申请上海交通大学博士学位论文

情感脑机交互研究

论文作者 郑伟龙

学　　号 0120339018

导　　师 吕宝粮　教授

专　　业 计算机科学与技术

答辩日期 2018 年 6 月

Submitted in total fulfillment of the requirements for the degree of Doctor
in Computer Science and Technology

Affective Brain-Computer Interactions

Wei-Long Zheng

Advisor
Professor Bao-Liang Lu

Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, P. R. China

June, 2018

上海交通大学
学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：_____

日　期：_____年_____月_____日

上海交通大学
学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

保　密 □，在 _____ 年解密后适用本授权书。

不保密 □。

(请在以上方框内打 ✓)

学位论文作者签名：_____　　指导教师签名：_____

日　期：_____年 ____月 ____日　　日　期：_____年 ____月 ____日

情感脑机交互研究

摘 要

情绪在日常生活人与人交流中扮演着重要角色。除了逻辑智能，情感智能也被认为是人类智能的重要组成部分。情感智能是指机器感知，理解和调控人的情绪的能力。然而，现有人机交互系统仍然缺乏情感智能。情感脑机交互研究的目的是通过构建情感计算模型来建立人与机器的情感交流通路。

在本论文中，我们探讨了情感脑机交互的理论基础，模型，算法，实现技术，实验验证以及原型应用。主要工作包括以下三个方面：

1）我们利用脑电，眼电和眼动信号以及深度神经网络构建了多模态情绪识别和警觉度估计系统。相对于传统浅层模型，深度神经网络能有效提高识别性能，并揭示情绪识别中关键频段和关键脑区，从而给出在实际应用中具有更少电极的配置方案。我们通过跨个体的在不同时间的多次实验，揭示了对于三类情绪（高兴、悲伤和中性）的稳定神经模式。我们发现高兴情绪在颞叶脑区具有更强的 beta 频段和 gamma 频段的脑电信号反应，中性和悲伤情绪的神经模式比较相似，中性情绪在顶叶和枕叶脑区具有更强的 alpha 频段脑电信号反应，而悲伤情绪在顶叶和枕叶脑区具有更强的 delta 频段脑电信号反应以及前额脑区更强的 gamma 频段脑电信号反应。

2）我们提出了利用脑电和眼动信号的多模态情绪识别框架，实现了对人的内在认知状态和外在潜意识活动的建模。我们探讨了特征层融合，决策层融合以及双模态深度自编码器等多模态信息融合方法。实验结果表明，多模态融合方法在三类和四类情绪识别中分别有 10% 和 15% 的准确率提升。另外，我们揭示了脑电和眼动信号在情绪识别中的互补特性。为了克服跨被试个体差异性问题和脑电信号的非平稳特性，我们引入了迁移学习算法构建基于脑电的个性化情感模型。其中，直推式参数迁移算法取得了 76.31% 的最高性能，比不用域适应的基线方法提高了将近 20% 的准确率。

3）我们开发了利用脑电和前额眼电构建多模态警觉度估计系统，并在实验室模拟环境和真实驾驶环境进行了验证。我们提出了新的前额眼电电极配置和利用眼动仪眼镜进行警觉度标注的新方法。为了对随时间变化的警觉度动态特性建模，我们引入了两种时序依赖性模型，连续条件随机场和连续条件神经场。我们首先在商业用湿电极脑电采集系统采集的脑电和眼电数据上进行了实验室模拟环境的实验评估。为了提高警觉度估计系统的可穿戴性和实用性，我们利用柔性干电极和超大规模集成电路技术开发了可穿戴

前额眼电采集设备。在实验室模拟环境和不同天气条件下的真实驾驶环境，实验结果验证了我们提出的警觉度估计系统的有效性。

在上述研究的过程中，我们开发了三个公开的情绪识别和警觉度估计数据集，分别是三类情绪数据集 SEED（高兴、悲伤和中性），四类情绪数据集 SEED-IV（高兴，悲伤，恐惧和中性）和警觉度估计数据集 SEED-VIG，这些数据集得到了国内外近 300 多高校和研究机构的使用。

关键词： **情感计算，情绪识别，情感脑机交互，警觉度估计，脑电信号，眼电信号，眼动信号**

Affective Brain-Computer Interactions

ABSTRACT

Emotion plays an important role in human and human communications in our daily life. Besides logical intelligence, emotional intelligence is considered as an important part of human intelligence, which represents the ability to perceive, understand, and response to emotion. However, the existing human-computer interaction systems still lack emotional intelligence. Affective brain-computer interactions aim to narrow the communication gap between human and machine by developing computational models of emotion.

In the thesis, we explore the theoretical basis, models, algorithms, implementation technologies, experimental validation, and prototype applications of affective brain-computer interactions. The main contributions are as follows,

1) Multimodal emotion recognition and vigilance estimation with EEG, EOG, and eye movements are developed using deep neural networks. Deep neural networks are adopted to improve the performance in comparison with conventional models. The critical frequency bands and brain regions are revealed and the optimal electrode placement with fewer electrodes and more feasibility is determined for emotion recognition. The neural patterns of three emotions (happy, sad, and neutral) are identified with commonality across individuals and sessions. We have found the following characteristics of EEG signals for emotion recognition: a) the beta band and gamma band responses at the temporal areas increased for happy emotion; b) the neural patterns of neutral emotion and sad emotion were similar; and c) neutral emotion had higher alpha band responses at parietal and occipital sites, while sad emotion had higher delta band responses at parietal and occipital sites and higher gamma band responses at prefrontal sites.

2) We propose multimodal emotion recognition framework using EEG and eye movements to model both internal cognitive states and external subconscious behaviors of human with various modality fusion strategies, including feature-level fusion, decision-level fusion, and bimodal deep auto-encoder. The experimental results demonstrate that significant improvements with about 10% accuracy for three emotion recognition and with about 15% accuracy for four emotion recognition were obtained in comparison with single modality. The complementary characteristics of EEG and eye movements are investigated. To deal with individual difference

across subjects and non-stationary characteristic of EEG signals, we introduce transfer learning for constructed personalized EEG-based affective models. The transductive parameter transfer algorithm achieves the best performance with 76.31% accuracy and a significant improvement of about 20% accuracy in comparison with the baseline without domain adaptation.

3) We develop mutimodal vigilance estimation systems with EEG and forehead EOG in both laboratory stimulations and real driving environments. The novel forehead EOG setup and vigilance annotation using eye tracking glasses are proposed. In order to capture vigilance dynamics with temporal evolutions, the temporal dependency models (CCRF and CCNF) are introduced. We first perform the evaluations of the computational models with the EEG and EOG data recorded by commercial EEG recording systems with wet electrodes. In order to improve the wearability and feasibility of vigilance estimation, we develop a wearable device with flexible dry electrodes and large-scale integrated circuits for the forehead EOG recordings. The experimental results on both laboratory driving simulations and real-world driving environments demonstrate the efficiency of the proposed vigilance estimation system.

We develop three publicly available datasets for emotion recognition and vigilance estimation, SJTU Emotion EEG Dataset (SEED) for happy, sad, and neutral emotions, SEED-IV for happy, sad, fear, and neutral emotions, and SEED-VIG for continuous vigilance estimation. These datasets have received more than 300 applications from universities and research institutes all over the world.

KEY WORDS: Affective Computing, Emotion Recognition, Affective Brain-Computer Interactions, Vigilance Estimation, EEG, EOG, Eye Movements

Contents

List of Figures **xv**

List of Tables **xviii**

Chapter 1 Introduction **1**

1.1 Motivation 1
1.2 Contributions 4
1.3 Thesis Overview 6

Chapter 2 Research Background **9**

2.1 Emotion Definition and Emotion Models 9
2.1.1 Discrete Model 10
2.1.2 Continuous Model 10
2.2 Brain Mechanism of Emotion 12
2.3 Electroencephalography 15
2.4 Emotion Elicitation and Emotion Experiment 17
2.5 Emotion Recognition 19
2.5.1 EEG-based Emotion Recognition 19
2.5.2 Multimodal Emotion Recognition 24
2.5.3 Public Emotion EEG Datasets 27
2.6 Driving Fatigue and Vigilance Estimation 28
2.7 Summary 30

Chapter 3 Experimental Setups **33**

3.1 SJTU Emotion EEG Dataset (SEED) for Three Emotions 33
3.1.1 Emotion Stimuli 33
3.1.2 Subjects 34
3.1.3 Experiment Protocol 35
3.2 SJTU Emotion EEG Dataset (SEED-IV) for Four Emotions 36
3.2.1 Emotion Stimuli 38

3.2.2 Subjects . . . 38
3.2.3 Experiment Protocol . . . 39
3.3 Data Processing for Multimodal Emotion Recognition . . . 40
3.3.1 Feature Extraction for EEG . . . 40
3.3.2 Feature Smoothing for EEG . . . 43
3.3.3 Dimensionality Reduction for EEG . . . 44
3.3.4 Feature Extraction for Eye Movements . . . 46
3.4 Multimodal Vigilance Estimation Dataset (SEED-VIG) . . . 47
3.4.1 Experimental setup . . . 47
3.4.2 Vigilance Annotation . . . 51
3.5 Wearable Device for Vigilance Estimation . . . 52
3.5.1 Flexible Dry Electrodes . . . 53
3.5.2 EOG Acquisition Board . . . 55
3.5.3 Laboratory Driving Simulations . . . 57
3.5.4 Real-World Driving Experiments . . . 58
3.6 Summary . . . 60

Chapter 4 EEG-based Emotion Recognition 63
4.1 EEG-based Emotion Classification Using Deep Neural Networks . . . 63
4.1.1 Introduction . . . 63
4.1.2 Deep Belief Networks . . . 64
4.1.3 Classifier Training . . . 66
4.1.4 Classification Performance . . . 67
4.1.5 Critical Frequency Bands and Channels . . . 70
4.1.6 Electrode Reduction . . . 73
4.2 Stable EEG Patterns over Time for Emotion Recognition . . . 76
4.2.1 Introduction . . . 76
4.2.2 Discriminative Graph Regularized Extreme Learning Machine . . . 77
4.2.3 Experiment Results on DEAP Data . . . 78
4.2.4 Experiment Results on SEED Data . . . 82
4.2.5 Neural Signatures and Stable Patterns . . . 84
4.2.6 Stability of the Emotion Recognition Model over Time . . . 88
4.3 Summary . . . 90

Chapter 5 Multimodal Emotion Recognition with EEG and Eye Movements 91
5.1 Introduction . . . 91
5.2 Multimodal Deep Learning . . . 93
5.3 Modality Fusion Methods . . . 94
5.4 Experimental Results on SEED for Three Emotions . . . 96
5.4.1 Eye Movement-Based Emotion Recognition . . . 96
5.4.2 Performance of Modality Fusion . . . 98
5.4.3 Analysis of Complementary Characteristics . . . 100
5.5 Experimental Results on SEED-IV for Four Emotions . . . 102
5.5.1 EEG-Based Emotion Recognition . . . 103
5.5.2 Analysis of Modality Fusion and Complementary Characteristics . . . 103
5.5.3 Analysis of Stability Across Sessions . . . 108
5.6 Summary . . . 108

Chapter 6 Personalizing Affective Models with Transfer Learning 111
6.1 Introduction . . . 111
6.2 Transfer Learning . . . 114
6.2.1 Transfer Component Analysis . . . 114
6.2.2 Kernel Principle Component Analysis . . . 115
6.2.3 Transductive Parameter Transfer . . . 116
6.3 Experiment Setup . . . 119
6.4 Experiment Results . . . 120
6.5 Heterogeneous Knowledge Transfer From Eye Tracking To EEG . . . 124
6.5.1 Introduction . . . 124
6.5.2 Spatiotemporal Scanpath Analysis . . . 127
6.5.3 Heterogeneous Transfer Learning . . . 128
6.5.4 Evaluation Details . . . 129
6.5.5 Experimental Results . . . 130
6.6 Summary . . . 133

Chapter 7 Multimodal Vigilance Estimation: From Simulated To Real Scenarios 135
7.1 Introduction . . . 135
7.2 Feature Extraction . . . 137
7.2.1 Preprocessing for Forehead EOG . . . 137

7.2.2 Feature Extraction for Forehead EOG . . . 139
7.2.3 Forehead EEG Signal Extraction . . . 141
7.2.4 Feature Extraction from EEG . . . 143
7.3 Incorporating Temporal Dependency into Vigilance Estimation . . . 144
7.4 Evaluation Metrics . . . 145
7.5 Experimental Results on SEED-VIG . . . 146
7.5.1 Forehead EOG-based Vigilance Estimation . . . 146
7.5.2 EEG-based Vigilance Estimation . . . 147
7.5.3 Modality Fusion with Temporal Dependency . . . 148
7.5.4 Complementary Characteristics . . . 152
7.6 Experimental Setups with Wearable Device . . . 153
7.7 Experimental Results on Wearable Device . . . 154
7.7.1 Laboratory Driving Simulations . . . 154
7.7.2 Real-World Driving Experiments . . . 158
7.8 Discussion . . . 159
7.9 Summary . . . 163

Chapter 8 Conclusions and Future Work **165**
8.1 Summary of Contributions . . . 165
8.2 Future Work . . . 167

Acknowledgements **193**

Publications **195**

Project Paticipation **201**

List of Patents **203**

Resume **205**

List of Figures

1–1 Facial expression of various emotions. 2
1–2 The organization of the thesis. 6

2–1 The wheel of emotions proposed by Plutchik[47], retrieved from WikiPedia [1]. . . 11
2–2 An illustration of the arousal-valence emotion model. 12
2–3 Key brain regions that are associated with emotion processing[68]. 14
2–4 EEG recordings, retrieved from Saint Luke's Health System [2]. 15
2–5 Illustrations of EEG waveforms, including Delta waves, Theta waves, Alpha waves, Beta waves, and spikes [3]. 16
2–6 Self-Assessment Manikins (SAM) with the two affective dimensions: valence and arousal. 19
2–7 Closed-loop affective brain-computer interactions. 20
2–8 Various modalities that are commonly used in health informatics[133] 25

3–1 The selected Chinese films in the emotion experiments. 35
3–2 The protocol used in the emotion experiment. 36
3–3 The experimental scene in the emotion experiments. 37
3–4 The layout of 62 EEG electrodes on the cap. 37
3–5 The SMI ETG eye tracking glasses used in the study and the pupillary image captured in one experiment. 38
3–6 The mean rating distributions of the different film clips on the arousal-valence plane for four emotions. The ratings are clustered into four classes: happy, sad, fear, and neutral emotions. 39
3–7 The protocol of SEED-IV for four emotions. 40
3–8 The EEG electrode layout of 62 channels. Six symmetrical temporal electrodes (*FT7*, *FT8*, *T7*, *T8*, *TP7*, and *TP8*) are selected in SEED-IV. 40
3–9 The hardware setup. The six symmetrical temporal electrodes above the ears are used for EEG recordings. The eye movement parameters are extracted from the wearable eye-tracking glasses. 41
3–10 The electrode layout of the left and right hemispheres. 43

3–11 The comparison of EEG feature before and after smoothing by the LDS approach. The EEG signal is obtained from the channel FP1 in gamma band of one subject in the SEED experiments. 45

3–12 Removing the luminance influences of pupil responses using PCA: (a) the raw pupil diameters to the same video clip from four subjects, (b) the first principle component as the estimate of the light reflex, and (c) pupil response associated with emotions only after subtracting the first principle component. 47

3–13 Illustration of various eye movement parameters: pupil diameter, fixation dispersion, saccade amplitude, saccade duration, and blink. 49

3–14 The simulated driving system and the experimental scene for multimodal vigilance estimation. (a) The simulated driving scenes, (b) One participant with EEG, forehead EOG, and eye tracking setups in the experiments, and (c) The simulated environment with a large LCD screen and a vehicle without without unnecessary engine and other components. 50

3–15 The electrode placements for the traditional and forehead EOG setups in vigilance estimation experiments. Electrode four is the shared electrode of both setups. 50

3–16 The electrode placements for the EEG setups in vigilance estimation experiments. Here, 12-channel and 6-channel EEG signals were recorded from the posterior site (red colour) and temporal site (green colour), respectively. 51

3–17 The designed dry fabric-based electrode. 53

3–18 The contact impedance of different frequencies for the designed conductive fabric electrodes. 54

3–19 The raw signals recorded simultaneously by using the commercial wet electrode (upper) and the dry fabric electrode (lower) on the forehead site $FP1$. 55

3–20 EEG signals recorded by using the dry conductive fabric electrodes: (a) EEG signals recorded with eyes closed and (b) PSD of the recorded EEG signals. . . 56

3–21 The designed EOG acquisition board and its paired tablet user interface. 56

3–22 System diagram and signal flow of the designed EOG acquisition board. 57

3–23 The map of the real-world driving experiments. 59

3–24 The sample frames recorded from the scene camera of eye-tracking glasses as the first perspective. The real-world driving experiments were performed under various illumination and weather conditions, including sunny, cloudy, windy, rainy and night time. 59
3–25 The snapshot of the project website SJTU Emotion EEG Dataset (SEED). . . . 60

