 22–32, 2007. 16, 17

[50] NS Dias, RM Mendes, and JH Correia, "Subject age in p300 bci", in *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*. IEEE, 2005, pp. 579–582. 16, 19

[51] John Polich, José R Criado, et al., "Neuropsychology and neuropharmacology of p3a and p3b", *International Journal of Psychophysiology*, vol. 60, no. 2, pp. 172–185, 2006. 16

[52] Fabrizio Beverina, Giorgio Palmas, Stefano Silvoni, Francesco Piccione, and Silvio Giove, "User adaptive bcis: Ssvep and p300 based interfaces", *PsychNology Journal*, vol. 1, no. 4, pp. 331–354, 2003. 17

[53] Eric W Sellers, Dean J Krusienski, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw, "A p300 event-related potential brain–computer interface (bci): the effects of matrix size and inter stimulus interval on performance", *Biological psychology*, vol. 73, no. 3, pp. 242–252, 2006. 46, 47, 48, 51, 52, 57, 112

[54] Eric W Sellers and Emanuel Donchin, "A p300-based brain–computer interface: initial tests by als patients", *Clinical neurophysiology*, vol. 117, no. 3, pp. 538–548, 2006.

[55] Dennis J McFarland, William A Sarnacki, George Townsend, Theresa Vaughan, and Jonathan R Wolpaw, "The p300-based brain–computer interface (bci): effects of stimulus rate", *Clinical Neurophysiology*, vol. 122, no. 4, pp. 731–737, 2011. 46, 47, 52

[56] Johannes Höhne, Martijn Schreuder, Benjamin Blankertz, and Michael Tangermann, "A novel 9-class auditory erp paradigm driving a predictive text entry system", *Frontiers in neuroscience*, vol. 5, 2011. 17, 36, 47, 49, 120, 123, 140, 142, 143

[57] Matthias S Treder and Benjamin Blankertz, "Research (c) overt attention and visual speller design in an erp-based brain-computer interface", 2010. 17, 36, 49, 56, 140, 143

[58] John Polich, "P300 in clinical applications: meaning, method, and measurement.", *American Journal of EEG Technology*, 1991. 17

[59] Anne-Marie Brouwer and Jan BF Van Erp, "A tactile p300 brain-computer interface", *Frontiers in neuroscience*, vol. 4, 2010. 17, 18, 36, 47, 49

[60] CONNIE C DUNCAN-JOHNSON, Walton T Roth, and Bert S Kopell, "Effects of stimulus sequence on p300 and reaction time in schizophrenics", *Annals of the New York Academy of Sciences*, vol. 425, no. 1, pp. 570–577, 1984. 18

[61] John Polich, "P300, probability, and interstimulus interval", *Psychophysiology*, vol. 27, no. 4, pp. 396–403, 1990. 18

[62] John Polich, "Task difficulty, probability, and inter-stimulus interval as determinants of p300 from auditory stimuli", *Electroencephalography and*

Clinical Neurophysiology/Evoked Potentials Section, vol. 68, no. 4, pp. 311–320, 1987. 18, 19, 49, 112

[63] Connie C Duncan-Johnson and Emanuel Donchin, "On quantifying surprise: The variation of event-related potentials with subjective probability", *Psychophysiology*, vol. 14, no. 5, pp. 456–467, 1977.

[64] John Polich, Catherine Margala, et al., "P300 and probability: comparison of oddball and single-stimulus paradigms.", *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, vol. 25, no. 2, pp. 169, 1997. 18

[65] Terence W Picton, "The p300 wave of the human event-related potential", *Journal of clinical neurophysiology*, vol. 9, no. 4, pp. 456–479, 1992. 18, 49

[66] CEM Van Beijsterveldt and GCM Van Baal, "Twin and family studies of the human electroencephalogram: a review and a meta-analysis", *Biological psychology*, vol. 61, no. 1, pp. 111–138, 2002. 18

[67] Stephanie E Eischen and John Polich, "P300 from families", *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, vol. 92, no. 4, pp. 369–372, 1994. 18

[68] Anders M Fjell and Kristine B Walhovd, "P300 and neuropsychological tests as measures of aging: scalp topography and cognitive changes", *Brain Topography*, vol. 14, no. 1, pp. 25–40, 2001. 19

[69] Hasmet A Hanagasi, I Hakan Gurvit, Numan Ermutlu, Gulustu Kaptanoglu, Sacit Karamursel, Halil A Idrisoglu, Murat Emre, and Tamer Demiralp, "Cognitive impairment in amyotrophic lateral sclerosis: evidence from neuropsychological investigation and event-related potentials", *Cognitive Brain Research*, vol. 14, no. 2, pp. 234–244, 2002. 19, 112, 142

[70] Luciana Pelosi, Maria Holly, Teresa Slade, Michael Hayward, Geoff Barrett, and Lance D Blumhardt, "Event-related potential (erp) correlates of performance of intelligence tests", *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, vol. 84, no. 6, pp. 515–520, 1992. 19

[71] Yoshio Hirayasu, Mizue Samura, Hirokazu Ohta, and Chikara Ogura, "Sex effects on rate of change of p300 latency with age", *Clinical neurophysiology*, vol. 111, no. 2, pp. 187–194, 2000. 19

[72] Daniel E Everhart, Michael D Carpenter, Joseph E Carmona, Amy J Ethridge, and Heath A Demaree, "Adult sex-related p300 differences during the perception of emotional prosody and facial affect", *Psychophysiology*, vol. 40, no. S1, pp. S39, 2003. 19

[73] Alison M Morris, Yuen So, Kathryn A Lee, Andrea A Lash, and Charles E Becker, "The p300 event-related potential: the effects of sleep deprivation", *Journal of Occupational and Environmental Medicine*, vol. 34, no. 12, pp. 1143–1152, 1992. 19