4–1 (a) A RBM contains the hidden layer neurons connected to the visible layer neurons with weights W. (b) A DBN using supervised fine-tuning of all layers with backpropagation. (c) The graphical depiction of unrolled DBN using unsupervised fine-tuning of all layers with backpropagation. 65
4–2 The average accuracies of different classifiers using the DE features. 68
4–3 The confusion matrix of different classifiers on one experiment for one subject. Here the number inside the figures denotes the recognition accuracy in percentage. 69
4–4 The DE feature map in one experiment, where the time frames are on the horizontal axis, and the DE features are on the vertical axis. 71
4–5 The mean absolute weight distribution of the trained DBNs in the first layers with the DE features from the total frequency band. 72
4–6 The weight distribution of different brain regions in the five frequency bands . . 72
4–7 Four different profiles of selected electrode placements according to the high peaks in the weight distribution and electrode symmetry: (a) 4 channels; (b) 6 channels; (c) 9 channels; (d) 12 channels. 74
4–8 The rating distribution of DEAP on the arousal-valence plane (VA plane) for the four conditions (LALV, HALV, LAHV, and HAHV) 79
4–9 The average accuracies of GELM using different features obtained from five frequency bands and using a fusion method on SEED. 84
4–10 The spectrogram of the electrode position T7 in one experiment. As different emotions elicited, we can see that the spectrogram has different patterns. 85
4–11 The average spectrogram of the participants for each session at some electrodes, which shows the stable neural patterns over time in the temporal lobes and high-frequency bands (a red color indicates a high amplitude). 86
4–12 The average neural patterns for all participants and sessions for different emotions, which shows that neural signatures associated with happy, neutral and sad emotions do exist. 86

4–13 The results of dimensionality reduction using PCA and MRMR 87
4–14 Distribution of the top 20 participant-independent features selected using the correlation coefficients. One electrode location "FT8" that is not shown is from the alpha frequency band. 87

5–1 The multimodal framework with EEG and eye movements 92
5–2 Multimodal deep neural network architecture for combining EEG and eye movements. 94
5–3 Individual and average pupil diameters of each of nine participants for three emotion states. 97
5–4 Box plots of four eye movement parameters (dispersion, fixation duration, saccade duration, and saccade amplitude) for three emotional states. 98
5–5 Performance of single modality and different modality fusion strategies on SEED. 99
5–6 Confusion graph of eye movements and EEG, which shows their complementary characteristics for emotion recognition. (The numbers is the percentage of samples in class (arrow tail) that is classified as class (arrow head). Bolder lines mean higher values.) 101
5–7 The fusion framework of EEG and eye movements using deep neural networks on SEED-IV. 102
5–8 Visualization of the DE features and the corresponding labels in one experiment. Here, labels with 0, 1, 2, and 3 denote the ground truth, neutral, sad, fear and happy emotions, respectively. The numbers 1, 2, 3, 4, and 5 in the vertical coordinates respectively denote the five frequency bands: δ, θ, α, β, and γ. The dynamic neural patterns in high frequency bands (α, β, and γ) have consistent changes with the emotional labels during the whole experiment. 104
5–9 The box plot of the accuracies with each single modality (eye movements and EEG) and the two modality fusion approaches on SEED-IV. The red lines indicate the median accuracies. 106
5–10 The confusion matrices of single modality and multimodal fusion methods on SEED-IV. 107
5–11 Confusion graph of EEG and eye movements on SEED-IV, which shows their complementary characteristics for emotion recognition. (The numbers denote the percentage values of samples in the class (arrow tail) classified as the class (arrow head). Bolder lines indicate higher values.) 107

6–1 Illustration of the covariate-shift challenges of constructing EEG-based affective models. 112

6–2 The subject transfer framework of EEG-based emotion recognition. 113

6–3 The framework of the transductive parameter transfer (TPT) approach for subject transfer in EEG-based emotion recognition. 118

6–4 Comparison of KPCA and TCA approaches with varying dimensionality of the latent subspace. 121

6–5 Visualizations of the EEG features of one experiment in a 3-dimension latent space using TCA and KPCA, respectively. 121

6–6 The accuracies of the five different transfer learning methods (Generic classifiers, KPCA, TCA, T-SVM, and TPT) for each subject. 122

6–7 The confusion matrices of the five transfer learning methods. 124

6–8 The framework of the proposed multimodal transductive parameter transfer for leveraging heterogeneous knowledge from spatiotemporal scanpath patterns to enhance subject transfer in EEG-based affective models. 126

6–9 Illustration of scanpath sequence encoding. Here, the original sequence is *ABCDEFGH*. Taking fixation duration into account, the temporal binning is introduced and the original sequence is transformed into the following sequence: *ABBCDEEFFGH*. 127

6–10 Classification accuracies of KMM, Generic, GenSub, GFK, KPCA, TCA, ARTL, DE, EMD, Scanpath, Fuzzy, Sum, and Max. The first nine methods construct subject transfer models using EEG data solely. '*Scanpath*' denotes TPT approach with scanpath-based kernels. The last three methods combine subject transfer models using EEG based and scanpath based kernels. 131

6–11 The DE-based, EMD-based, and scanpath-based kernels. 132

7–1 The flowchart and structure of our proposed wearable vigilance estimation system. 137

7–2 The electrode placements for the traditional and forehead EOG setups in vigilance estimation experiments. Electrode four is the shared electrode of both setups. 138

7–3 Comparison of traditional EOG and forehead EOG using minus operation and ICA separation strategies. Here, (a) and (d) are traditional VEO and HEO; (b) and (e) are extracted VEO_f and HEO_f from forehead EOG using the minus operation; and (c) and (f) are extracted VEO_f and HEO_f from forehead EOG using the ICA approach. 139

7–4 The blink detected by using continuous wavelet transform. 140

7–5 The saccade detected by using continuous wavelet transform. 141

7–6 (a) The decomposed independent components from the four forehead channels (Nos. 4-7) using ICA. IC 1 and IC 2 are EOG components for eye activities. (b) The reconstructed forehead EEG by filtering out the EOG components. It can be observed that strong alpha activities are verified under eye closure conditions. 143

7–7 The mean neural patterns of awake and drowsy states as well as the difference between these two states. From the average neural patterns, we observe that drowsy states have higher theta and alpha frequency activities and lower gamma frequency activities in temporal and parietal areas ($p < 0.01$, ANOVA). 149

7–8 The mean COR and mean RMSE of each single modality and different modality fusion strategies. 150

7–9 The continuous vigilance estimation of different methods in one experiment. As shown, the predictions from our proposed approaches are almost consistent with the true subjects' behaviours and cognitive states. 151

7–10 Confusion graph of forehead EOG and posterior EEG, which shows their complementary characteristics for vigilance estimation. Here, the numbers denote the percentage values of samples in the class (arrow tail) classified as the class (arrow head). Bolder lines indicate higher values. 153

7–11 The performances of CCRF and CCNF with varying sequence length in the simulated driving experiments. 155

7–12 The detailed performances of different models: SVR, CCRF, and CCNF for individual subjects in the simulated driving experiments. 156

7–13 The continuous vigilance estimation of different methods in one experiment (Subject 6) in the simulated driving experiments. As shown, the predictions from our proposed system are almost consistent with the true subjects' vigilance states. 157

7–14 The performances of CCRF and CCNF with varying sequence length in the real-world driving experiments. 159
7–15 The detailed performances of different models, SVR, CCRF, and CCNF, for individual subjects in the real-world driving experiments. 160
7–16 The continuous vigilance estimation of different methods in one experiment (Subject 9) in the real-world driving experiments. 161

List of Tables

2–1 The summarization of the affective video databases[51]. 18
2–2 Various studies on emotion classification using EEG (continued on next page). . 21
2–3 Various studies on emotion classification using EEG (continued from previous page). 22

3–1 Details of the film clips used in our emotion experiment 34
3–2 The details of the extracted eye movement features. 48
3–3 Summary of Main Components and Technical Parameters of the Proposed Forehead EOG Acquisition System. 57

4–1 The details of parameters used in different classifiers 67
4–2 The mean accuracies and standard deviations (%) of SVM and DNN for different kinds of features 68
4–3 The mean accuracies and standard deviations (%) of SVM for different profiles of electrodes sets 75
4–4 The ratings of arousal and valence for the quadrants of VA space. 80
4–5 The mean accuracy rates (%) of SVM and GELM classifiers for different features obtained from separate and total frequency bands in DEAP. 80
4–6 Comparison of the various studies using EEG in the DEAP dataset. 81
4–7 The means and standard deviations of accuracies (%) for the PSD, DE, DASM, RASM, ASM and DCAU features from the total frequency bands on SEED. . . 82
4–8 The Performance (%) of Different Feature Smoothing Methods on SEED. . . . 83
4–9 The Mean Accuracies (%) of the Five Classifiers (KNN, LR, SVM and GELM) on SEED. 83
4–10 The average accuracies (%) of our emotion model across sessions. 88
4–11 The classification accuracies (%) of the training and test data from different sessions using GELM. 89

5–1 Classification accuracies (%) of different eye movement features. ('FLF' means feature level fusion by combining all the eye movement features) 98

5–2 Classification accuracies (%) of different features and their feature level fusion from EEG. 99
5–3 Confusion matrices of single modality and different modality fusion strategies. 100
5–4 The mean accuracy rates (%) of different electrode setups for the two different features obtained from the separate and total frequency bands. 105

6–1 The mean accuracies and standard deviations of the five different transfer learning approaches. 122

7–1 The details of the extracted 36 EOG features. 142
7–2 The mean RMSE, the mean COR, and their standard deviations with different separation methods. Here, the numbers in the first and second rows are the averages and standard deviations, respectively. 147
7–3 The average and standard deviations of COR and RMSE for different EEG features. Here, the numbers in the first and second rows are the averages and standard deviations, respectively. 148

Chapter 1

Introduction

1.1 Motivation

Artificial Intelligence (AI) research has made great progress in various domains from computer vision, natural language processing, speech recognition, to autonomous cars, robots, and precise medicine in recent years and it has been drawing more and more attention and interest from both industry and scientific communities since last AI winter. One of the major breakthrough techniques is deep learning. Inspired by hierarchical structures of mammalian visual cortex, deep learning trains various deep architectures to extract intermediate representations by using layer-wise unsupervised learning[1, 2]. Recently, AlphaGo[3] developed by Google Deepmind became the first program to defeat a world champion (Lee Sedol) in the game of Go with deep neural networks. More recently, AlphaGo Zero, a latest version of AlphaGo[4], can quickly surpass the human levels even without human knowledge, which is considered a major breakthrough for AI. Another example is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in computer vision[5]. Deep learning algorithms surpassed the human-level performance at image recognition in ImageNet[6].

Although great progress has been achieved in artificial intelligence in the past decades, especially in numerical computing, optimization and statistical methods, there are still some limitations and challengers about reasoning, planning, learning and creativity in comparison with human intelligence. Among these high level cognition, emotion plays a crucial role in natural communications in our daily life. Emotion has been suggested to interact and influence with other domains of cognition, e.g., attention, memory, reasoning, and decision making[7]. Therefore, besides logical intelligence, **emotional intelligence** is considered as an important part of human intelligence. Emotion intelligence represents the ability to perceive, understand and response to emotions, which was firstly proposed by Salovery and Mayer[8]. However, at present, both academia and industry mainly focus on logical intelligence, and the research of emotional intelligence is still in its infancy. For general artificial intelligence, logical intelligence and emotional intelligence should not be developed in isolation and should considered in a unified framework. Current Human-Computer Interactions (HCI) paradigms are mainly based

on specific rules or pre-designed intrusions. By introducing affective factors, HCI systems can adaptively respond to users' emotional states and enhance the natural interactions between human and machine.

What is an emotion? This is a very important question when you would like to study emotion. However, there is still no common definition in research community. James Gross, a famous psychologist from Stanford University, concluded that there are only two points of agreement for emotion[9]. One is that emotion is a collection of psychological states that include subjective experience, expressive behaviors, and peripheral physiological responses. And the other is that emotions are a central feature in any psychological model of human mind. Emotion research is an interdisciplinary field across neuroscience, cognitive science, psychology, and computer science. The study about computational models of emotion has been rapidly developed as an interdisciplinary research field called **Affective Computing** (AC)[10–12] proposed by Rosalind W Picard from MIT Media Lab, which aims to narrow the communication gap between human and machine.

Figure 1–1 Facial expression of various emotions.

Emotion recognition is the primary and important phase in the framework of emotion intelligence. Since emotion is a complex psychological states including various expressive behaviors and peripheral physiological responses, researchers utilized various modalities, such as facial expressions, voice, heart rates, and skin conductance etc., as indicators of emotions[13–15]. One of the twenty big questions about the future of humanity reported by *Scientific American* is whether we can use wearable technologies to detect human emotions [1]. Among these modalities, electroencephalography (EEG) is considered as an efficient and reliable neural signal for representing different emotions, since EEG can objectively record our brain activities and human brain is the central nervous system for emotion precessing.

[1] `http://www.scientificamerican.com/article/20-big-questions-about-thefuture-of-humanity/`

Brain-Computer Interfaces (BCIs) or **Brain-Machine Interfaces** (BMIs) aim to construct a direct communication pathway between brain and external devices through recording specific neural activity, constructing cognitive models, decoding users' intentions and mental states and developing closed-loop neuro-feedbacks[16–19]. Based on the measurement of brain activity, BCIs can be categorized into invasive[20, 21] (e.g., electrocorticogram (ECoG)) and non-invasive[22] (e.g., functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and functional near-infrared spectrography (fNIRS)) technologies. EEG-based brain-computer interfaces are the popular study in the non-invasive category because of the high temporal resolution of EEG[23]. Most current BCI systems focus on neuroprosthetics applications based on motor imagery and visual evoked potential[24]. Zander and Kothe extended BCIs to new definitions and divided BCI technology into active, reactive, and passive BCIs[25]. They explored the potential BCI applications that not only enable disabled users to send direct commands for communication and control, but also recognize the mental states of users and adaptively respond to situations for healthy users. In this framework, the detection and prediction of convert aspects of user states e.g., emotion and attention, are included in passive BCI systems and also called cognitive monitoring.

Affective Brain-Computer Interactions (aBCIs) refer to hybrid BCI systems with the ability to detect, process, and respond to users' affective and cognitive states using various modalities[26, 27]. Combining conventional BCI systems with cognitive monitoring provides context information about users' mental states while it does not increase the cost significantly. Human factors are important and necessary for robust systems that need to interact with human, especially for human-in-the-loop systems. Emotion and vigilance are two vital components among cognitive states, which are the focus of this research. While great progress has been achieved in aBCIs, there still exist many critical challenges and open questions in Affective BCIs. Brain activity has been suggested with emotion. Although the underlying neural mechanisms of emotion processing has been extensively studied in neuroscience and psychology, the computational modeling of the emotion process is yet well explored. How can we deal with artifacts and noises in uncontrolled environments? How can machines response to the recognized affective states and bring users to a desired affective state? How can we elicit and measure emotions in social setting. How can we develop adaptive Affective BCI systems that address individual differences and changing environments? How can we introduce the contextual information to Affective BCIs? What are the neural patterns or signatures for different affective states and how is the stability of computational models over time? In this study, we will explore some parts of these

questions.

Affective brain-computer interactions have broad potential applications across various domains. For example, mental diseases are reported to be relevant to emotions, such as depression, autism, game addiction, and attention deficit hyperactivity disorder[28–30]. An efficient quantified measurement for emotion can provide active feedback with objective evaluations for disease treatments, instead of subjective feedbacks with questionnaires.

According to the report of the National Highway Traffic Safety Administration (NHTSA), an agency of the US Department of Transportation, vigilance decrement or attention lapse has long been recognized as the critical factor responsible for thousands of deaths and injuries each year in the public traffic community[31]. Prolonged driving on monotonous roads often leads to drivers fatigue and control performance reduction. For some dangerous occupations, vigilance decrement[32] or a momentary lapse of attention[33, 34] might severely endanger pubic transportation safety. Considerable progress has been achieved in various areas over the past decades. However, there is still a lack of efficient quantitative measurements for vigilance or attention, especially in real-world driving environments. The main reason for this difficulty is that vigilance states are intrinsic mental states that involve temporal evolution rather than a time point. Moreover, real-world applications require continuous vigilance estimation with high temporal resolution[34]. In comparison with approaches based on external behaviors, such as head nodding, yawning, and eye closure, EEG modality is one of the most popular signals used for vigilance estimation because of its advantages of high temporal resolution and noninvasive low-cost properties. And EEG can even predict attention lapses in a convert setting up to 10 s in advance before external behaviors[35].

Emotional communications are important components in human and human communications. Emotional human-computer interaction is an important but still unsolved problem in the field of artificial intelligence. Integrating emotional intelligence and logical intelligence as the general framework of artificial intelligence has important scientific value and wide social applications.

1.2 Contributions

In this doctoral thesis, we mainly explore the theoretical basis, implementation methods, experimental validation, and prototype applications of Affective Brain-Computer Interactions. The contributions and achievements of the current thesis can be summarized in the following items:

- We investigated the stable neural patterns over time for emotion recognition from EEG.

The experimental results revealed that neural signatures for three emotions (happy, sad and neutral) do exist and that EEG patterns at critical frequency bands and brain regions are relatively stable within and between sessions.

- We applied deep belief networks to emotion recognition using multichannel EEG signals and enhanced the classification performance in comparison with conventional shallow models. We further investigated the critical frequency bands and brain areas associated with emotion recognition by analyzing the weight distributions learned from the trained deep neural networks. We investigated different electrode set reductions and defined the optimal electrode placement which outperforms original full channels with less computational cost and more feasibility in real world applications.
- We proposed multimodal emotion recognition framework using EEG and eye movements to model both internal cognitive states and external subconscious behaviors with various modality fusion strategies. The complementary characteristics of EEG and eye movements were discussed.
- To deal with individual difference across subjects and non-stationary characteristic of EEG, we adopted transfer learning algorithms in a transductive setup for constructing personalized EEG-based affective models.
- We proposed a multimodal approach to estimating vigilance by combining EEG and forehead EOG and incorporating the temporal dependency of vigilance into model training. We designed a novel forehead electrode setup for the simultaneous recording EEG and EOG. We implemented a wearable prototype device that integrates flexible dry electrodes and an acquisition board and performed systematic experiments not only in laboratory simulations but also in real-world scenarios under different weather conditions.
- We developed three publicly available EEG datasets for emotion recognition, namely, SJTU Emotion EEG Dataset (SEED) for three emotions and four emotions (SEED-IV) [1] and vigilance estimation, namely, SEED-VIG. To the best of our knowledge, this is the first public EEG dataset for the analysis of the stability of neural patterns regarding emotion recognition. The datasets have drawn more and more attention and have received over 300 applications from all over the world.

The mentioned contributions were published in 13 articles as first author in interactional journals and conference proceedings. See also the Publication section for the full listing. The study about stable neural patterns over time for emotion recognition from EEG was reported with

[1] http://bcmi.sjtu.edu.cn/%7Eseed/

title “How One Intelligent Machine Learned to Recognize Human Emotions” in MIT Technology Review in 2016 [1] and the paper[36] published in IEEE Transactions on Autonomous Mental Development received the IEEE TAMD Outstanding Paper Award from IEEE Computational Intelligence Society.

1.3 Thesis Overview

Figure 1–2 depicts the organization of the thesis. The remainder of the thesis is organized as follows. In Chapter 2, we provide an overview of the literature in the fields of emotion models, neural mechanisms about emotion, and affective brain-computer interactions. Chapter 3 describes the detailed experimental setups of emotion recognition and vigilance estimation in laboratory simulations and real-world scenarios. The details of preprocessing, feature extraction, feature smoothing for EEG and eye movements are also introduced. Chapter 4 presents

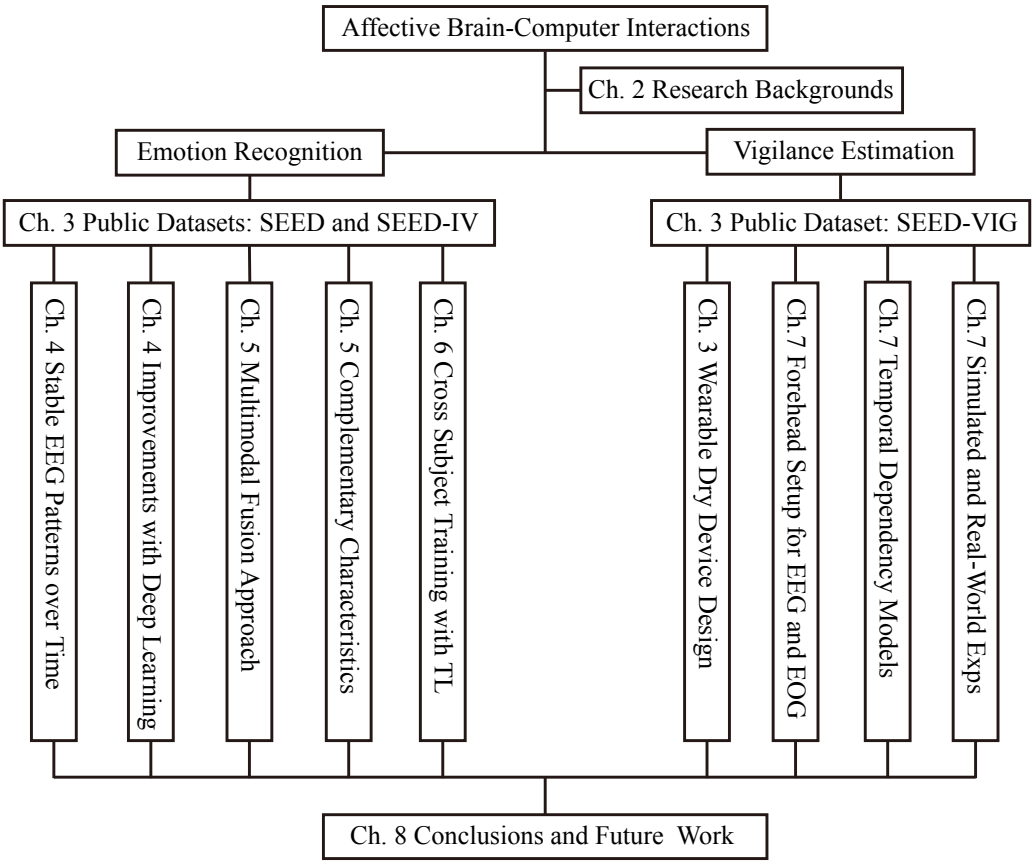


Figure 1–2 The organization of the thesis.

[1] https://www.technologyreview.com/s/545986/how-one-intelligent-machine-learned-to-recognize-human-emotions/

the experimental results and discussions about EEG-based emotion recognition. In Chapter 5, we present the experimental results and discussions about multimodal emotion recognition using EEG and eye movements. The performance evaluation of transfer learning for tackling subject-to-subject variations and the heterogeneous knowledge transfer between eye tracking and EEG are systematically described in Chapter 6. Chapter 7 presents the experimental results and discussions about multimodal vigilance estimation using EEG and forehead EOG in laboratory simulations and real-world scenarios under different weather conditions. In Chapter 8, the conclusions and future work are summarized.

Chapter 2

Research Background

In this Chapter, we describe the background of affective brain-computer interaction research. Section 2.1 introduces emotion definition and emotion models in psychology. Section 2.2 gives a brief review of relevant brain mechanism and neural correlates of emotion, which provides fundamental theoretical support for aBCIs. A brief introduction of electroencephalography is given in Section 2.3. In order to record relevant emotion data in the laboratory, emotion elicitation and emotion experiment are necessary (Section 3–3). An overview of the research about EEG-based emotion recognition and multimodal emotion recognition as well as public emotion datasets is given in Section 2.5. Finally, Section 2.6 introduces the motivations of vigilance estimation for driving fatigue and recent progress in this field. Section 3.6 concludes the Chapter.

2.1 Emotion Definition and Emotion Models

What is an emotion? More than a century ago, the psychologist William James also asked the question and he thought that the body language and physical instinct are in concordance with emotional experience[37]. Pleasure, sadness, fear, anxiety, anger, disappointment, disgust, excitement, shyness, pleasure, stress, etc., these are all part of our usual vocabulary of emotions. Although our emotional vocabulary is very rich, it is hard to translate into specific states and variables that can be directly quantified and measured in the research. In fact, one of the most important difficulties in the study of emotions lies in how to define and obtain real emotional labels[38]. Emotions are influenced by time, context, space, culture and individual differences[39]. Different researchers set up different emotional models from different research backgrounds and perspectives. Researchers studying animal models usually use reward and punishment, whereas human studies use “basic” emotions or complex social emotions in psychiatric[40]. Scherer proposed that “emotion is defined as an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism”[38]. Kleinginna and Kleinginna proposed the definition of emotion as a complex set of interactions among subjective and objective factors[41]. The details of the history and relevant concepts are reviewed in

the literature[9, 38, 42–44].

At present, researchers from human-computer interactions and affective computing mainly focus on two types of emotional models: discrete model and continuous model. However, the definition and models of emotion are still controversial. Defining emotions in the basic and continuous models still does not describe all emotions very well. However, they provide an initial framework for us to study emotion scientifically so that we can conclude some effective conclusions about emotions.

2.1.1 Discrete Model

Discrete model theory supports that emotions are composed of a variety of discrete basic and universal emotions and the rest of complex emotions are the combination of these basic emotions. Different psychologists proposed different sets of basic emotions and the numbers of these basic emotions vary from two to fourteen.

One of the most famous discrete model of emotion is proposed by Ekman. By studying facial expressions of different cultures, Ekman found some basic emotions are universal and he proposed six basic emotion, namely, fear, anger, surprise, disgust, joy, and sadness[45]. Here, the term basic includes three aspects: unique characteristics regrading physiological and behavioral responses, a learned response to deal with various tasks, and elemental[46]. The limitation of this model is that the number of emotion categories is too small in comparison with the diversity of emotion. For example, Ekman's basic emotion model contains only one positive emotion, which might cause difficulty in positive evaluations. Plutchik developed another classic basic model, the "wheel of emotions", which consists of eight primary emotions: joy, sadness, anger, fear, trust, disgust, surprise, and anticipation[47] (Figure 2–1). Mohammad and Turney from Canada's National Research Council released the Word-Emotion Association Lexicon with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) using crowdsourcing[48].

2.1.2 Continuous Model

The continuous model (also called dimension model) defines emotions in a coordinate system composed of multiple dimensions. The most famous one is Russell Circumplex's model proposed in 1980 by Russell[49]. It describes all human emotions as a two-dimensional coordinate system established by Valence and Arousal (see Figure 2–2). Another three-dimensional model

[1] https://en.wikipedia.org/wiki/File:Plutchik-wheel.svg

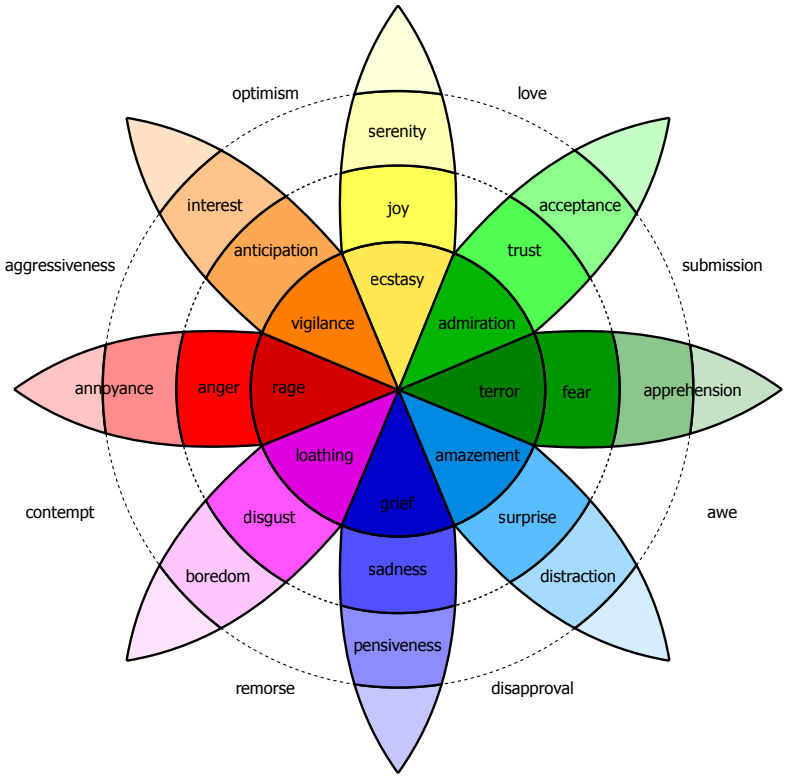


Figure 2–1 The wheel of emotions proposed by Plutchik[47], retrieved from WikiPedia [1].

is the valence-arousal-dominance space (i.e., the pleasure-arousal-dominance space) proposed by Russell and Mehrabian[50]. Valence ranges from negative to positive, whereas arousal ranges from calm to excited, and dominance ranges from dominated to in control.

All emotional states can be represented by a point coordinate in the low-dimensional coordinate system. The model narrows the original almost infinite research space to the two-dimensional space. Using a dimensional approach, researchers can more specifically assess and quantify stimulus-triggered emotional responses that have been adopted by many researchers[26, 51, 52]. However, Fontaine and colleagues proposed a different view that the world of emotions is not two-dimensional and four dimensions are needed to represent the similarities and differences of emotions[53]. These dimensions are evaluation-pleasantness, potency-control, activation-arousal, and unpredictability. Recently, Cowen and Keltner[54] analyzed the self-reports of emotional states elicited by more than two thousand video clips and they found 27 distinct categories of emotion. They reported that categorial labels (discrete labels) can better

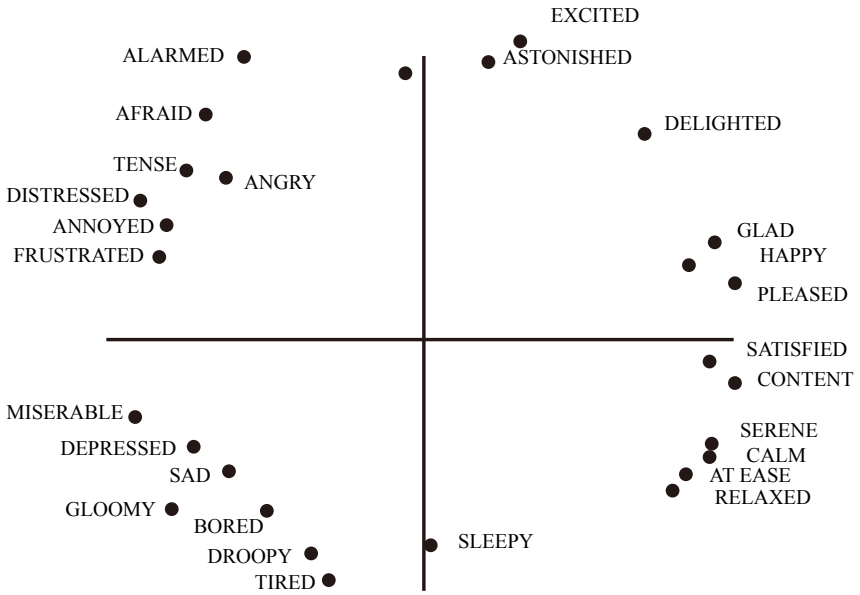


Figure 2–2 An illustration of the arousal-valence emotion model.

describe the subjective experience than dimensional labels (e.g., valence and arousal). The distribution of these emotions has fuzzy boundaries with continuous gradients, which correspond to affective dimensions in continuous models.

2.2 Brain Mechanism of Emotion

One of the goals of affective neuroscience is to examine whether patterns of brain activity for specific emotions exist and whether these patterns are to some extent common across individuals. Various studies have examined the neural correlations of emotions. It seems that processing modules for specific emotion do not exist. However, neural signatures of specific emotions, as a distributed pattern of brain activity[55], may exist. Davidson proposed there are differences between the left and the right hemispheres for positive and negative emotions[56, 57]. Mauss and Robinson[58] proposed that the emotional state is likely to involve circuits rather than any brain region considered in isolation. To affective computing researchers, identifying neural patterns that are both common across participants and stable across sessions can provide valuable information for emotion recognition based on brain imaging data.

Dalgleish presented a detailed review about the historical development of affective neuroscience[59]. Charles Darwin[60] and James William[37] are regarded as two fathers of affective neuroscience. Charles Darwin made two important contributions: animals emotions are ho-

mologue for human emotions and basic emotions exist across species and cultures, which have a profound influence on analyzing animals emotions to understand human emotions. Later, James-Lange theory of emotions[61] proposed the bodily theory of emotion, which was challenged by the Cannon-Bard theory[62]. James Papez[63] proposed the central neural circuitry of emotion (i.e., Papez circuit). He proposed a series of neural connections across sensory cortex, hippocampus, hypothalamus, anterior thalamus and cingulate cortex with two streams: stream of thinking and stream of feelings. Another model of emotional brain was proposed by Paul Maclean (Maclean's limbic system[64]), which has been dominant in affective neuroscience. For the details of these relevant theories, we recommend readers refer to the literature[59, 65–67]. From these theories, we can see that there are many structures involved in the emotional brain and these structures are highly distributed and interacted in different brain regions. Figure 2–3 shows the key brain regions that are associated with emotion processing[68]. For example, amygdala has been found to be an important component that is associated with fear emotion in neuroscience[69]. Janak and Tye presented a review article about amygdala studies from circuits to behaviors[70]. Tye and colleagues investigated the amygdala circuits of anxiety and they found that distinct amygdala projections can control opposite emotional valence[71]. Therefore, targeting specific neural circuits with projections and interactions, rather than simply cell types, is very important for understanding the neural mechanisms of emotion.

With the development and emergence of new technologies and methods, nowadays we have more efficient approaches to monitoring brain activity in both invasive and non-invasive ways. Recently, researchers in neuroscience use advanced brain imaging techniques to find the neural circuits and processing mechanisms that underlie the emotion experience. Functional Magnetic Resonance Imaging (fMRI) provides new evidence into emotional phenomena. fMRI can provide the functional neuroanatomy of emotion and point out which parts of brain are crucial in emotion processing[72].