[74] Mark W Geisler and John Polich, "P300, food consumption, and memory performance", *Psychophysiology*, vol. 29, no. 1, pp. 76–85, 1992. 19

[75] AP Anokhin, AB Vedeniapin, EJ Sirevaag, LO Bauer, SJ OConnor, S Kuperman, B Porjesz, T Reich, H Begleiter, J Polich, et al., "The p300 brain potential is reduced in smokers", *Psychopharmacology*, vol. 149, no. 4, pp. 409–413, 2000. 19

[76] Andréa Camaz Deslandes, Heloisa Veiga, Maurício Cagy, Roberto Piedade, Fernando Pompeu, and Pedro Ribeiro, "Effects of caffeine on visual evoked potencial (p300) and neuromotor performance", *Arquivos de neuro-psiquiatria*, vol. 62, no. 2B, pp. 385–390, 2004.

[77] Laura Costa, Lance Bauer, Samuel Kuperman, Bernice Porjesz, Sean OConnor, Victor Hesselbrock, John Rohrbaugh, and Henri Begleiter, "Frontal p300 decrements, alcohol dependence, and antisocial personality disorder", *Biological Psychiatry*, vol. 47, no. 12, pp. 1064–1071, 2000. 19

[78] SC Kleih, F Nijboer, S Halder, and A Kübler, "Motivation modulates the p300 amplitude during brain–computer interface use", *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1023–1031, 2010. 19, 47, 50

[79] Chad E Lakey, Daniel R Berry, and Eric W Sellers, "Manipulating attention via mindfulness induction improves p300-based brain–computer interface performance", *Journal of neural engineering*, vol. 8, no. 2, pp. 025019, 2011. 19, 47, 50

[80] Edward K Vogel and Steven J Luck, "The visual n1 component as an index of a discrimination process", *Psychophysiology*, vol. 37, no. 02, pp. 190–203, 2000. 19

[81] Steven J Luck, HJ Heinze, GR Mangun, and Steven A Hillyard, "Visual event-related potentials index focused attention within bilateral stimulus arrays. ii. functional dissociation of p1 and n1 components", *Electroencephalography and clinical Neurophysiology*, vol. 75, no. 6, pp. 528–542, 1990. 20

[82] S Johannes, TF Münte, HJ Heinze, and GR Mangun, "Luminance and spatial attention effects on early visual processing", *Cognitive Brain Research*, vol. 2, no. 3, pp. 189–205, 1995. 20

[83] Steven J Luck, Geoffrey F Woodman, and Edward K Vogel, "Event-related potential studies of attention", *Trends in cognitive sciences*, vol. 4, no. 11, pp. 432–440, 2000. 20

[84] MD Rugg, AD Milner, CR Lines, and R Phalp, "Modulation of visual event-related potentials by spatial and non-spatial visual selective attention", *Neuropsychologia*, vol. 25, no. 1, pp. 85–96, 1987. 20

[85] Dan Foti, Greg Hajcak, and Joseph Dien, "Differentiating neural responses to emotional pictures: Evidence from temporal-spatial pca", *Psychophysiology*, vol. 46, no. 3, pp. 521–530, 2009. 20

[86] Alexandra Fort, Julien Besle, Marie-Hélène Giard, and Jacques Pernier, "Task-dependent activation latency in human visual extrastriate cortex", *Neuroscience letters*, vol. 379, no. 2, pp. 144–148, 2005. 21

[87] M Carrillo-de-la Peña, S Rodríguez Holguín, M Corral, and F Cadaveira, "The effects of stimulus intensity and age on visual-evoked potentials (veps) in normal children", *Psychophysiology*, vol. 36, no. 6, pp. 693–698, 1999. 21

[88] Emanuel Donchin, Kevin M Spencer, and Ranjith Wijesinghe, "The mental prosthesis: assessing the speed of a p300-based brain-computer interface", *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 2, pp. 174–179, 2000. 23, 24, 28

[89] Gerwin Schalk, "Brain–computer symbiosis", *Journal of neural engineering*, vol. 5, no. 1, pp. P1, 2008. 30

[90] DV Moretti, F Babiloni, F Carducci, F Cincotti, E Remondini, PM Rossini, S Salinari, and C Babiloni, "Computerized processing of eeg–eog–emg artifacts for multi-centric studies in eeg oscillations and event-related potentials", *International Journal of Psychophysiology*, vol. 47, no. 3, pp. 199–216, 2003. 30

[91] Kin Foon Kevin Wong, Andreas Galka, Okito Yamashita, and Tohru Ozaki, "Modelling non-stationary variance in eeg time series by state space garch

model", *Computers in biology and medicine*, vol. 36, no. 12, pp. 1327–1335, 2006. 30

[92] G Townsend, BK LaPallo, CB Boulay, DJ Krusienski, GE Frye, CK Hauser, NE Schwartz, TM Vaughan, JR Wolpaw, and EW Sellers, "A novel p300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns", *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109, 2010. 33, 38, 40, 47, 49, 55

[93] A Furdea, S Halder, DJ Krusienski, D Bross, F Nijboer, N Birbaumer, and A Kübler, "An auditory oddball (p300) spelling system for brain-computer interfaces", *Psychophysiology*, vol. 46, no. 3, pp. 617–625, 2009. 35

[94] P Brunner, S Joshi, S Briskin, JR Wolpaw, H Bischof, and G Schalk, "Does the'p300'speller depend on eye gaze?", *Journal of neural engineering*, vol. 7, no. 5, pp. 056013, 2010. 36, 49