Nummenmaa and colleagues studied the synchrony of brain activity across subjects in social interactions using functional MRI[73]. And they found that high arousal synchrony and negative valence synchrony support similar attention to the environment and emotional sensations across subjects, respectively. Kragel *et al.* utilized functional MRI to classify seven emotional states: content, amusement, surprise, fear, anger, sad, and neutral[74]. Vuilleumier and Pourtois investigated the brain mechanisms during emotion face perception using functional MRI and EEG signals[75]. They demonstrated that emotion face perception is related to an interactive network with distributed activity in time and space. Recently, Tanaka and colleagues[30] reported an

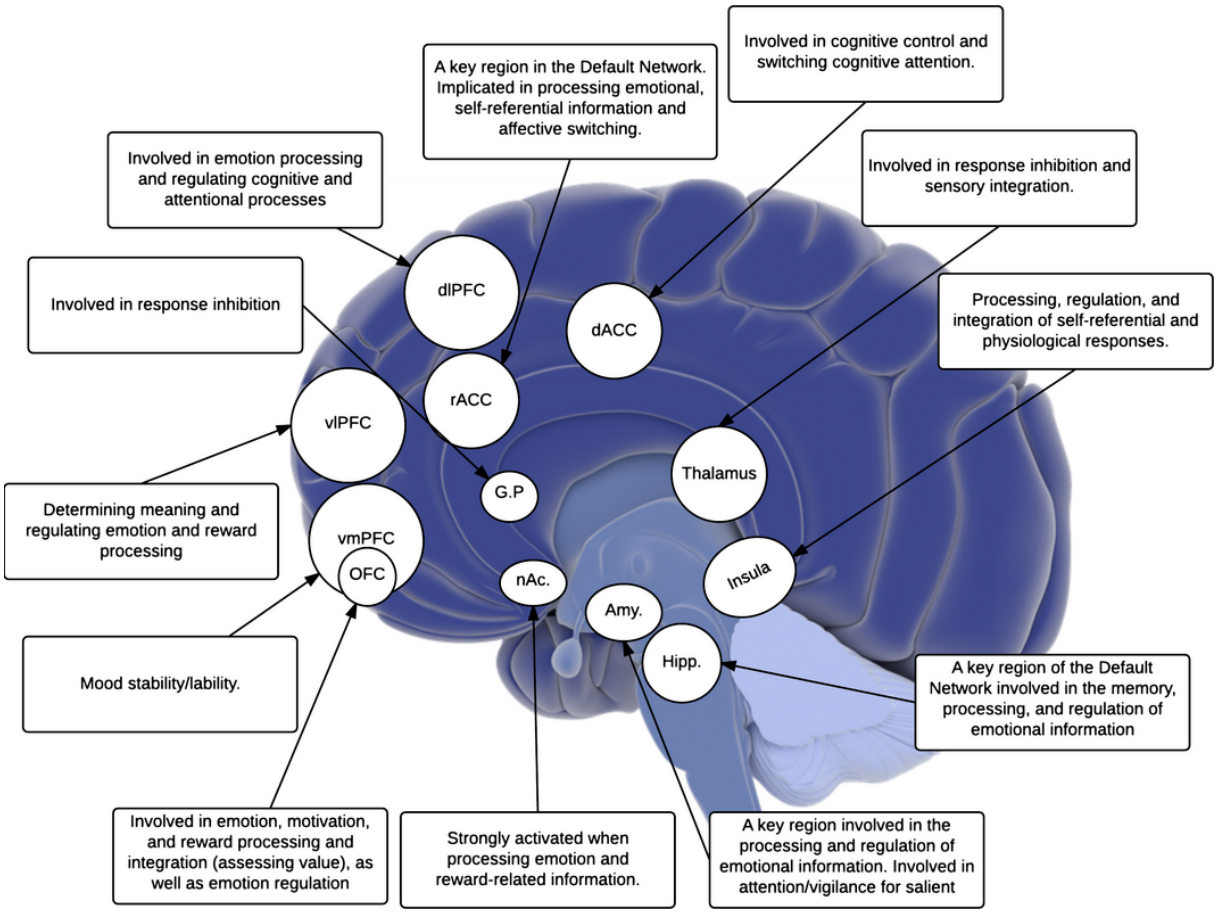


Figure 2–3 Key brain regions that are associated with emotion processing[68].

interesting result about neural mechanisms of relation between economic inequality and depression. They demonstrated that functional MRI patterns in the amygdala and hippocampus can predict present and future depression indices and economic inequality has importance influence on human emotional states. Saarimäki *et al.*[76] revealed the neural signatures of six basic emotions: disgust, fear, happiness, sadness, anger, and surprise. They found that the emotion specific neural signatures are distributed in the networks of cortical and subcortical areas. There is no direct one-to-one mapping between specific emotions and brain regions.

From the above studies, we can see that various brain activities are involved in different stages of emotion processing, including perception, interaction, and regulation. These findings provide the theoretical support and have profound influence for developing affective brain-computer interactions.

2.3 Electroencephalography

Among the techniques of recording brain activity, electroencephalogram (EEG)[77, 78] is one of the most popular approaches, especially in BCIs. The electrical signals of brain activity can be measured with electrodes on the scalp without any surgery (as shown in Figure 2–4). There are some electrical fields with electrochemical signals passing from one neuron to the other. When many these signals are passed simultaneously, these electrical fields sum and become strong enough to be measured from outside the head. Therefore, EEG signals are the smoothed version of the local field potential with high temporal resolution and low spatial resolution. With the fast developments of wearable devices and dry electrodes[79, 80], it is now possible to record EEG signals outside the laboratory. For example, the commercial Emotiv EPOC headset with fourteen electrodes is popular and relatively cheap for various applications[81]. Many advanced dry electrodes and embedded systems are developed to handle the wearability, portability, and practical use of BCI systems in real world environments. However, EEG signals are usually mixed with much noise when recorded and have low signal to noise ratio (SNR), which need advanced signal processing for analysis. Moreover, given the topographical distribution from EEG, it is still very difficult to tell the active brain regions, especially for the deep regions in the brains[82, 83].

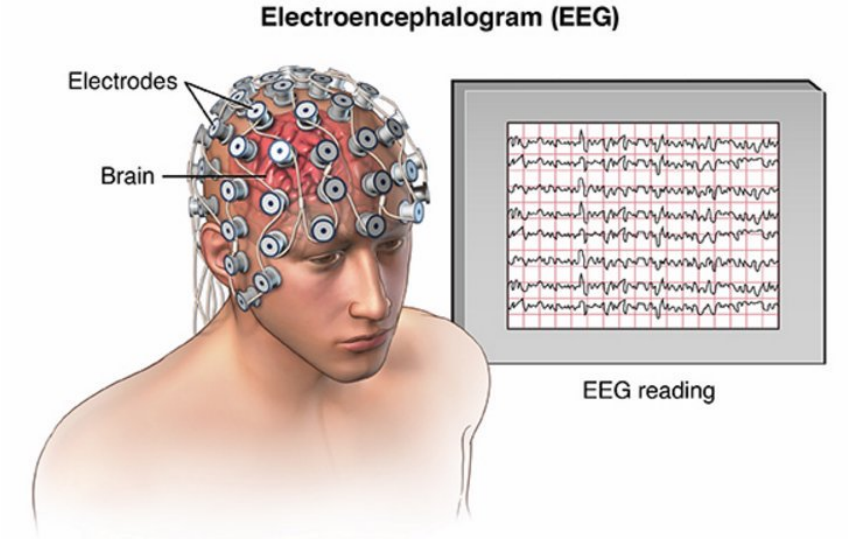


Figure 2–4 EEG recordings, retrieved from Saint Luke's Health System [2].

EEG provides an efficient and direct measurement of the "inner" states of individuals in

[2]https://www.saintlukeshealthsystem.org/health-library/electroencephalogram-eeg

population levels. Previous findings support that there exist some tight links between specific EEG oscillations and different cognitive states, e.g., perception, motivation, and emotion. These specific EEG oscillations are observed with multiple spatial and temporal scales[77]. There are usually five frequency bands: Delta (1-3Hz), Theta (4-7 Hz), Alpha (8-13 Hz), Beta (14-30 Hz) and Gamma (31-50 Hz), for EEG analysis (Figure 2–5). For example, different EEG oscillations are used for sleep stage annotations in sleep research[84, 85]. Alpha oscillations are observed during relaxation and drowsy states. Theta oscillations with sleep spindles and K-complexes occur in non-rapid eye movement (NREM) sleep stage. EEG alpha and beta activities reflect attention and emotion processing[86]. The utilization of EEG allows us to relate the neural oscillations with various cognitive states and possibly find the underlying neural mechanisms.

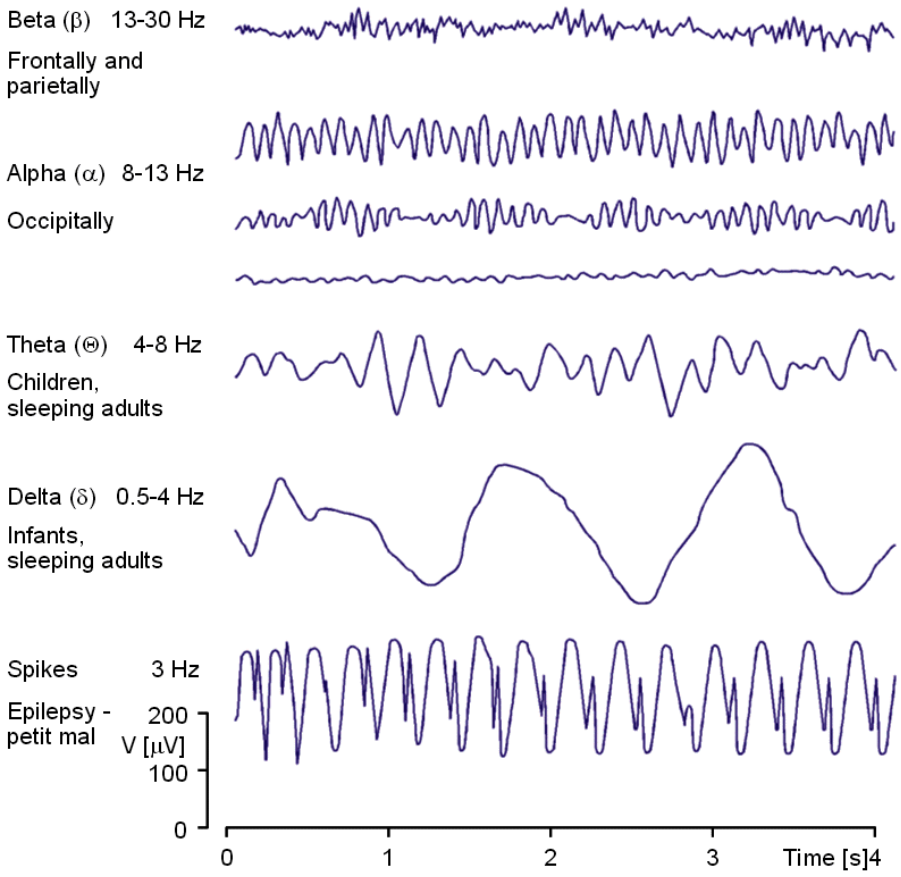


Figure 2–5 Illustrations of EEG waveforms, including Delta waves, Theta waves, Alpha waves, Beta waves, and spikes [3].

[3]http://www.bem.fi/book/13/13.htm

2.4 Emotion Elicitation and Emotion Experiment

In emotion research, one of the technical challenges is how to elicit subjects' target emotions and collect relevant signals in laboratory environments. At present, there are mainly three types of emotion elicitation: 1) spontaneous emotion, participants are exposed in the real world and have spontaneous emotions affected by the surrounding environments; 2) active emotion, participants make themselves in a particular emotional state through mental imagery and memory in a controlled lab environment; 3) evoked emotion, participants evoke emotion through specific emotional stimulation. There are many kinds of stimuli for emotion elicitation, e.g, image, text, music and video. Among these methods, videos are found to be more effective in stable affective processing for longer periods of time and expose subjects to more real-life scenarios, which attract more and more interests in this field[87, 88].

Development and assessment of effective emotion eliciting stimuli provide an important methodological tool for emotion research. However, there are significant individual differences across participants for emotion elicitation. The responses to the same stimuli might vary for different cultural backgrounds, ages and genders. How to qualitatively evaluate and select appropriate emotion eliciting stimuli for target emotions is the critical problem in emotion experiment design. With the great efforts of researchers in relevant research fields in recent decades, there are some standard public emotion stimuli databases with systematical evaluations.

In order to objectively evaluate the elicited emotional states in response to various picture stimuli, Lang and colleagues developed the international affective picture system (IAPS)[89]. They selected some subjects of different countries, genders, ages to rate the picture materials in the database in terms of valence, arousal, and dominance. This work provides the fundamental basis for emotion annotation in emotion experiments. Moreover, they also developed the international affective digital sound system (IADS) to elicit various emotions with sound[90]. Krumhansl developed a music dataset with six music clips for eliciting happy, fear, and sad emotions[91]. Movie clips with both visual and audio stimuli are also one of the popular approaches for emotion eliciting. Many current studies adopt movie clips as emotion stimuli. Philippot *et al.* constructed a film clip dataset with twelve clips for six emotions: happy, angry, disgusted, fear, calm, and sad[92]. Tomarken *et al.* presented another film clip dataset with eight clips for eliciting positive, sad, angry, and disgusted emotions[93]. Gross *et al.* from Stanford recruited 494 English native speaking participants to rate 250 film clips and developed a film clip dataset of 16 clips for eliciting eight emotions: happy, angry, disgusted, fear, sad, surprised, and clam[87]. However, the data sizes of these preliminary studies are limited.

Table 2–1 The summarization of the affective video databases[51].

Name	Size	Emotional Labels
HUMAINE[96]	50 clips from 5 seconds to 3 minutes long	Frame-by-frame level with ratings of intensity, arousal, valence, dominance, predictability, etc.
FilmStim[88]	70 film excerpts from 1 to 7 minutes long	Self-reporting emotions, differential emotions scale (DES), the Positive and Negative Affect Schedule (PANAS)
DEAP[97]	120 one-minute music videos	Self-assessment of arousal, valence and dominance
MAHNOB-HCI[98]	20 film excerpts from 35 to 117 seconds long	Emotional keyword, arousal, valence, dominance
EMDB[99]	52 non-auditory film clips of 40 seconds long	Global ratings of arousal, valence, dominance dimensions
VIOLENT SCENES DATASET[100]	25 full-length movies	Movie segments containing physical violence
LIRIS-ACCEDE[51]	9800 excerpts from 8 to 12 seconds long	Ratings of valence and arousal dimensions

In recent years, more databases with much larger size have been proposed. Baveye and colleagues released a large scale public video database for affective content analysis called LIRIS-ACCEDE[51]. The database consists of 9800 video clips for downloading and each has a duration between 8 and 12 seconds. Affective annotations were achieved using crowdsourcing techniques with 1517 volunteers from 89 countries. Schaefer and colleagues developed a large database called FilmStim of 824 emotional film clips for eliciting fear, angry, sad, disgusted, amused, tender, and neutral emotions[88]. And they recruited 364 participants to rate each clip with several criteria such as self-reported emotion, differential emotions scale (DES)[94] and the Positive and Negative Affect Schedule (PANAS)[95]. The popular affective video databases are summarized in Table 2–1.

The protocols of emotion experiments usually contains many trials in each session. Each

trial has a phase of emotion elicitation with stimuli following by self-assessment about feelings. Self-Assessment Manikins (SAM) is one of the popular self-reporting tools in the literature[101]. SAM contains different dimensions: arousal, valence, and dominance in emotion models and does not need any words for understanding (as shown in Figure 2–6).

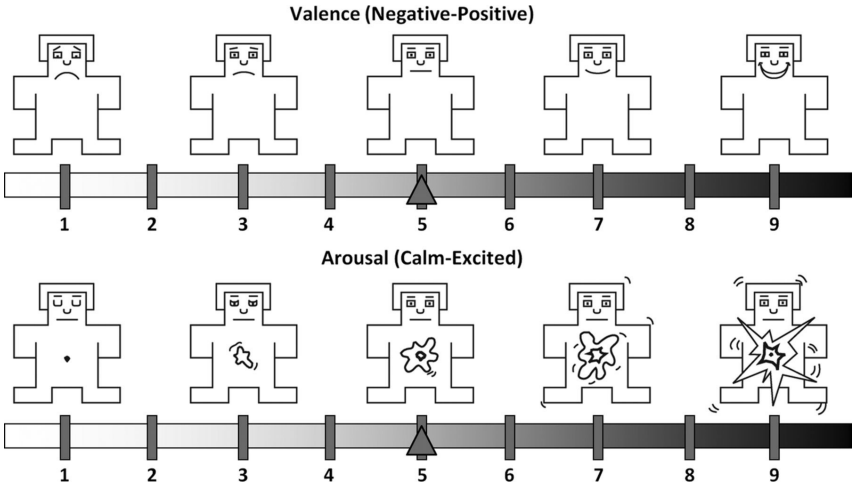


Figure 2–6 Self-Assessment Manikins (SAM) with the two affective dimensions: valence and arousal.

2.5 Emotion Recognition

2.5.1 EEG-based Emotion Recognition

Affective brain-computer interactions aims to narrow the communication gap between highly emotional humans and emotionally challenged machines by developing intelligent systems to detect, process, and respond to users' emotional states using direct brain activities. Affective brain-computer interactions are closed-loop affective computing systems with combination of both detection and feedbacks. Figure 2–7 shows the illustration of the closed-loop affective brain-computer interactions. There are several phases in this cycle. The brain activities are recorded when participants are exposed to the stimuli or real-world scenarios. The raw data are further preprocessed for filtering noise and artifacts. The efficient features are extracted and then we train a classifier or regression model based on these features. After recognizing participants' emotional states, efficient feedbacks are introduced for emotion regulation. Nowadays, most researches of affective computing pertain to the assumption that emotions are passive and disregard the effects of emotion regulation when studying emotion processing. However, emotion processing is a dynamic system and emotion regulation plays an important role in emotion

processing, which should gain considerable research attention[102].

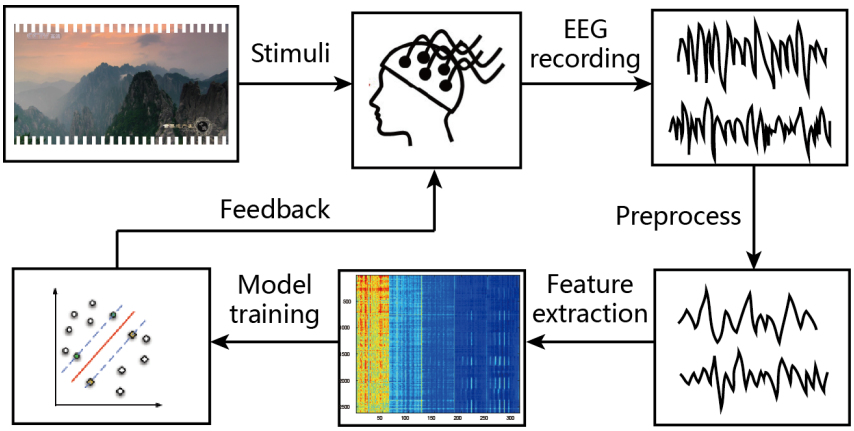


Figure 2–7 Closed-loop affective brain-computer interactions.

Emotion recognition is the primary and important phase in affective brain-computer interaction cycle. Recently, Alarcao and Fonseca gave a very detailed survey about emotion recognition with EEG data, including the comparisons and discussions of stimuli, feature extraction, and classifier[103]. Kim *et al.* also presented a survey of computational models of emotion recognition using EEG[104]. Jenke *et al.* focused on the systematical comparisons about feature extraction and selection for EEG-based emotion recognition and they found advanced feature extraction and selection with multivariate methods are superior than conventional methods[105]. There are various EEG features: 1) time domain features: event related potentials, statistics, fractal dimension, higher order crossings; 2) frequency domain features: band power, higher order spectra; 3) time-frequency domain: Hilbert-Huang spectrum discrete wavelet transform; 4) asymmetry and electrode combinations. Wang *et al.* investigated the performance of three EEG features, namely power spectrum features, wavelet features and nonlinear dynamical features[106]. They visualized the trajectory of emotion changes with manifold learning. Mühl and colleagues presented a survey of affective brain-computer interfaces and discussed the state-of-the-art and challenges of affect detection from brain activities[27]. With the great efforts of many researchers from various studies, various approaches to EEG-based emotion recognition have been proposed. We organize some of these studies in Tables 2–2 and 2–3. These studies demonstrated the efficiency and feasibility of EEG-based emotion recognition.

One of the research goals of affective brain-computer interactions is to investigate whether neural patterns of specific emotions across individuals exist. Some findings about EEG correlates of emotions have been proposed, for example, the lateralization effect. The EEG activities

Table 2–2 Various studies on emotion classification using EEG (continued on next page).

Study	Stimuli	#Chan.	Method Description	Emotion states	Accuracy	Pattern Study
[107]	IAPS, IADS	3	Power of alpha and beta, then PCA, 5 participants, classification with FDA	Valence and arousal	Valence: 92.3%, arousal: 92.3%	×
[108]	IAPS	2	Amplitudes of four frequency bands, 17 participants, evaluated KNN, Bagging	Valence (12), arousal (12) and dominance (12)	Valence: 74%, arousal: 74%, and dominance: 75%	×
[109]	Video	62	Wavelet features of alpha, beta and gamma, 20 participants, classification with KNN and LDA	disgust, happy, surprise, fear and neutral	83.26%	×
[110]	Music	24	Power spectral density and asymmetry features of five frequency bands, 26 participants, evaluated SVM	Joy, anger, sadness, and pleasure	82.29%	√
[111]	IAPS	8	Spectral power features, 11 participants, KNN	Positive, negative and neutral	85%	×
[112]	IAPS	4	Asymmetry index of alpha and beta power, 16 participants, SVM	Four quadrants of the valence-arousal space	94.4% (user-dependent), 62.58% (user-independent)	×
[97]	Video	32	Spectral power features of five frequency bands, 32 participants, Gaussian naive Bayes classifier	Valence (2), arousal (2) and liking (2)	Valence: 57.6%, arousal: 62% and liking: 55.4%	×
[113]	Music	14	Time-frequency (TF) analysis, 9 participants, KNN, QDA and SVM	Like and dislike	86.52%	√
[98]	Video	32	Power spectral density features of five frequency bands, modality fusion with eye track, 24 participants, SVM	Valence (3) and arousal (3)	Valence: 68.5%, arousal: 76.4%	×

Table 2–3 Various studies on emotion classification using EEG (continued from previous page).

Study	Stimuli	#Chan.	Method Description	Emotion states	Accuracy	Pattern Study
[114]	Video	62	Power spectrum features, wavelet features, nonlinear dynamical features, 6 participants, SVM	Positive and negative	87.53%	√
[105]	IAPS	64	Higher order crossings, higher order spectra and Hilbert-Huang Spectrum features, 16 participants, QDA	Happy, curious, angry, sad, quiet	36.8%	√
[115]	Music	19	Asymmetry measures and connectivity measures, 31 participants, principal component analysis	Pleasantness, energy, tension, anger, fear, happiness, sadness, and tenderness	/	√
[116]	Video	14	Power spectrum features and asymmetry, 30 participants	Valence and arousal	86.63%	×
[117]	Video	64	Correlation, coherence, and phase synchronization features, QDA	Positive, negative and neutral	82%	√
[118]	Video	32	Time-frequency analysis, 9 participants, LDA, SVM	Amused, disgusted, sad and neutral	97.2%	×
[119]	Pictures	62	Common spatial patterns, 10 participants, SVM	Happy and sad	93.5%	×

IAPS and IADS stand for the International Affective Picture System and the International Affective Digital Sounds, respectively. The numbers given in parentheses denote the numbers of categories for each dimension. Pattern study indicates whether the work reveals the neural activities (critical brain areas and critical frequency bands) that share commonality across participants or sessions. The classifiers include K Nearest Neighbors (KNN), Fisher's Discriminant Analysis (FDA), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Bagging, and Support Vector Machine (SVM).

of the left prefrontal areas enhance for positive emotions, whereas the right prefrontal areas have lower EEG responses for negative emotions[57]. Muller *et al.* found negative emotion has higher gamma responses (30-50 Hz) in the left temporal areas[120]. Nie and colleagues reported that subject-independent EEG features are mainly located in the right occipital and parietal lobes in the alpha band, the central site in beta band, and the left frontal lobe and right temporal lobe in the gamma band[121]. Mu and Lu presented that the gamma activity is associated with emotion recognition[119]. Balconi et al. found that frequency oscillations are highly influenced by valence and arousal. In comparison with stimuli with low activation, the frequency oscillations have significant changes for high activation stimuli[122]. Besides the relation between the response differences of the left and right hemispheres in EEG, some studies also provide evidences for the relation between event-related potentials (ERPs) and emotion[123, 124].

Although some findings about the neural patterns (critical frequency bands and brain regions) of different emotions have been proposed (see Tables 2–2 and 2–3), whether these neural patterns are common and stable across subjects and across sessions still needs further investigation. In clinical applications, the internal consistency and test-retest stability of EEG have been studied[125, 126]. The stability and reliability of EEG patterns in these studies are mainly based on some statistical parameters, e.g., correlation confidences within classes, rather than machine learning approaches. However, few studies have been conducted about the stability of neural patterns for EEG-based emotion recognition.

Lan and colleagues presented a preliminary study on the stability of EEG features in each channel with correlation analysis[127]. However, only a small group of participants was included and the number of EEG channels was limited with fourteen. The intra-class correlation coefficients of EEG features from the same channels and the same emotions were computed individually, rather than evaluating different features together as a group. Lin *et al.* explored the dat-to-day EEG variability for emotion recognition[128]. And they found that the data distributions of same emotion across days tended to scatter wider than the data distributions of different emotion within a day. The individual variability and temporal variability of EEG signals pose a technical challenge for building robust affective brain-computer interaction systems in real-world applications. The performance of computational models over time still remains an unsolved problem. Therefore, one of my major contributions in this thesis is to investigate the stale EEG patterns over time with machine learning algorithms. The factors like individual difference and temporal evolution should be considered to make emotion recognition models automatically adaptable. One possible way to dealing with these problems is to adopt transfer

learning techniques[129].

Conventional machine learning approaches depend on the assumption that training data and test data should be subject to the same data distribution. In EEG-based emotion recognition, it is useful to improve the model generalization by collecting data from more various environments and subjects for training. However, it is very expressive to record high quality labeled EEG for building subject-specific models. Therefore, we consider how to learn basic knowledge about emotion recognition from history data and transfer knowledge for future predictions. Transfer learning, a subfield of machine learning, has been proposed for dealing with these problems and demonstrated efficient in various tasks[129]. Jayaram *et al.* presented a systematical survey about transfer learning in brain-computer interfaces[130]. Samek *et al.* proposed an approach to transferring nonstationarity information across subjects and achieved significantly increased performance[131].

2.5.2 Multimodal Emotion Recognition

Emotion is a very complex psycho-physiological process, which is manifested via internal physiological responses and external expressions. Signals from different modalities describe different aspects of emotion and complementary information from different modalities can be integrated to build a more robust emotion recognition model compared to current unimodel approaches[132]. Multimodal fusion technologies can integrate various sensor data with multiple levels of feature representations in various ways and make final decisions based on these multidimensional information. There are many modality signals that are associated with emotion processing, e.g., facial expression, voice, gesture, and physiological signals. Figure 2–8 shows various modalities that are commonly used in health informatics. In comparison with unimodal approaches, multimodal approaches have the following advantages: 1) developing more accurate models; 2) possible solutions to missing data; 3) potential solutions to noisy channels. Besides these opportunities, there are also a number of challenges, e.g., how to efficiently combine various sensor data and construct multimodal models, how to handle missing data and different sampling rates.

There are mainly two kinds of conventional multimodal fusion strategies in the literature: feature-level fusion and decision-level fusion. Feature-level fusion extracts features of individual modality and directly concatenate them into a larger feature vector as a input of a classifier. For decision-level fusion, several basic classifiers are learned from different modalities and the outputs are combined to obtain final predictions using some rules (e.g., maximal and sum rules) or learning algorithms (e.g., fuzzy integral).

Figure 2–8 Various modalities that are commonly used in health informatics[133]

D'mello and Kory presented a review and meta-analysis about multimodal emotion detection[134]. Most studies focus on the fusion of visual and auditory modalities at first, whereas more other modalities about emotion processing receive increasing interest in recent years. Koelstra utilized facial expression and EEG for multimodal affective implicit tagging using feature-level fusion and decision-level fusion[135]. Nicolaou *et al.* proposed to fuse facial expression, shoulder gesture, and audio signals for continuous emotion predictions in the valence and arousal space[136]. They used the bidirectional Long Short-Term Memory neural networks for capturing the past and future relations. The output-associative fusion approach was proposed in order to incorporate correlations and covariances between the affective dimensions. Lin and colleagues proposed to fuse both EEG signals and musical contents for the classification of valence and arousal and investigated the relations between these two modalities[137]. Wagner *et al.* focused on the challenge of missing data in multimodal emotion recognition. They compared and discussed the performance of various fusion methods on the conditions of temporarily unavailable modalities[138]. Tsalamlal *et al.* presented a study about combining facial expression and touch for multimodal affective sensing[139]. Schirmer and Adolphs conducted the study about emotion

perception with multiple modalities: facial expression, voice, tactile, EEG and fMRI. They reviewed their similarities and differences across different modalities and proposed multisensory fusion methods in different phases[140].

Recently, more multimodal fusion frameworks with deep neural networks have been proposed for leveraging the advantages of different modalities. Ngiam and colleagues proposed a multimodal deep learning framework for learning the high-level shared representations over audio and image modalities[141]. There was an autoencoder network for each individual modality. Tzirakis and colleagues proposed a novel multimodal deep neural network for end-to-end emotion recognition[142]. They constructed the speech network and visual network with convolutional neural network and deep residual network, respectively. A two-layer LSTM was trained on the top of these two networks for modeling the contextual structures. Despite the above progress, how to construct EEG-based multimodal emotion recognition with deep neural networks remains largely unexplored.

Cognitive psychology studies support that emotion is a complex process with generation, development, and lessening, caused by specific scenarios. We should consider specific contexts, rather than simply judge according to some external behaviors and ignore the sentiment and context capacities for emotion recognition. Humans interact with the surrounding environments or stimuli through various modalities and emotion perception should consider the specific context[143]. For example, for the same tearing behaviors, it might be for extreme happiness or extreme sadness dependent on specific scenarios. Therefore, developing context aware emotional intelligence receives more and more interests from researchers. Ptaszynski and colleagues proposed to introduce contextual analysis to emotion processing for building emotional intelligence[144].

Among the technologies, eye tracking can naturally provide various eye movement parameters and observe users' subconscious behaviors, which provide important clues for users' current activity contexts[145]. Eye movements have long been studied for measuring users' behaviors and cognitive states in various tasks[146–148]. Many previous studies support that pupillary responses are associated with cognitive and emotional processing. Partala *et al.* studied the pupillary responses under different affective stimuli and they found that the pupil size of negative stimuli was significantly larger than that of neutral and positive stimuli[149]. Pupillary changes are larger during emotionally arousing picture viewing and associated with autonomic activity measured by heart rate and skin conductance[150].

The combination of brain signals and eye movements is an efficient approach for measur-

ing users cognitive states, e.g., emotion and attention. Many studies have utilized these two different techniques together for obtaining complementary information from different aspects. Zekveld *et al.* used fMRI to investigate the neural correlates of pupil size in the listening load task[151]. Their results showed that pupil dilation can be a indicator for studying various brain regions in speech perception. Soleymani and colleagues used EEG, pupillary response and gaze distance for emotion recognition and achieved the classification accuracies of 68.5% and 76.4% for valence and arousal, respectively[98]. Langer and colleagues developed a large dataset with both EEG and eye tracking data of 126 individuals for investigating information processing in the developing brain[152]. The tasks contained both the active and passive paradigms and the corresponding cognitive tests and questionnaires were included. Wang *et al.* studied the dynamic functional connectivity of vigilance fluctuation using fMRI[153]. They used eye tracking techniques for measuring the degree of eye closure as a indicator of vigilance states. In this way, they did not need the usage of an intrusive stimuli or behavior probe as commonly used in the literature. Laeng *et al.* presented a brief review about the correlation between the activity of the locus coeruleus and pupillary dilation[154]. Murphy and colleagues found that using pupil diameter can help understand the role of the locus coeruleus in human cognition[155]. Pupillary response provides a possible estimation of the mental states and a very promising tool in various psychology studies.

Considering the feasibility of eye tracking in real-world applications, nowadays eye tracking techniques can be well embedded in recent popular user-centered wearable devices, e.g., virtual reality and augmented reality. Many wearable eye tracking glasses have been developed and there are some commercial products (e.g., SMI eye tracking glasses [1]). At the Consumer Electronics Show 2016 in Las Vegas, a startup company called Looxid Labs presented a prototype headset combining two-electrode frontal EEG and eye tracking for device controls [2]. Krafka and colleagues developed GazeCapture to implement eye tracking using mobile devices with low cost and no additional hardware[156].

2.5.3 Public Emotion EEG Datasets

Although there already have been some public emotion eliciting databases, the publicly available emotion datasets are limited, especially for affective brain-computer interactions. There are two popular public emotion EEG datasets: MAHNOB HCI[98] and DEAP[97]. MAHNOB HCI [3]

[1] https://www.smivision.com/eye-tracking/product/eye-tracking-glasses/

[2] http://looxidlabs.com/

[3] https://mahnob-db.eu/hci-tagging/

consists of 32-channel EEG signals, peripheral physiological signals, face and body video, eye gaze and audio data of 27 participants while watching 20 emotional videos. DEAP [1] includes EEG and peripheral physiological signals of 32 subjects and each video has ratings of valence, arousal, like, dominance, and familiarity. In these datasets, each participant performed the emotion experiments only once. Therefore, it is very difficult to investigate the sable EEG patterns of different emotion and the adaptivity of computational models over time. In this thesis, we develop a new emotion EEG dataset called SEED for studying the stability of neural patterns for emotion recognition. Different from MAHNOB HCI and DEAP, SEED dataset contains EEG data from sessions in different days for the same participants. And we study many scientific questions (e.g., critical frequency bands and brain regions, stability of EEG patterns over time) in the research field of affective brain-computer interactions.

2.6 Driving Fatigue and Vigilance Estimation

Besides emotion, vigilance is another important component among various cognitive states. Vigilance refers to the ability to endogenously maintain the focus or attention. Various working environments require sustained high vigilance, particularly for some dangerous occupations such as truck drivers and high-speed railways. In these situations, vigilance decrement or attention lapse can cause severe accidents in the public transportation. Driving fatigue is reported as a major factor in fatal road accidents every year. The National Highway Traffic Safety Administration (NHTSA) of the US government held a forum for discussion about the problem of driving fatigue[31]. For actively enhancing driving safety, it is attractive to design an affective user interface for monitoring drivers' emotional and cognitive states and response to drivers to regulate their states via multimedia. As the blood alcohol tests in drunk driving, researchers aim to find a standardized method for measuring human vigilance states. However, it is still challenging to reliably quantify drivers vigilance states.

There are many research challenges in this field. Vigilance states are dynamic changing process of intrinsic mental states with temporal evolution instead of a time point. This process cannot simply be treated as a function of the duration of time when engaged in tasks. It is difficult to evaluate vigilance states without using an intrusive stimulus or a behavior probe[153]. Real-world applications require continuous vigilance estimation with high temporal resolution or accurate predictions in advance[34].

[1] http://www.eecs.qmul.ac.uk/mmv/datasets/deap/

Considerable progress has been achieved in vigilance estimation over the past decades. Various approaches have been proposed in the literature[157–159]. Dong *et al.* summarized a detailed review of driving vigilance monitoring systems[160]. Most of these vigilance estimation are based on external behaviors, e.g., head nodding, yawning, and eye closure, and internal physiological signals. The popular approach to vigilance estimation is called PERCLOS[161], which detects eye states and computes the percentage of eye closure using cameras. This approach is easy for setups and has low cost for implementation. However, the performance can be highly degraded due to the illustration changes and heavy occlusion. Moreover, it is very difficult for predictions in advance using solely external behaviors. How to identify reliable and efficient biomarkers for vigilance estimation remains still a challenge in relevant research fields.