[95] L Acqualagnav, Matthias Sebastian Treder, Martijn Schreuder, and Benjamin Blankertz, "A novel brain-computer interface based on the rapid serial visual presentation paradigm", in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 2686–2689. 36, 49

[96] Laura Acqualagna and Benjamin Blankertz, "Gaze-independent bci-spelling using rapid serial visual presentation (rsvp)", *Clinical Neurophysiology*, 2013. 47, 49

[97] Matthias Sebastian Treder, Nico Maurice Schmidt, and Benjamin Blankertz, "Gaze-independent brain–computer interfaces based on covert attention and feature attention", *Journal of neural engineering*, vol. 8, no. 6, pp. 066003, 2011. 56, 123, 140, 142, 143

[98] Fabio Aloise, Pietro Aricò, Francesca Schettini, Angela Riccio, Serenella Salinari, Donatella Mattia, Fabio Babiloni, and Febo Cincotti, "A covert attention p300-based brain–computer interface: Geospell", *Ergonomics*, vol. 55, no. 5, pp. 538–551, 2012. 123, 140, 142, 143

[99] Yang Liu, Zongtan Zhou, and Dewen Hu, "Gaze independent brain–computer speller with covert visual search tasks", *Clinical Neurophysiology*, vol. 122, no. 6, pp. 1127–1136, 2011.

[100] Gabriel Pires, Urbano Nunes, and Miguel Castelo-Branco, "Gibs block speller: Toward a gaze-independent p300-based bci", in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 6360–6364. 36

[101] Ivo Käthner, Carolin A Ruf, Emanuele Pasqualotto, Christoph Braun, Niels Birbaumer, and Sebastian Halder, "A portable auditory p300 brain–computer interface with directional cues", *Clinical Neurophysiology*, 2012. 36

[102] Johannes Höhne, Konrad Krenzlin, Sven Dähne, and Michael Tangermann, "Natural stimuli improve auditory bcis with respect to ergonomics and performance", *Journal of Neural Engineering*, vol. 9, no. 4, pp. 045003, 2012. 47, 49

[103] Martijn Schreuder, Thomas Rost, and Michael Tangermann, "Listen, you are writing! speeding up online spelling with a dynamic auditory bci", *Frontiers in neuroscience*, vol. 5, 2011. 36, 49, 120, 123, 140, 142, 143

[104] NJ Hill and Bernhard Schölkopf, "An online brain–computer interface based on shifting attention to concurrent streams of auditory stimuli", *Journal of neural engineering*, vol. 9, no. 2, pp. 026011, 2012. 36, 47

[105] Haihong Zhang, Cuntai Guan, and Chuanchu Wang, "Asynchronous p300-based brain–computer interfaces: A computational approach with statistical

models", *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 6, pp. 1754–1763, 2008. 36

[106] Alexander Lenhardt, Matthias Kaper, and Helge J Ritter, "An adaptive p300-based online brain–computer interface", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 16, no. 2, pp. 121–130, 2008. 40

[107] Tao Liu, Leslie Goldberg, Shangkai Gao, and Bo Hong, "An online brain–computer interface using non-flashing visual evoked potentials", *Journal of neural engineering*, vol. 7, no. 3, pp. 036003, 2010.

[108] Johannes Hohne, Martijn Schreuder, Benjamin Blankertz, and Michael Tangermann, "Two-dimensional auditory p300 speller with predictive text system", in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 4185–4188. 49

[109] Jing Jin, Brendan Z Allison, Eric W Sellers, Clemens Brunner, Petar Horki, Xingyu Wang, and Christa Neuper, "An adaptive p300-based control system", *Journal of neural engineering*, vol. 8, no. 3, pp. 036006, 2011. 36

[110] Martijn Schreuder, Johannes Hohne, Matthias Treder, Benjamin Blankertz, and Michael Tangermann, "Performance optimization of erp-based bcis using dynamic stopping", in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 4580–4583. 36

[111] Martijn Schreuder, Johannes Hhne, Benjamin Blankertz, Stefan Haufe, Thorsten Dickhaus, and Michael Tangermann, "Optimizing event-related potential based braincomputer interfaces: a systematic evaluation of dynamic stopping methods", *Journal of Neural Engineering*, vol. 10, no. 3, pp. 036025, 2013. 36

[112] Rajesh C Panicker, Sadasivan Puthusserypady, and Ying Sun, "An asynchronous p300 bci with ssvep-based control state detection", *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 6, pp. 1781–1788, 2011. 37

[113] Caterina Cinel, Riccardo Poli, and Luca Citi, "Possible sources of perceptual errors in p300-based speller paradigm", *Biomedizinische technik*, vol. 49, pp. 39–40, 2004. 38, 51, 55

[114] Scott Gavett, Zachary Wygant, Setare Amiri, and Reza Fazel-Rezai, "Reducing human error in p300 speller paradigm for brain-computer interface", in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 2869–2872. 38, 55

[115] Reza Fazel-Rezai and Kamyar Abhari, "A region-based p300 speller for brain-computer interface", *Electrical and Computer Engineering, Canadian Journal of*, vol. 34, no. 3, pp. 81–85, 2009. 38, 47, 51, 56

[116] Mathew Salvaris and Francisco Sepulveda, "Perceptual errors in the farwell and donchin matrix speller", in *Neural Engineering, 2009. NER'09. 4th International IEEE/EMBS Conference on*. IEEE, 2009, pp. 275–278. 38

[117] Jane E Raymond, Kimron L Shapiro, Karen M Arnell, et al., "Temporary suppression of visual processing in an rsvp task: An attentional blink", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 18, no. 3, pp. 849–860, 1992. 38