Among various modalities, EEG signal is considered as a promising approach to estimating vigilance because of its high temporal resolution and noninvasive low-cost properties[33, 162–165]. EEG signals directly reflect human brain activity of various cognitive states. Compared with existing approaches using external behaviors and physiological changes, EEG has intrinsic potential to allow vigilance estimation in advance[34]. Martel and colleagues found there were increased alpha responses (8-14 Hz) for vigilance decrement and they developed a lapse detection approach with up to 10 s before onset[35]. Many studies observe that posterior alpha oscillations are dominant during eye closure (called "idling rhythm")[166] and then exhibits a slight decrease from alertness to drowsiness[167]. Shi and Lu presented a study of EEG-based vigilance estimation using extreme learning machines with L_1 norm and L_2 norm[164]. Lin *et al.* developed a closed-loop system combining both lapse monitoring and feedback assessment with EEG spectra[168]. They also presented how to estimate the efficiency of the warning feedbacks for regulating drowsy states. Rosenberg *et al.* tried to pinpointed the neuromarker for sustained attention with brain functional connectivity and developed the sustained attention network for predictions[169]. O'Connell *et al.* examined the temporal dynamics of EEG for attentional lapse and they found the specific neural signatures of EEG with about 20 s before a performance error[170].

Based on the above findings about the neural mechanisms of vigilance and attention, researchers have developed various wearable approaches and prototypes for vigilance estimation in real-world applications. For example, Lin and colleagues designed a wireless and wearable EEG system for drivers vigilance estimation and performed evaluations in a virtual driving environment[171, 172]. More improved high performance dry electrodes and EEG systems have been developed in recent years[79, 173, 174].

Besides EEG signals, electrooculogram (EOG) signals are shown effective in detecting eye states that are relevant to vigilance estimation. EOG has a higher signal-to-noise ratio and much lower cost for recording implementations. EOG signals can capture most characteristic eye movements, e.g., blink, saccade, and fixation. Various vigilance estimation approaches with EOG have been proposed in the literature[175–177]. Among various eye movement parameters, spontaneous eyelid closures are found to be an efficient indicator for non-intrusive vigilance estimation[153]. Damousis and Tzovaras presented a fuzzy combination of various eyelid activity parameters for predicting sleep onset and preventing related accidents[178]. Zhang and colleagues fused multimodal modalities including eye gaze, EEG, and peripheral signals to measure cognitive load in driving tasks for subjects with autism spectrum disorder[179]. The approach allowed the personalization of driving skill training with cognitive load evaluations.

However, most proposed vigilance estimation systems are designed and evaluated in the controlled laboratory environments[180, 181]. Very few studies have been performed in real-world environments, e.g., real driving under different weather conditions, due to the more complex challenges compared with indoor simulations. Papadelis *et al.* designed an on-board electrophysiological system for monitoring sleepiness[175]. They examined the relation between neurophysiological signal statistics and driving time and errors. Healey and Picard used physiological signals for stress detection in real-world driving scenarios[182]. A recent study conducted by Wang and colleagues presented real time driving fatigue detection and performed evaluations in the driving route of about 252 km long[183]. But the channels of EEG recordings were limited with only two electrodes (O1 and O2). More innovative technologies including both sensors and algorithms should be developed for filling the reality gap between laboratory simulations and real-world environments.

2.7 Summary

The Chapter presents a background review of the related research in the thesis. We firstly introduce the emotion definition and emotion models. Some concepts in emotion research are still controversial and we follow some common definitions in the thesis. The neural mechanisms of emotion are introduced and provide theoretic support for building affective brain-computer interaction systems. Some details about popular emotion experiment protocols are also described. Then, we summarized the studies about emotion recognition using EEG and multimodal approaches. The availability of relevant public emotion EEG datasets is also presented. Moreover, we introduce the research background and the recent progress of the other applica-

tion vigilance estimation. The current challenges of the related research are discussed and our proposed approaches to handling these problems will be shown in the rest of the thesis.

Chapter 3

Experimental Setups

In this Chapter, we present the detailed experimental setups and the data processing for emotion recognition and vigilance estimation. Section 3.1 presents an EEG dataset for three-class emotions (happy, sad, and neutral) called SJTU Emotion EEG Dataset (SEED) including the emotion elicitation materials, subject information, experimental protocols, and data recordings. Based on SEED, another EEG dataset called SEED-IV for four emotions (happy, sad, fear, and neutral) is developed and described in Section 3.2. The data processing for multimodal emotion recognition is introduced in Section 3.3, including feature extraction and feature smoothing for EEG and eye movements. We develop a simulated driving system in the laboratory and release a multimodal dataset with EEG and forehead EOG for vigilance estimation (SEED-VIG) in Section 3.4. Moreover, we implement a wearable prototype with flexible dry electrodes and an acquisition board and perform evaluations in both simulated and real driving environments (Section 3.5). These datasets are the basic supports for our further analysis in affective brain-computer interactions. In order to enhance cooperations in this research field, these datasets are freely available to the academic community via the project website of SEED[1]. Section 3.6 concludes the Chapter. The work in this Chapter has been published at IEEE Transactions on Affective Computing[184], IEEE Transactions on Cybernetics[185], and Journal of Neural Engineering[186].

3.1 SJTU Emotion EEG Dataset (SEED) for Three Emotions

3.1.1 Emotion Stimuli

One of the challenges in affective brain-computer interactions is how to design efficient emotion elicitation stimuli and collect emotion data. As we review in Section 2 of Chapter 2, there are many kinds of emotion stimuli used in emotion research, e.g., image, music, and video. Among these elicitation approaches, emotional film clips with both scene and audio are shown reliable and effective for evoking subjects' emotions in the laboratory[87, 88]. The contexts in emotional

[1] http://bcmi.sjtu.edu.cn/%7Eseed/

Table 3–1 Details of the film clips used in our emotion experiment

No.	Labels	Film clip sources	#clips
1	Sad	Tangshan Earthquake	2
2	Sad	Back to 1942	3
3	Happy	Lost in Thailand	2
4	Happy	Flirting Scholar	1
5	Happy	Just Another Pandora's Box	2
6	Neutral	World Heritage in China	5

film clips expose subjects to real-life scenarios and elicit strong emotional changes. Therefore, we used emotional film clips as stimuli in our emotion experiments. Only Chinese film clips were selected since native culture factors might affect elicitation.

We manually selected a pool of emotional film clips from famous Chinese films in the preliminary study. For further evaluations of these clips, twenty participants were recruited to assess their emotions using scores (1-5) and keywords (happy, sad, and neutral) after watching each clip. The film clips were selected according to the following criteria: 1) the length of the whole experiment should not be too long in case it will give the participants visual fatigue; 2) the videos should be understood without explanation; and 3) the videos should elicit a single desired target emotion. Based on the ratings from the twenty participants, 15 film clips for happy, neutral, and sad emotions were selected as stimuli from the material pool with the mean ratings of higher than or equal to 3 points. Each emotion had five corresponding emotional clips. The duration of each film clip was about 4 minutes. Each film clip was further edited in order to create coherent emotion elicitation. The details of these film clips used in the experiments are listed in Table 3–1. Figure 3–1 presents the posters of the selected Chinese films.

3.1.2 Subjects

Fifteen participants (7 males and 8 females; age range: 19-28 years old, mean: 23.27, std: 2.37), different from those in film clip selection, participated in the data recording experiments. None of them have any history of mental diseases. In order to investigate the neural signatures and stable patterns across sessions and individuals, each participant was required to perform the experiments for three sessions. The time interval between two sessions was one week or longer.

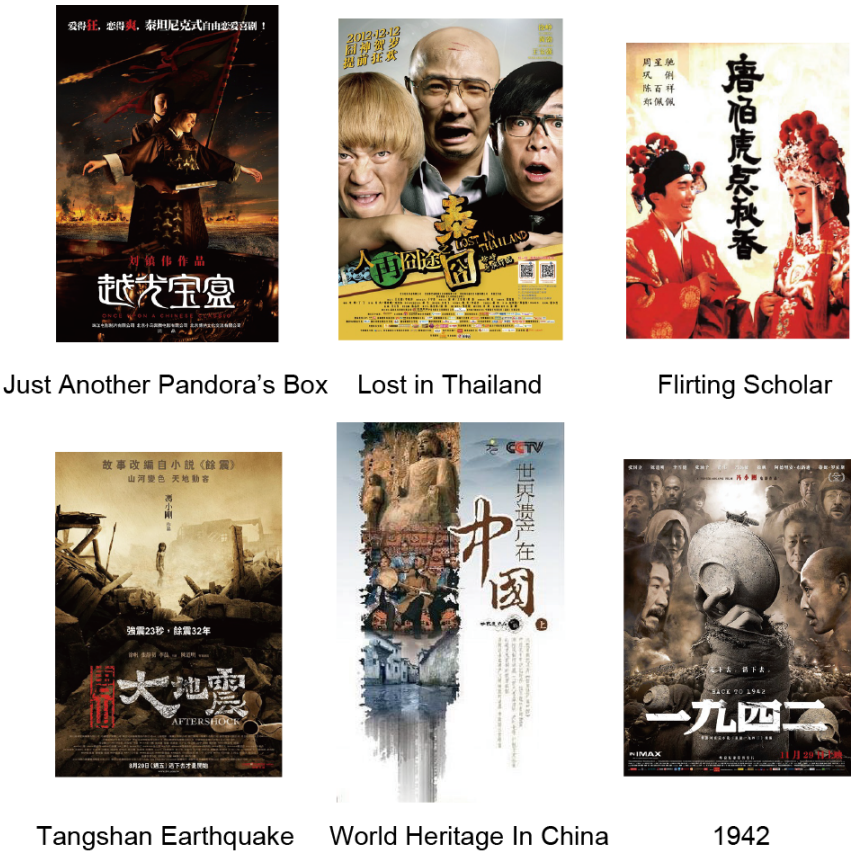


Figure 3–1 The selected Chinese films in the emotion experiments.

All the participants were native Chinese students from Shanghai Jiao Tong University with self-reported normal or corrected-to-normal vision and normal hearing. Before the experiments, the participants were informed about the research background and instructed to sit comfortably, watch the forthcoming movie clips attentively without diverting their attention from the screen, and refrain as much as possible from overt movements. We used Eysenck Personality Questionnaire (EPQ)[187] for assessing the personality traits of a person. The questionnaire contains three independent dimensions: Extraversion/Introversion, Neuroticism/Stability and Psychoticism/Socialisation.

3.1.3 Experiment Protocol

The experiments were performed in a quiet environment in the morning or early in the afternoon. Figure 3–2 presents the detailed protocol of the emotion experiments. In total, there were 15 trials (film clips) for each experiment. Each trial had a 15-s hint of start before each clip and 10-s

feedback after each clip. For the feedback, participants were told to report their own emotional reactions to each film clip by completing the questionnaire immediately after watching each clip. Inspired by the previous study[92], the questions were as follows: (1) what they had actually felt in response to viewing the film clip; (2) how they felt at the specific time they were watching the film clips; (3) whether they had watched the movie before; and (4) whether they had understood the film clips. They also rated the intensity of subjective emotional arousal using a 5-point scale according to what they actually felt during the trials.

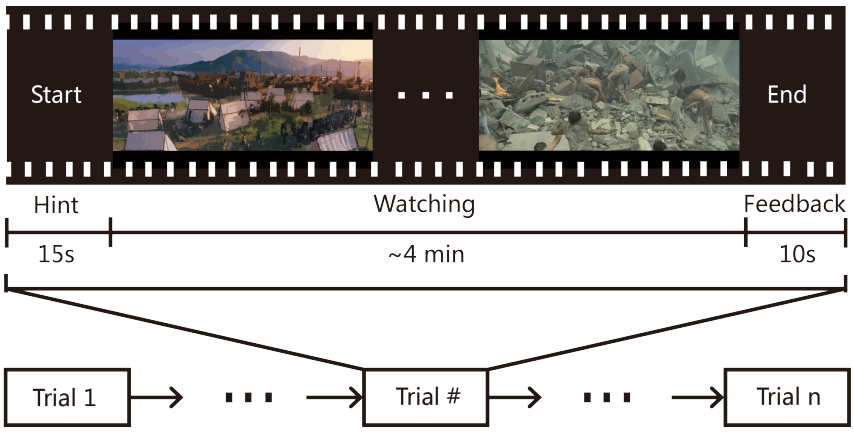


Figure 3–2 The protocol used in the emotion experiment.

Figure 3–3 shows the experimental scene in the emotion experiments. Facial videos, EEG and eye tracking data were recorded simultaneously. EEG was recorded using an ESI NeuroScan System[1] at a sampling rate of 1000 Hz from a 62-channel active AgCl electrode cap according to the international 10-20 system. The layout of EEG electrodes on the cap is shown in Figure 3–4. The impedance of each electrode had to be less than 5 kΩ. The frontal face videos were recorded from the camera mounted in front of the participants. Facial videos were encoded in AVI format with a frame rate of 30 frames per second and a resolution of 160×120. Eye tracking was recorded using SMI ETG eye tracking glasses [2] as shown in Figure 3–5.

3.2 SJTU Emotion EEG Dataset (SEED-IV) for Four Emotions

The experimental setup of SJTU Emotion EEG Dataset (SEED-IV) for four emotions is similar to SEED dataset. SEED-IV is developed based on SEED with more emotion categories, stimuli databases, and systematical evaluations. The development goal of SEED-IV is not only

[1] http://compumedicsneuroscan.com/

[2] https://www.smivision.com/eye-tracking/product/eye-tracking-glasses/

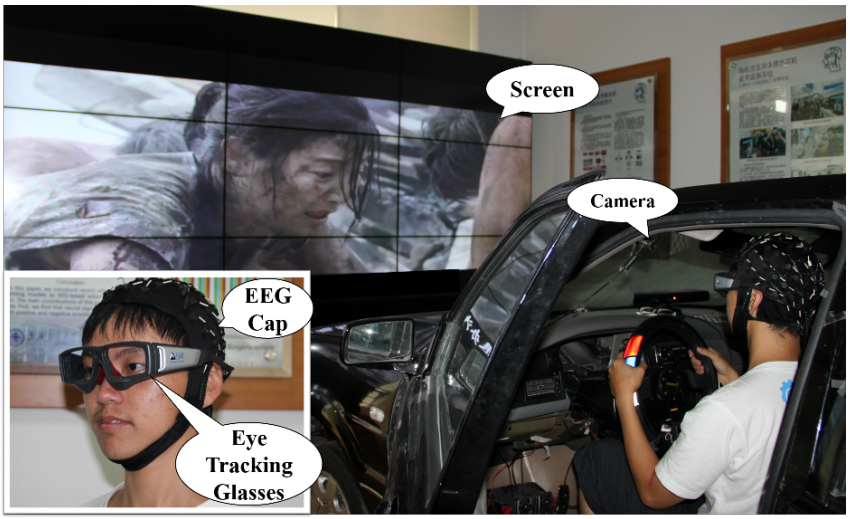


Figure 3–3 The experimental scene in the emotion experiments.

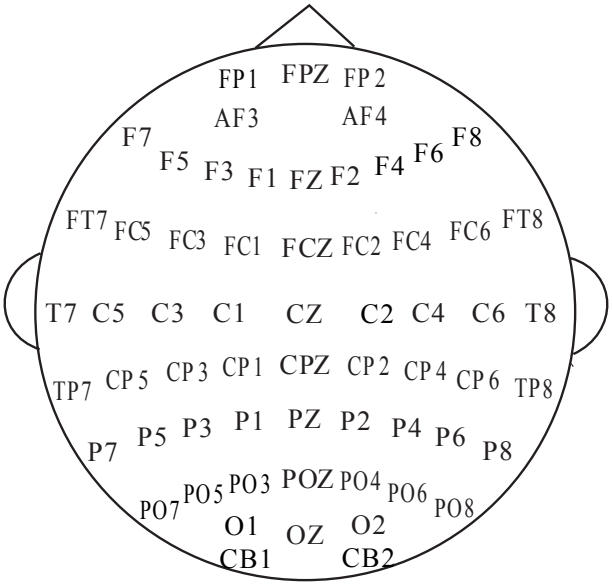


Figure 3–4 The layout of 62 EEG electrodes on the cap.

to explore the representation characteristics of various modalities for more emotions, but also to investigate the feasibility and wearability of affective brain-computer interactions. In this dataset, we focus on how to obtain the comparable performance with much fewer electrodes for setups.

Figure 3–5 The SMI ETG eye tracking glasses used in the study and the pupillary image captured in one experiment.

3.2.1 Emotion Stimuli

Before the experiments, we selected totally 168 film clips as our material pool for four emotions (happy, sad, fear, and neutral). Forty-four participants (22 females, all college students) were recruited to assess their emotions when watching the film clips with keywords of the four discrete emotions (happy, sad, neutral, and fear) and ratings out of ten points (from -5 to 5) for two dimensions: valence and arousal. The valence scale ranges from sad to happy. The arousal scale ranges from calm to excited. The mean rating distributions of the different film clips are shown in Figure 3–6. As shown in this figure, there are significant differences between the conditions in terms of the ratings of valence and arousal, reflecting the successful elicitation of the targeted emotions in the laboratory environments. Finally, 72 film clips were selected from the pool of materials that received the highest match across participants. The stimuli of these selected clips generally resulted in the elicitation of the four target emotions. The duration of each film clip was approximately two minutes.

3.2.2 Subjects

A total of 15 healthy, right-handed participants (7 males and 8 females, age between 20 and 24 years) joined the experiments. Before each experiments, the participants were informed of the purpose and procedure of the experiments and also the harmlessness of the whole experiments. Each participants performed the experiments three times in different days. The stimuli used in three sessions were completely different. Therefore, there were totally 45 experiments in this dataset.

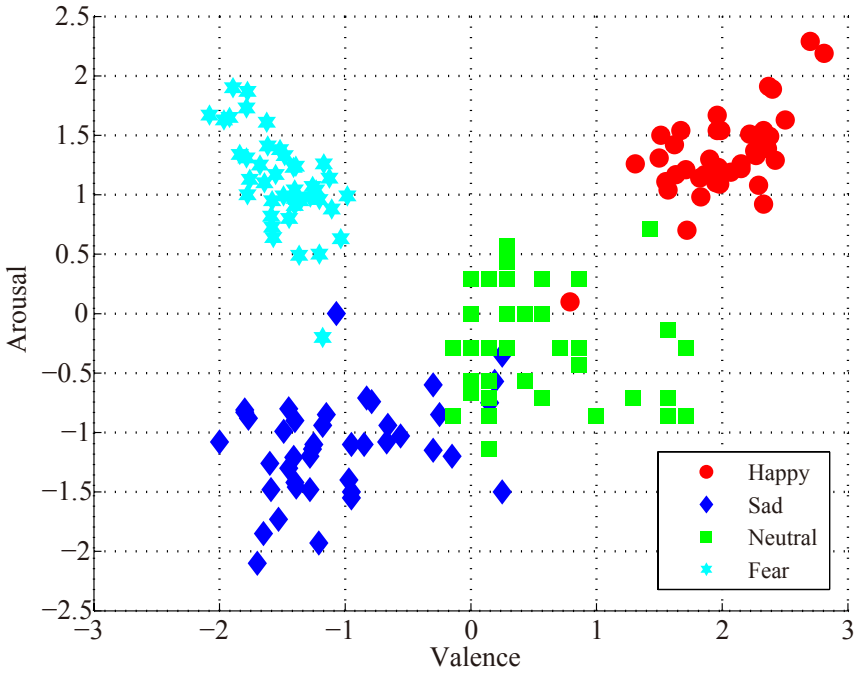


Figure 3–6 The mean rating distributions of the different film clips on the arousal-valence plane for four emotions. The ratings are clustered into four classes: happy, sad, fear, and neutral emotions.

3.2.3 Experiment Protocol

To avoid repetition, each film clip was presented only once. To investigate the stability of our model over time, we designed three different sessions for each participant on different days. Each session consisted of 24 trials (6 trials per emotion), and the stimuli for the three sessions were completely different. Figure 3–7 presents the detailed protocol of our designed emotion experiments. Each film clip had a 5 s hint for starting and a 45 s self-assessment with the PANAS scales[95] after each clip. The participants were asked to watch the emotional clips and elicit the corresponding emotions. The ratings of the subjects were based on how they actually felt while watching the clips rather than what they thought the film clips should be. According to the feedback, if the participants failed to elicit the correct emotions or the arousal emotions were not strong enough, the data were discarded.

In order to investigate the performance of electrode reduction, we selected the six symmetrical electrodes in the temporal areas above the ears. These electrode placements were FT7, FT8, T7, T8, TP7, and TP8 of the international 10–20 system shown in Figure 3–8. These electrodes can be easily embedded in a wearable headset or spectacle frames. In order to compare the performance of electrode reduction with the original full electrodes, we also simultaneously

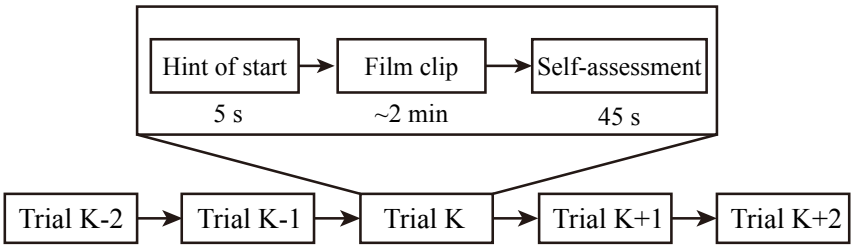


Figure 3–7 The protocol of SEED-IV for four emotions.

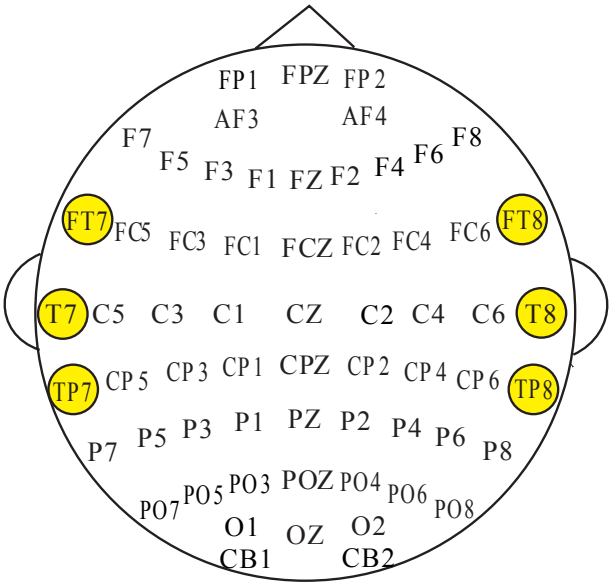


Figure 3–8 The EEG electrode layout of 62 channels. Six symmetrical temporal electrodes (*FT7*, *FT8*, *T7*, *T8*, *TP7*, and *TP8*) are selected in SEED-IV.

recorded 62-channel EEG data with the international 10-20 system using the ESI NeuroScan System and eye movements using SMI ETG eye-tracking glasses. Figure 3–9 shows the EEG and eye tracking setup in the experiments.

3.3 Data Processing for Multimodal Emotion Recognition

3.3.1 Feature Extraction for EEG

For EEG data preprocessing, the raw EEG signals are down-sampled to data with a 200 Hz sampling rate. We visually check the EEG data and manually remove the data that are severely contaminated by EMG and EOG. In order to filter the noise and artifacts, the EEG data are processed with a bandpass filter between 0.5 Hz and 70 Hz.

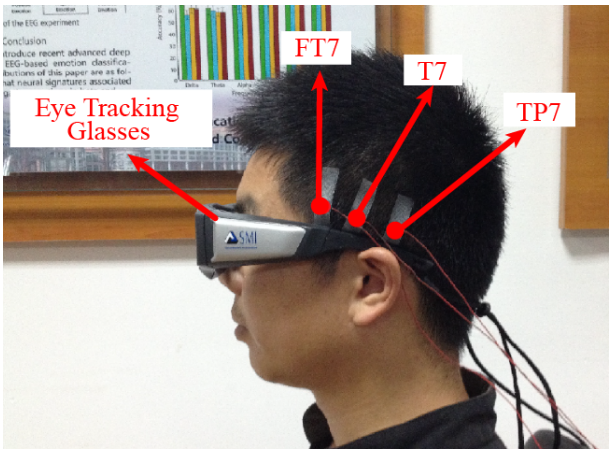


Figure 3–9 The hardware setup. The six symmetrical temporal electrodes above the ears are used for EEG recordings. The eye movement parameters are extracted from the wearable eye-tracking glasses.

Based on our previous studies[188, 189], we extracted six efficient EEG features and electrode combinations from the preprocessed EEG. These features include power spectral density (PSD), differential entropy (DE), differential asymmetry (DASM), rational asymmetry (RASM), asymmetry (ASM) and differential caudality (DCAU) features. We compute the traditional PSD features in the five frequency bands (delta: 1-3 Hz, theta: 4-7 Hz, alpha: 8-13 Hz, beta: 14-30 Hz, gamma: 31-50 Hz) using Short Time Fourier Transform (STFT) with a 1-s-long non-overlapping Hanning window.

The differential entropy feature is firstly proposed by Shi and colleagues[188, 189]. Differential entropy extends the concept of discrete Shannon entropy to measure the complexity of a continuous random variable[190]. Duan *et al.* proposed that differential entropy has the balance ability of discriminating EEG patterns between low and high frequency bands compared with the conventional PSD features. The calculation formula of DE is defined as follows:

$$h(X) = -\int_X f(x)log(f(x))dx, \tag{3–1}$$

where s is a continuous variable and $f(x)$ is the corresponding probabilistic distribution. Following this calculation, it is difficult to compute the DE values if the probabilistic distribution is unknown. However, how to accurately estimate the probabilistic distribution is still an open question. If a random variable obeys the Gaussian distribution $N(\mu, \sigma^2)$, the calculation of

differential entropy can be simplified with the following formulation,

$$\begin{aligned} h(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} \log\frac{1}{\sqrt{2\pi\sigma^2}} \\ \exp\frac{(x-\mu)^2}{2\sigma^2} dx = \frac{1}{2}\log 2\pi e\sigma^2, \end{aligned} \tag{3–2}$$

where x is a variable, and π and e are constants. According to the proof of Shi and colleagues[188], DE feature is equivalent to the logarithmic spectral energy for a fixed-length EEG sequence in a certain frequency band. Therefore, we can calculate the DE features in the five frequency bands with the time complexity $O(KN\log N)$, where K is the electrode number and N is the size of samples. We calculate the DE features in the five frequency bands.

Based on the previous findings about the lateralization between the left and right hemispheres for emotion processing in Chapter 2, we also extract the asymmetry features. We compute the differential asymmetry (DASM) and rational asymmetry (RASM) features as the differences and ratios between the DE features of 27 pairs of hemispheric asymmetry electrodes [1]. The electrode layout of the left and right hemispheres is presented in Figure 3–10. DASM and RASM features are defined as follows, respectively,

$$DASM = DE(X_{left}) - DE(X_{right}) \tag{3–3}$$

and

$$RASM = DE(X_{left})/DE(X_{right}), \tag{3–4}$$

where X_{left} and X_{right} denote the pairs of electrodes on the left and right hemispheres, respectively. The DASM and RASM features are concatenated as the ASM features. Besides the lateralization of emotion processing, previous studies find the spectral differences along the frontal and posterior brain regions. In order to characterize the spectral band asymmetry with respect to caudality (in the frontal-posterior direction), we define the DCAU features as the differences between the DE features of 23 pairs of frontal-posterior electrodes [2]. DCAU can be calculated as

$$DCAU = DE(X_{frontal}) - DE(X_{posterior}), \tag{3–5}$$

[1] Fp1-Fp2, F7-F8, F3-F4, FT7-FT8, FC3-FC4, T7-T8, P7-P8, C3-C4, TP7-TP8, CP3-CP4, P3-P4, O1-O2, AF3-AF4, F5-F6, F7-F8, FC5-FC6, FC1-FC2, C5-C6, C1-C2, CP5-CP6, CP1-CP2, P5-P6, P1-P2, PO7-PO8, PO5-PO6, PO3-PO4, and CB1-CB2

[2] FT7-TP7, FC5-CP5, FC3-CP3, FC1-CP1, FCZ-CPZ, FC2-CP2, FC4-CP4, FC6-CP6, FT8-TP8, F7-P7, F5-P5, F3-P3, F1-P1, FZ-PZ, F2-P2, F4-P4, F6-P6, F8-P8, FP1-O1, FP2-O2, FPZ-OZ, AF3-CB1, and AF4-CB2

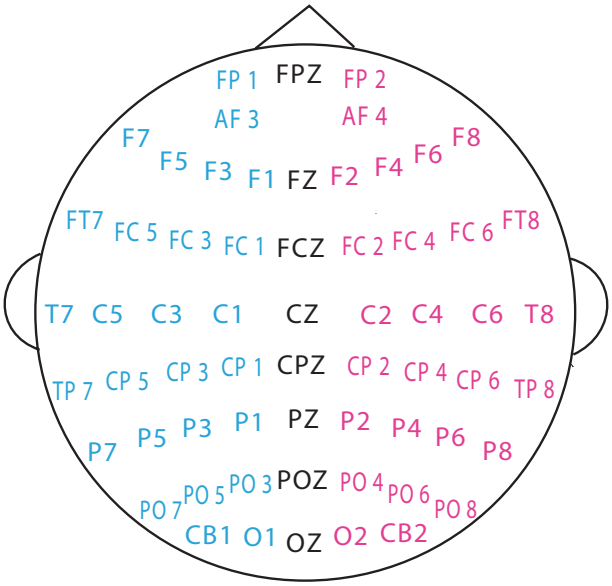


Figure 3–10 The electrode layout of the left and right hemispheres.

where $X_{frontal}$ and $X_{posterior}$ denote the pairs of electrodes on the frontal and posterior areas, respectively. The total dimensions of PSD, DE, DASM, RASM, ASM and DCAU features are 310 (62 electrodes $\times$ 5 bands), 310 (62 electrodes $\times$ 5 bands), 135 (27 electrode pairs $\times$ 5 bands), 135 (27 electrode pairs $\times$ 5 bands), 270 (54 electrode pairs $\times$ 5 bands), and 115 (23 electrode pairs $\times$ 5 bands), respectively.

3.3.2 Feature Smoothing for EEG

Emotion is a complex psychophysiological process with temporal evolution. The dynamic characteristics of emotion changes and how the observed EEG is generated from a hidden emotional states should be considered. Here, we have an assumption that the emotional state is defined in a continuous space and that emotion changes gradually in a short time. Therefore, we introduce the linear dynamic system (LDS) algorithm to filtering the components that are not associated with emotion recognition[191, 192]. We also perform the feature smoothing using the conventional moving average approach for comparison.

The linear dynamic system (LDS) algorithm aims to model the relation between the observations and the hidden states as well as the transitions between different states. A linear dynamic system can be defined as follows,

$$x_t = z_t + w_t \tag{3–6}$$

and

$$z_t = Az_{t-1} + v_t, \tag{3–7}$$

where x_t denotes the observed variables, z_t denotes the hidden emotion variables, A is a transition matrix, w_t is Gaussian noise with mean $\bar{w}$ and variance Q, and v_t is Gaussian noise with mean $\bar{v}$ and variance R. These equations can also be expressed in equivalent form in terms of Gaussian conditional distributions,

$$p(x_t|z_t) = N(x_t|z_t + \bar{w}, Q), \tag{3–8}$$

and

$$p(z_t|z_{t-1}) = N(z_t|Az_{t-1} + \bar{v}, R). \tag{3–9}$$

The initial state is assumed to be

$$p(z_1) = N(z_1|\pi_0, S_0). \tag{3–10}$$

The above model is parameterized by $\theta = \{A, Q, R, \bar{w}, \bar{v}, \pi_0, S_0\}$. θ can be determined using maximum likelihood through the Expectation Maximization algorithm based on the observation sequence x_t. To infer the latent states z_t from the observation sequence x_t, the marginal distribution, $p(z_t|X)$, must be calculated. The latent state can be expressed as

$$z_t = E(z_t|X), \tag{3–11}$$

where E denotes the expectation. This marginal distribution can be achieved by using the message propagation method[193]. We use the cross-validation method to determine the prior parameters in the LDS algorithm. The EEG feature before and after smoothing by the LDS algorithm is shown in Figure 3–11.

3.3.3 Dimensionality Reduction for EEG

The extracted features might contain some uncorrelated components with emotion recognition and cause performance degradation of classifiers. EEG is usually interfered with noise and artifacts when recorded. Additionally, high feature dimensionality may cause the "curse of dimensionality" problem[194]. Therefore, feature selection and feature reduction are usually performed to reduce the computational complexity and improve the robustness of computational models. EEG signals contain meaningful information in terms of critical channels and frequency bands for understanding underlying neural mechanisms beyond cognitive states and behaviors.

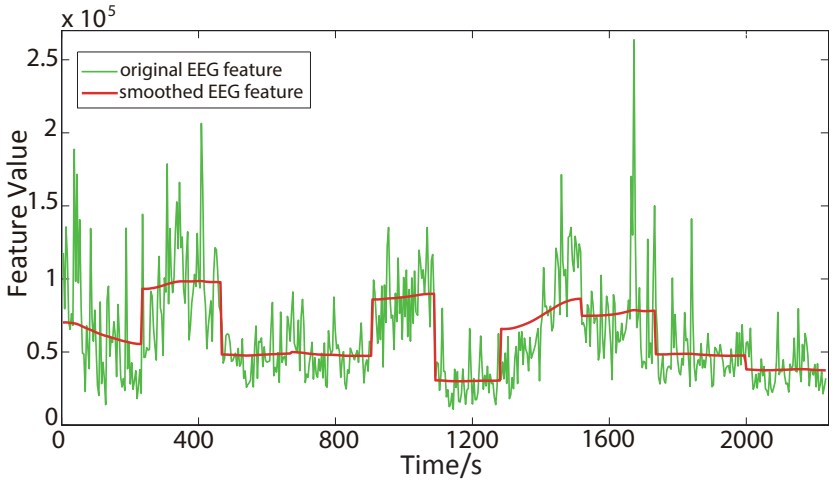


Figure 3–11 The comparison of EEG feature before and after smoothing by the LDS approach. The EEG signal is obtained from the channel FP1 in gamma band of one subject in the SEED experiments.

Feature selection and feature reduction for EEG signals require domain knowledge. However, conventional feature reduction methods, e.g., principal component analysis (PCA) and Fisher projection do not preserve these physiological information after transformation.

Here, we apply the minimal redundancy maximal relevance (MRMR) algorithm to select a critical feature subset from the initial feature set[195]. The MRMR algorithm uses mutual information as the relevance measure with the max-dependency criterion and minimal redundancy criterion. Max-Relevance searches for features satisfying Eq. 3–12 with the mean value of all the mutual information values between the individual feature x_d and class c as follows,

$$\max D(S,c), D = \frac{1}{|S|}\sum_{x_d \in S} I(x_d;c), \tag{3–12}$$

where S represents the feature subset to select. When two features highly depend on each other, the respective class-discriminative power would not change much if one of them is removed. Therefore, the following minimal redundancy condition can be added to select for mutually exclusive features,

$$\min R(S), R = \frac{1}{|S|^2}\sum_{x_{di},x_{dj} \in S} I(x_{di},x_{dj}). \tag{3–13}$$

Combining with the above two constraints, the criterion of the minimal-redundancy-maximal-relevance can be defined as follow,

$$\max \varphi(D,R), \tag{3–14}$$

where $\varphi = D - R$. In practice, an incremental search method is used to find the near-optimal K features.