[118] SMM Martens, NJ Hill, J Farquhar, and B Schölkopf, "Overlap and refractory effects in a brain–computer interface speller based on the visual p300 event-related potential", *Journal of neural engineering*, vol. 6, no. 2, pp. 026003, 2009. 38, 47, 48, 51

[119] Nancy G Kanwisher et al., "Repetition blindness: Type recognition without token individuation", *Cognition*, vol. 27, no. 2, pp. 117–143, 1987. 38

[120] Luca Citi, Riccardo Poli, and Caterina Cinel, "Documenting, modelling and exploiting p300 amplitude changes due to variable target delays in donchin's speller", *Journal of neural engineering*, vol. 7, no. 5, pp. 056006, 2010. 38, 47, 51, 52

[121] Päivi Majaranta, I Scott MacKenzie, Anne Aula, and Kari-Jouko Räihä, "Effects of feedback and dwell time on eye typing speed and accuracy", *Universal Access in the Information Society*, vol. 5, no. 2, pp. 199–208, 2006. 40

[122] Yu-Te Wang, Yijun Wang, and Tzyy-Ping Jung, "A cell-phone-based brain–computer interface for communication in daily life", *Journal of neural engineering*, vol. 8, no. 2, pp. 025018, 2011. 41

[123] Yu Mike Chi, Yu-Te Wang, Yijun Wang, Christoph Maier, Tzyy-Ping Jung, and Gert Cauwenberghs, "Dry and noncontact eeg sensors for mobile brain–computer interfaces", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 20, no. 2, pp. 228–235, 2012. 41

[124] Dennis J McFarland, William A Sarnacki, and Jonathan R Wolpaw, "Should the parameters of a bci translation algorithm be continually adapted?", *Journal of neuroscience methods*, vol. 199, no. 1, pp. 103–107, 2011. 41

[125] Jing Jin, Eric W Sellers, and Xingyu Wang, "Targeting an efficient target-to-target interval for p300 speller brain–computer interfaces", *Medical & biological engineering & computing*, vol. 50, no. 3, pp. 289–296, 2012. 47, 52

[126] Johannes Hohne and Michael Tangermann, "How stimulation speed affects event-related potentials and bci performance", in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 1802–1805. 47

[127] Brendan Z Allison and Jaime A Pineda, "Effects of soa and flash pattern manipulations on erps, performance, and preference: implications for a bci system", *International journal of psychophysiology*, vol. 59, no. 2, pp. 127–140, 2006. 46, 47, 52

[128] Yu Zhang, Qibin Zhao, Jing Jin, Xingyu Wang, and Andrzej Cichocki, "A novel bci based on erp components sensitive to configural processing of human faces", *Journal of neural engineering*, vol. 9, no. 2, pp. 026018, 2012. 47, 48

[129] Qibin Zhao, Akinari Onishi, Yu Zhang, Jianting Cao, Liqing Zhang, and Andrzej Cichocki, "A novel oddball paradigm for affective bcis using emotional faces as stimuli", in *Neural Information Processing*. Springer, 2011, pp. 279–286.

[130] Qibin Zhao, Yu Zhang, Akinari Onishi, and Andrzej Cichocki, "An affective bci using multiple erp components associated to facial emotion processing", in *Brain-Computer Interface Research*, Christoph Guger, Brendan Z. Allison, and Gnter Edlinger, Eds., SpringerBriefs in Electrical and Computer Engineering, pp. 61–72. Springer Berlin Heidelberg, 2013. 48

[131] Jing Jin, Brendan Z Allison, Tobias Kaufmann, Andrea Kübler, Yu Zhang, Xingyu Wang, and Andrzej Cichocki, "The changing face of p300 bcis: a comparison of stimulus changes in a p300 bci involving faces, emotion, and movement", *PloS one*, vol. 7, no. 11, pp. e49688, 2012. 47

[132] Bo Hong, Fei Guo, Tao Liu, Xiaorong Gao, and Shangkai Gao, "N200-speller using motion-onset visual response", *Clinical neurophysiology*, vol. 120, no. 9, pp. 1658–1666, 2009. 47, 48

[133] Sulamith Schaeff, Matthias Sebastian Treder, Bastian Venthur, and Benjamin Blankertz, "Exploring motion veps for gaze-independent communication", *Journal of Neural Engineering*, vol. 9, no. 4, pp. 045006, 2012.

[134] Jing Jin, Brendan Z Allison, Xingyu Wang, and Christa Neuper, "A combined brain–computer interface based on p300 potentials and motion-onset visual evoked potentials", *Journal of neuroscience methods*, vol. 205, no. 2, pp. 265–276, 2012.

[135] Yang Liu, Zongtan Zhou, and Dewen Hu, "Comparison of stimulus types in visual p300 speller of brain-computer interfaces", in *Cognitive Informatics (ICCI), 2010 9th IEEE International Conference on*. IEEE, 2010, pp. 273–279. 47, 48

[136] Michael Tangermann, Martijn Schreuder, Sven Dähne, Johannes Höhne, Sebastian Regler, Andrew Ramsay, Melissa Quek, John Williamson, and Roderick Murray-Smith, "Optimized stimulation events for a visual erp bci", *Int J Bioelectromagnetism Volume*, vol. 13, no. 3, pp. 119–120, 2011. 47

[137] Michael Tangermann, Johannes Höhne, Martijn Schreuder, Max Sagebaum, Benjamin Blankertz, Andrew Ramsay, and Roderick Murray-Smith, "Data driven neuroergonomic optimization of bci stimuli", 2011. 47, 48

[138] James W Covington and John Polich, "P300, stimulus intensity, and modality", *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, vol. 100, no. 6, pp. 579–584, 1996. 47, 48

[139] Kouji Takano, Tomoaki Komatsu, Naoki Hata, Yasoichi Nakajima, and Kenji Kansaku, "Visual stimuli for the p300 brain–computer interface: a comparison of white/gray and green/blue flicker matrices", *Clinical neurophysiology*, vol. 120, no. 8, pp. 1562–1566, 2009. 47, 48

[140] DS Klobassa, TM Vaughan, P Brunner, NE Schwartz, JR Wolpaw, C Neuper, and EW Sellers, "Toward a high-throughput auditory p300-based brain-computer interface", *Clinical neurophysiology: official journal of the*

International Federation of Clinical Neurophysiology, vol. 120, no. 7, pp. 1252, 2009. 47

[141] Febo Cincotti, "Tactile, visual, and bimodal p300s: Could bimodal p300s boost bci performance?", *SRX Neuroscience*, vol. 2010, 2010. 47

[142] George Townsend, Jessica Shanahan, David B Ryan, and Eric W Sellers, "A general p300 brain-computer interface presentation paradigm based on performance guided constraints", *Neuroscience Letters*, 2012. 47, 49

[143] Jing Jin, Brendan Z Allison, Eric W Sellers, Clemens Brunner, Petar Horki, Xingyu Wang, and Christa Neuper, "Optimized stimulus presentation patterns for an event-related potential eeg-based brain–computer interface", *Medical & biological engineering & computing*, vol. 49, no. 2, pp. 181–191, 2011. 47, 49

[144] Cuntai Guan, Manoj Thulasidas, and Jiankang Wu, "High performance p300 speller for brain-computer interface", in *Biomedical Circuits and Systems, 2004 IEEE International Workshop on*. IEEE, 2004, pp. S3–5. 47, 58, 112

[145] Eric W Sellers, Andrea Kubler, and Emanuel Donchin, "Brain-computer interface research at the university of south florida cognitive psychophysiology laboratory: the p300 speller", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 221–224, 2006. 47, 52, 57

[146] Brendan Z Allison and Jaime A Pineda, "Erps evoked by different matrix sizes: implications for a brain computer interface (bci) system", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, no. 2, pp. 110–113, 2003. 47, 48, 51, 52, 112

[147] Reza Fazel-Rezai and Kamyar Abhari, "A comparison between a matrix-based and a region-based p300 speller paradigms for brain-computer interface", in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE, 2008, pp. 1147–1150. 47, 48, 51

[148] M Salvaris and F Sepulveda, "Visual modifications on the p300 speller bci paradigm", *Journal of neural engineering*, vol. 6, no. 4, pp. 046011, 2009. 47, 51, 52, 53

[149] Yohei Sakai and Tohru Yagi, "Alphabet matrix layout in p300 speller may alter its performance", in *Biomedical Engineering International Conference (BMEiCON), 2011*. IEEE, 2012, pp. 89–92. 47, 51, 52

[150] Michael Tangermann, Johannes Hohne, Heiko Stecher, and Martijn Schreuder, "No surprise-fixed sequence event-related potentials for brain-computer interfaces", in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 2501–2504. 47, 49

[151] Manson Cheuk-Man Fong, James William Minett, Thierry Blu, and William Shi-Yuan Wang, "Brain–computer interface (bci): is it strictly necessary to use random sequences in visual spellers?", in *Proceedings of the 10th asia pacific conference on Computer human interaction*. ACM, 2012, pp. 109–118. 47, 49

[152] Elizabeth A Felton, Justin C Williams, Gregg C Vanderheiden, and Robert G Radwin, "Mental workload during brain–computer interface training", *Ergonomics*, vol. 55, no. 5, pp. 526–537, 2012. 47, 50

[153] Honglai Xu, Dan Zhang, Minhui Ouyang, and Bo Hong, "Employing an active mental task to enhance the performance of auditory attention-based brain–computer interfaces", *Clinical Neurophysiology*, 2012. 47, 50

[154] Jing Guo, Shangkai Gao, and Bo Hong, "An auditory brain–computer interface using active mental response", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 18, no. 3, pp. 230–235, 2010. 47, 50

[155] Craig J Gonsalvez and John Polich, "P300 amplitude is determined by target-to-target interval", *Psychophysiology*, vol. 39, no. 3, pp. 388–396, 2002. 51, 112

[156] GE Frye, CK Hauser, G Townsend, and EW Sellers, "Suppressing flashes of items surrounding targets during calibration of a p300-based brain–computer interface improves performance", *Journal of neural engineering*, vol. 8, no. 2, pp. 025024, 2011. 52

[157] Jack T Holladay, "Proper method for calculating average visual acuity", *Journal of Refractive Surgery*, vol. 13, pp. 388–391, 1997. 53

[158] GERALD Westheimer, "Visual acuity", *Annual Review of Psychology*, vol. 16, no. 1, pp. 359–380, 1965. 53

[159] Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Michael Schröder, John Williamson, Roderick Murray-Smith, and Klaus-Robert Müller, "The berlin brain-computer interface presents the novel mental typewriter hex-o-spell.", 2006. 56

[160] DB Ryan, GE Frye, G Townsend, DR Berry, S Mesa-G, NA Gates, and EW Sellers, "Predictive spelling with a p300-based brain–computer interface: Increasing the rate of communication", *Intl. Journal of Human–Computer Interaction*, vol. 27, no. 1, pp. 69–84, 2010. 59, 60, 120, 122, 142

[161] Tobias Kaufmann, Stefan Völker, Laura Gunesch, and Andrea Kübler, "Spelling is just a click away–a user-centered brain–computer interface including auto-calibration and predictive text entry", *Frontiers in Neuroscience*, vol. 6, 2012. 59, 61, 120, 142

[162] WJ Teahan, "Probability estimation for ppm", in *In Proceedings NZCSRSC'95. Available from http://www. cs. waikato. ac. nz/wjt.* Citeseer, 1995. 60

[163] Heidi Horstmann Koester and Simon P Levine, "Learning and performance of able-bodied individuals using scanning systems with and without word prediction", *Assistive Technology*, vol. 6, no. 1, pp. 42–53, 1994. 65

[164] Horabail Venkatagiri, "Effect of window size on rate of communication in a lexical prediction aac system", *Augmentative and Alternative Communication*, vol. 10, no. 2, pp. 105–112, 1994. 65

[165] Joseph N Mak, Dennis J McFarland, Theresa M Vaughan, Lynn M McCane, Phillippa Z Tsui, Debra J Zeitlin, Eric W Sellers, and Jonathan R Wolpaw, "Eeg correlates of p300-based brain–computer interface (bci) performance in people with amyotrophic lateral sclerosis", *Journal of neural engineering*, vol. 9, no. 2, pp. 026014, 2012. 70

[166] Ryan Rifkin, Sayan Mukherjee, Pablo Tamayo, Sridhar Ramaswamy, Chen-Hsiang Yeang, Michael Angelo, Michael Reich, Tomaso Poggio, Eric S Lander, Todd R Golub, et al., "An analytical method for multiclass molecular cancer classification", *Siam Review*, vol. 45, no. 4, pp. 706–723, 2003. 73

[167] Xin Zhou and David P Tuck, "Msvm-rfe: extensions of svm-rfe for multiclass gene selection on dna microarray data", *Bioinformatics*, vol. 23, no. 9, pp. 1106–1114, 2007. 73

[168] F Aloise, F Schettini, P Aricò, S Salinari, F Babiloni, and F Cincotti, "A comparison of classification techniques for a gaze-independent p300-based brain–computer interface", *Journal of Neural Engineering*, vol. 9, no. 4, pp. 045012, 2012. 75, 80, 82

[169] Nikolay V Manyakov, Nikolay Chumerin, Adrien Combaz, and Marc M Van Hulle, "Comparison of classification methods for p300 brain-computer interface on disabled subjects", *Computational intelligence and neuroscience*, vol. 2011, pp. 2, 2011. 82

[170] Steven Lemm, Benjamin Blankertz, Thorsten Dickhaus, and Klaus-Robert Müller, "Introduction to machine learning for brain imaging", *Neuroimage*, vol. 56, no. 2, pp. 387–399, 2011. 75

[171] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, Bruno Arnaldi, et al., "A review of classification algorithms for eeg-based brain–computer interfaces", *Journal of neural engineering*, vol. 4, 2007. 75, 82

[172] Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens, "An efficient p300-based brain–computer interface for disabled subjects", *Journal of Neuroscience methods*, vol. 167, no. 1, pp. 115–125, 2008. 75

[173] Jason Farquhar and N Jeremy Hill, "Interactions between pre-processing and classification methods for event-related-potential classification", *Neuroinformatics*, pp. 1–18, 2012. 75, 80

[174] Ryota Tomioka, Kazuyuki Aihara, and Klaus-Robert Müller, "Logistic regression for single trial eeg classification", *Advances in neural information processing systems*, vol. 19, pp. 1377–1384, 2007. 75

[175] Garett D Johnson and Dean J Krusienski, "Ensemble swlda classifiers for the p300 speller", in *Human-Computer Interaction. Novel Interaction Methods and Techniques*, pp. 551–557. Springer, 2009. 75, 80

[176] Hubert Cecotti and Axel Graser, "Convolutional neural networks for p300 detection with application to brain-computer interfaces", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 3, pp. 433–445, 2011. 75, 82

[177] Paul L Nunez, *Electric fields of the brain: the neurophysics of EEG*, Oxford University Press, 2006. 75, 76

[178] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw, "Bci2000: a general-purpose brain-computer interface

(bci) system", *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1034–1043, 2004. 81, 95

[179] Rupert Ortner, Fabio Aloise, Robert Prückl, Francesca Schettini, Veronika Putz, Josef Scharinger, Eloy Opisso, Ursula Costa, and Christoph Guger, "Accuracy of a p300 speller for people with motor impairments: a comparison", *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 214–218, 2011. 140, 143

[180] Marco Loog, Reinhold Haeb-Umbach, et al., "Multi-class linear dimension reduction by generalized fisher criteria.", in *INTERSPEECH*, 2000, pp. 1069–1072. 144

[181] Hyunsoo Kim, Barry L Drake, and Haesun Park, "Multiclass classifiers based on dimension reduction with generalized lda", *Pattern recognition*, vol. 40, no. 11, pp. 2939–2945, 2007. 144

Appendix

Appendix A

Event-related potentials (ERPs)

A.1 ERP responses of 8×8 RC, 3×3 RC, and 3×3 SC at all 9 electrodes

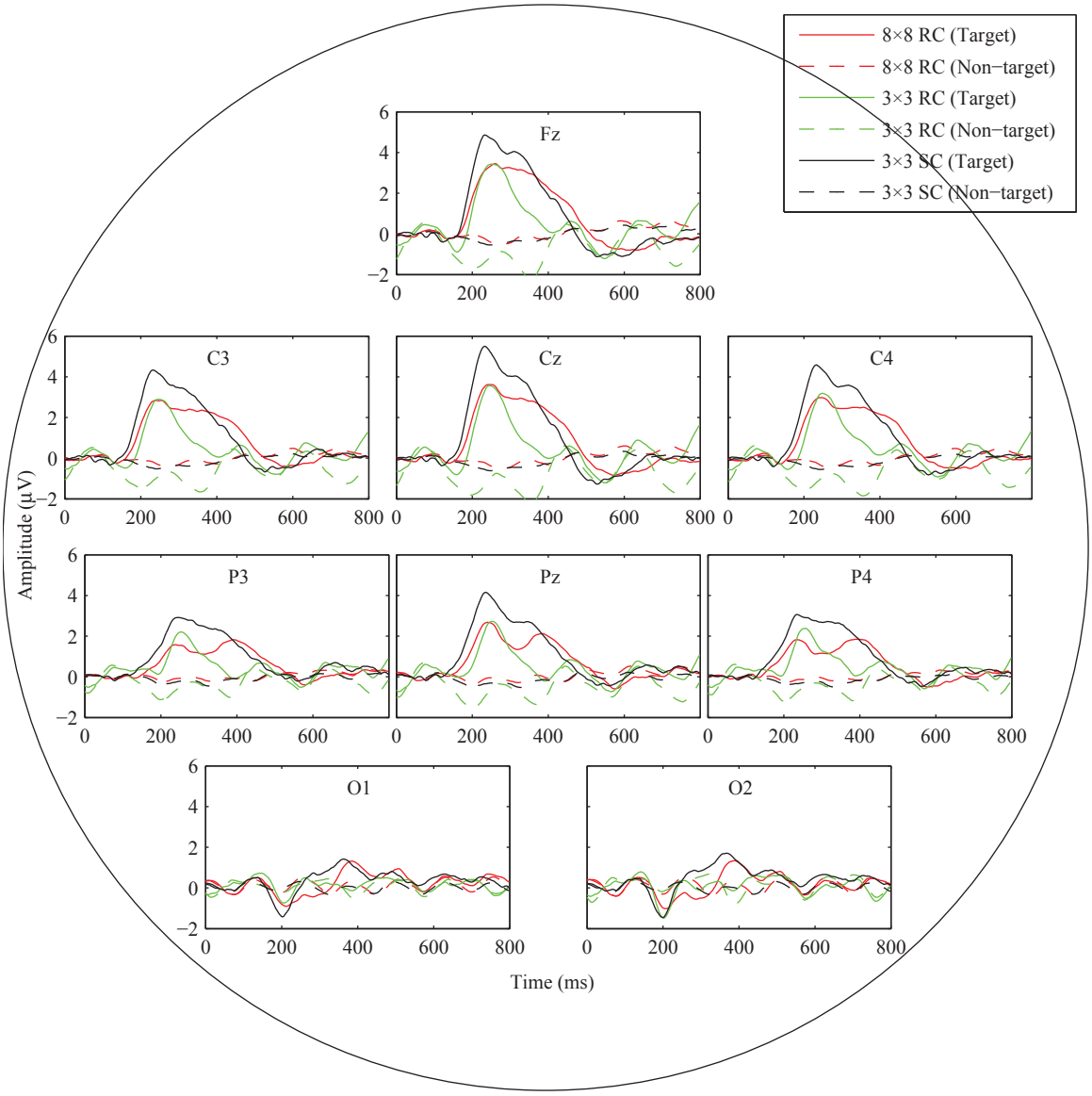


Figure A.1: Grand mean waveforms of Target (solid) and Non-target (dashed) responses for the 8×8 RC (red), 3×3 RC (green), and 3×3 SC (black) at electrodes Fz, C3, Cz, C4, P3, Pz, P4, O1, and O2.

A.2 ERP responses of *Step-1*, *Step-2*, and *Step-p* at all 9 electrodes

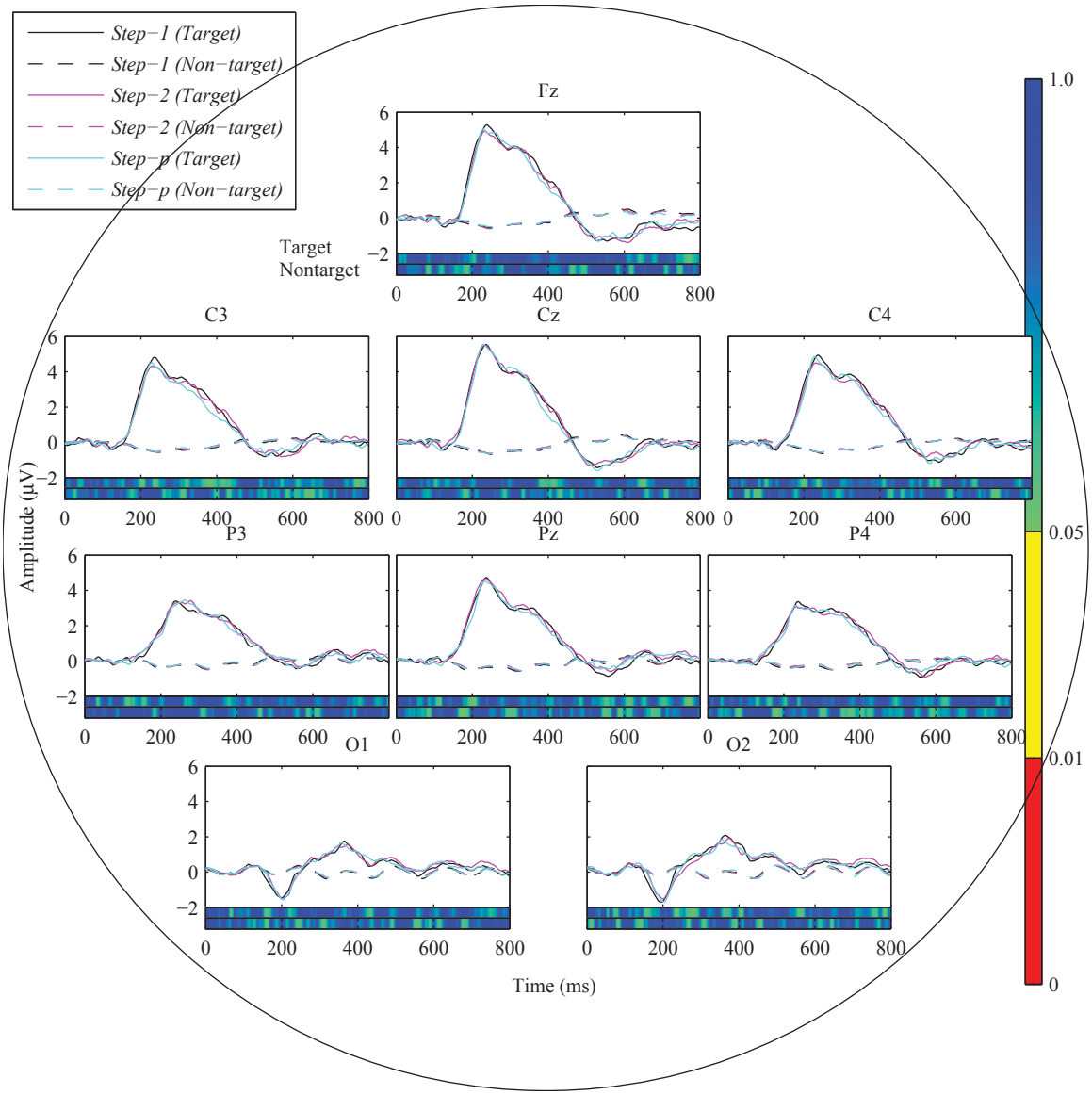


Figure A.2: Grand mean waveforms for *Step-1* (black), *Step-2* (pink), and *Step-p* (cyan) of Target (solid) and Non-target (dashed) at electrodes Fz, C3, Cz, C4, P3, Pz, P4, O1, and O2. Color-bars below each waveform are coded by sample-wised p-values of ANOVAs conducted between *Step-1*, *-2*, and *-p* for Target and Non-target responses, respectively. No significant difference is found between the three steps.

Appendix B

Questionnaire

NO.: _____ Age: _____ Gender: _____

1. Counting Tasks

8x8 RC:

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10

3x3 RC:

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10

3x3 SC:

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10

2. Visual Fatigue

8x8 RC: ( )
0 ---------- 10
Not at all very tiring

3x3 RC: ( )
0 ---------- 10
Not at all very tiring

3x3 SC: ( )
0 ---------- 10
Not at all very tiring

3. Concentration

8x8 RC: ( )
0 ---------- 10
Not at all very distracting

3x3 RC: ( )
0 ---------- 10
Not at all very distracting

3x3 SC: ( )
0 ---------- 10
Not at all very distracting

4. Preference

☐ 8x8 RC ☐ 3x3 TLP

Figure B.1: A questionnaire snapshot used in Study II and Study III. Subjects are asked to fill in part 2, 3, and 4. Other blanks are filled by experiment investigator.

Appendix C

Validation of Gaussian Normality of ERP Model Proposed in Chapter 4

Although the Gaussian normality can be simply explained by the law of large numbers, we validated the normality in this section. The proposed multi-class ERP models were trained in Study VI conducted in section 5.7.

In order to validate the normality of the ERP models for all the 20 subjects, Kolmogorov-Smirnov (KS) tests were conducted on classwise epoch-projections. No KS-test result can reject the null-hypothesis that the input variable fulfills Gaussian distribution. All p-values are shown in Table C.1.

Table C.1: p-values of Kolmogorov-Smirnov (KS) tests for validating normality of the proposed multi-class ERP model

Subject	***p*-value of Kolmogorov-Smirnov test**								
	-4N	**-3N**	**-2N**	**-1N**	**T**	**+1N**	**+2N**	**+3N**	**N**
#1	0.60	0.65	0.77	0.32	0.69	0.81	0.83	0.16	0.31
#2	0.44	0.98	0.98	0.47	0.72	0.93	0.57	0.99	0.98
#3	0.82	0.94	0.81	0.92	0.62	0.46	0.71	0.15	0.81
#4	0.94	0.66	0.87	0.55	0.96	0.68	0.88	0.46	0.96
#5	0.46	0.99	0.26	0.85	0.51	0.78	0.94	0.71	0.82
#6	0.98	0.88	0.76	0.30	0.33	0.76	0.95	0.64	0.67
#7	0.81	0.91	0.30	0.82	0.44	0.79	0.50	0.80	0.49
#8	0.98	0.80	0.86	0.94	0.75	0.66	0.93	0.51	0.85
#9	0.90	0.56	0.99	0.99	0.94	0.86	0.55	0.95	0.98
#10	0.89	0.74	0.80	0.93	0.62	0.68	0.78	0.85	0.72
#11	0.43	0.97	0.83	0.87	0.74	0.89	0.99	0.93	0.67
#12	0.80	0.95	0.95	0.36	0.92	0.88	0.94	0.54	0.92
#13	0.84	0.96	0.65	0.78	0.30	0.87	0.99	0.79	0.43
#14	0.59	0.77	0.98	0.78	0.55	0.71	0.65	0.59	0.60
#15	0.70	0.87	0.95	0.86	0.84	0.33	0.57	0.67	0.99
#16	0.59	1.00	0.74	0.57	0.09	0.63	0.89	0.91	0.56
#17	0.72	0.63	0.80	0.85	0.84	1.00	0.75	0.67	0.77
#18	0.91	0.26	0.62	0.16	0.92	0.98	0.46	0.28	0.28
#19	0.98	0.51	0.76	0.86	1.00	0.88	0.47	0.81	0.96
#20	0.79	0.33	0.94	0.87	0.78	0.92	0.95	0.67	0.75