3.3.4 Feature Extraction for Eye Movements

Eye tracking provides various eye movement parameters, e.g., pupil diameters, fixation details, saccade details, blink details, and event statistics. Previous studies support that pupil diameters are associated with emotion processing[150, 196]. However, the changes of pupil diameter can be highly influenced by the environmental luminance. It is essential to remove the light reflex components and obtain pupil information that are solely associated with emotion. Based on the assumption that the pupil changes of different participants in response to the same stimuli are similar, we build a light reflex model to approximately remove luminance influences using the principal component analysis (PCA) algorithm[98].

Suppose that $\mathbf{Y}$ is an $M \times N$ matrix that represents pupil diameters to the same video clip from N subjects and M samples. Then, $\mathbf{Y} = \mathbf{A} + \mathbf{B} + \mathbf{C}$, where $\mathbf{A}$ is luminance influences that are prominent, $\mathbf{B}$ is emotional influences that we want, and $\mathbf{C}$ is noise. We used PCA to decompose $\mathbf{Y}$ and computed the first principle component as the estimate of the light reflex. Let $\mathbf{Y}_{rest}$ be the emotion-relevant pupil response. We define $\mathbf{Y}_{rest} = \mathbf{Y} - \mathbf{Y}_1$. After subtracting the first principle component, the residual part contains the pupil response that is only associated with emotions. Figure 3–12 illustrates the results of removing the luminance influences using PCA. The eye movement data are further re-sampled to be aligned with the EEG modality.

After preprocessing the eye movement data, we extract various features from different parameters, such as pupil diameter, fixation, saccade, and blink. We extracted the PSD and DE features for the pupil diameter in X and Y axes in the four frequency bands (0-0.2 Hz, 0.2-0.4 Hz, 0.4-0.6 Hz, and 0.6-1 Hz) based on the previous study[98]. We also extracted the statistical parameters such as mean and standard deviation for pupil diameters, dispersion, fixation duration, blink duration and saccade. Nine event statistics such as blink frequency and saccade frequency are extracted for each trial. The details of the extracted eye movement features are shown in Table 3–2. The total number of dimensions of the eye movement features is 33. Figure 3–13 illustrates five eye movement parameters.

[1]The average saccade latency is the average value of the next saccade start time minus the last saccade end time.

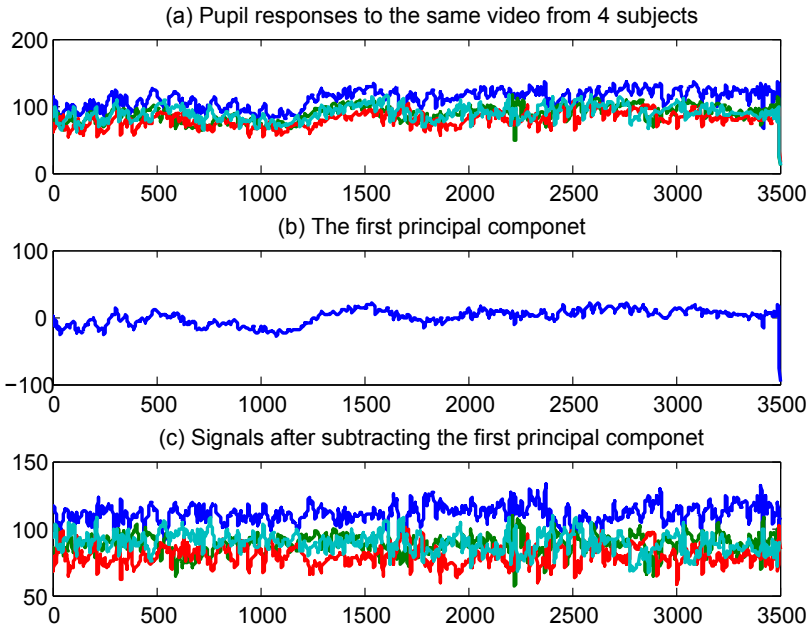


Figure 3–12 Removing the luminance influences of pupil responses using PCA: (a) the raw pupil diameters to the same video clip from four subjects, (b) the first principle component as the estimate of the light reflex, and (c) pupil response associated with emotions only after subtracting the first principle component.

3.4 Multimodal Vigilance Estimation Dataset (SEED-VIG)

3.4.1 Experimental setup

Besides multimodal emotion recognition dataset, we develop another multimodal vigilance estimation dataset (SEED-VIG) for driving fatigue detection. We built up a simulated driving system to collect EEG and EOG under different vigilance states in the laboratory. We designed the virtual-reality-based simulated driving scenes, including a four-lane highway road and various weather. The driving scenes were shown on a large LCD screen in front of a real vehicle. The simulated driving experiments were performed in the real vehicle without unnecessary engine and other components. The subjects were asked to drive the car using the steering wheel and gas pedal. The driving scenes were synchronously updated according to subjects' operations. The road was primarily designed straight and monotonous to induce fatigue more easily. There was no warning feedback to subjects throughout the whole experiments. The simulated driving system and the experimental scene were shown in Figure 3–14.

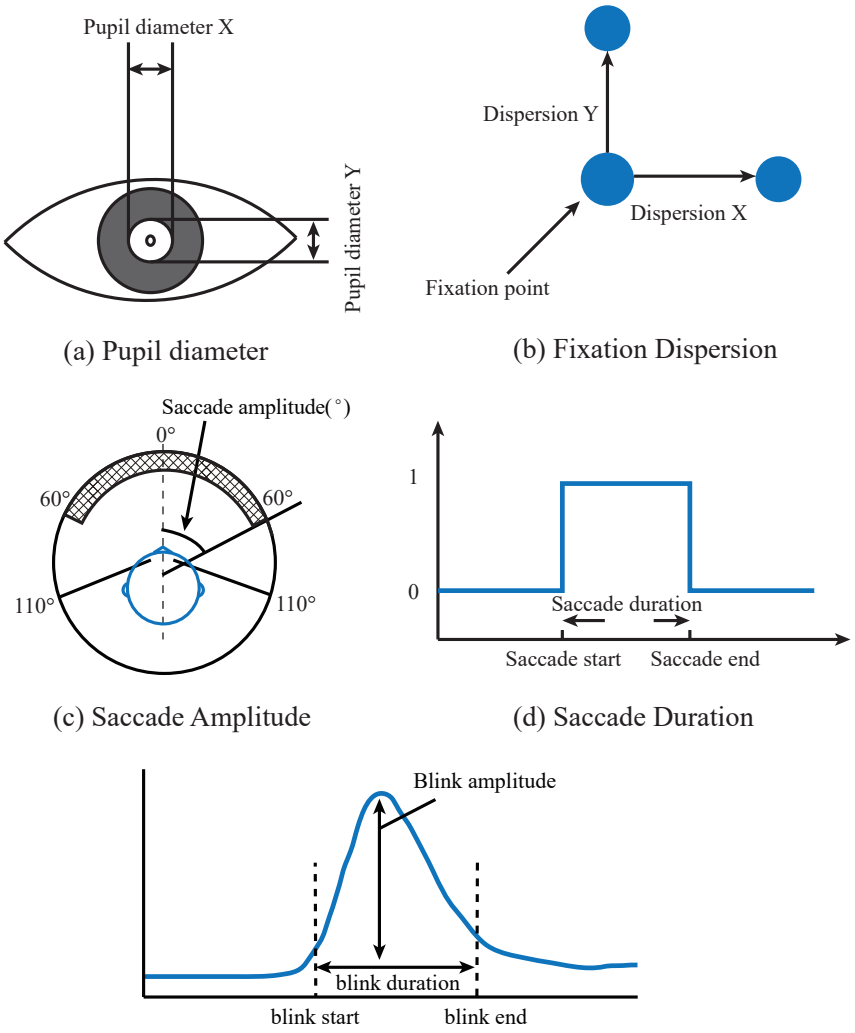
There were totally 23 subjects (mean age: 23.3 years, STD: 1.4 years, 12 females and 11

Table 3–2 The details of the extracted eye movement features.

Eye movement parameters	**Extracted features**
Pupil diameter (X and Y)	Mean, standard deviation and DE features in four bands: 0-0.2 Hz, 0.2-0.4 Hz, 0.4-0.6 Hz, and 0.6-1 Hz
Dispersion (X and Y)	Mean, standard deviation
Fixation duration (ms)	Mean, standard deviation
Blink duration (ms)	Mean, standard deviation
Saccade	Mean, standard deviation of saccade duration (ms) and saccade amplitude (°)
Event statistics	Blink frequency, fixation frequency, maximum fixation duration, total fixation dispersion, maximum fixation dispersion, saccade frequency, average saccade duration, average saccade amplitude, and average saccade latency[1].

males) participating in the experiments. Caffeine, tobacco, and alcohol were prohibited prior to participating in the experiments. All participants had normal or corrected-to-normal vision. Before the experiments, we had a pre-test for the participants to understand the instructions and ba familiar with the environments. In order to induce fatigue easily, we performed most fatigue driving experiments in the early afternoon after lunch (approximately 13:30) when the circadian rhythm of sleepiness reached its peak[197]. The whole experiment last for about 2 hours. The participants were observed to be more and more tired and drowsy with continuous driving in the simulated environments.

For data recording, we collected both EEG and forehead EOG signals simultaneously using the Neuroscan EEG system with a 1000 Hz sampling rate. For EOG setup, in our previous study, we designed a new forehead electrode placements for EOG recordings and showed the superiority of this EOG setup compared with the conventional EOG setup[198, 199]. The electrode placements of the forehead EOG and the conventional EOG are presented in Figure 3–15. For



(a) Pupil diameter (b) Fixation Dispersion

(c) Saccade Amplitude (d) Saccade Duration

(e) Blink amplitude

Figure 3–13 Illustration of various eye movement parameters: pupil diameter, fixation dispersion, saccade amplitude, saccade duration, and blink.

the EEG setup, we recorded 12-channel EEG signals from the posterior site [1] and 6-channel EEG signals from the temporal site[2] using the international 10-20 system as shown in Figure 3–16. For vigilance annotations, both eye movements and facial video were simultaneously recorded during the whole experiments using SMI ETG eye tracking glasses and a video camera in front of the participants, respectively. To enhance the cooperation in the related research fields, the developed SJTU multimodal vigilance estimation dataset (called SEED-VIG) can also be freely

[1] CP1, CPZ, CP2, P1, PZ, P2, PO3, POZ, PO4, O1, OZ, and O2

[2] FT7, FT8, T7, T8, TP7, and TP8

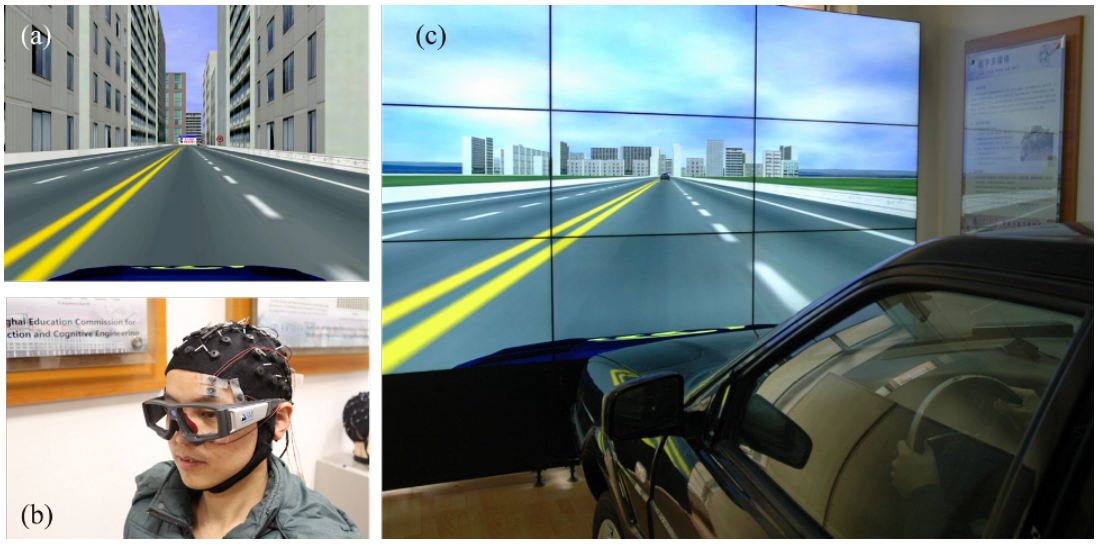
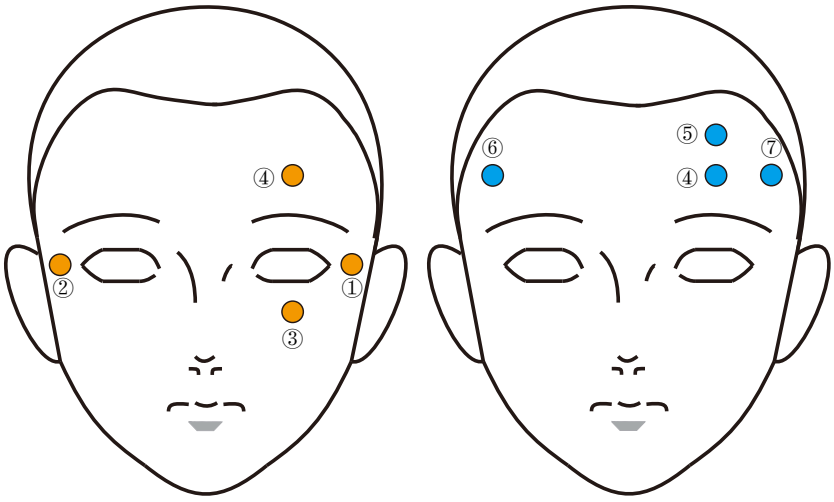


Figure 3–14 The simulated driving system and the experimental scene for multimodal vigilance estimation. (a) The simulated driving scenes, (b) One participant with EEG, forehead EOG, and eye tracking setups in the experiments, and (c) The simulated environment with a large LCD screen and a vehicle without without unnecessary engine and other components.



(a) Traditional EOG (b) Forehead EOG

Figure 3–15 The electrode placements for the traditional and forehead EOG setups in vigilance estimation experiments. Electrode four is the shared electrode of both setups.

downloaded for academic researchers via the project website of SEED.

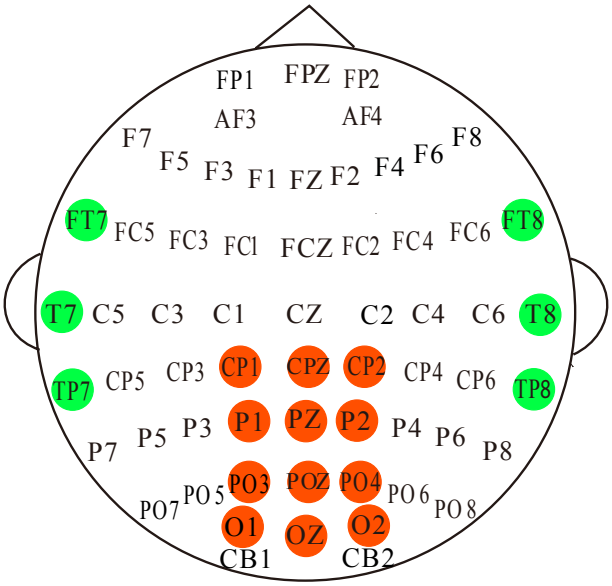


Figure 3–16 The electrode placements for the EEG setups in vigilance estimation experiments. Here, 12-channel and 6-channel EEG signals were recorded from the posterior site (red colour) and temporal site (green colour), respectively.

3.4.2 Vigilance Annotation

The vigilance fluctuations are the convert cognitive states, it is extremely technically difficult to obtain the ground-truth labels of various vigilance states, especially in continuous scales. Therefore, the primary challenge of vigilance estimation is how to quantitatively annotate the recorded sensor data for leveraging the superiority of supervised machine learning paradigms. Various vigilance annotation approaches have been proposed and evaluated in the literature, e.g., lane departure and local error rates[200, 201]. Lin and colleagues designed an event-related lane-departure experiment for driving fatigue detection[168, 202]. They used the response time of drivers to the random drifts of cars as the vigilance index. Shi and Lu used the local error rate of the subjects as the vigilance annotation, where the subjects were asked to press the corresponding buttons when seeing traffic signs with different colors[164]. These vigilance annotation methods are mainly based on the subjects' cognitive performance and can reflect the vigilance fluctuations in the simulated driving environments to some extent. However, these methods usually involve an intrusive stimulus or behavioral probe, which periodically interrupts subjects. Moreover, they are not appropriate for vigilance estimation in real-world driving scenarios, since dual tasks distract drivers attention and affect driving safety. It is attractive to develop an efficient nonintrusive vigilance annotation method.

There is another popular image-based vigilance estimation method called PERCLOS index (the percentage of eye closure) from the US Department of Transportation[161], which is widely used in the literature[160, 203, 204]. Spontaneous eyelid closures have been found to be an efficient indicator for vigilance states. PERCLOS methods use facial videos from cameras in the front to detect eyelid states. The performance of these computer vision systems can be highly degraded due to illumination changes and heavy occlusion. Therefore, some studies utilized near-infrared and depth images to reduce the influence of environmental changes[205, 206]. In this study, we use eye tracking glasses instead of facial videos to compute PERCLOS index for vigilance annotations[207]. The proposed vigilance annotation method can be used in both laboratory and real-world experiments.

In comparison with facial videos, eye tracking provides accurate and detailed various eye movements, such as blink, saccade, and fixation. The eye tracking-based PERCLOS index can be calculated from the percentage of the durations of blinks and 'CLOS' over a specified time interval as follows:

$$PERCLOS = \frac{blink + CLOS}{interval}, \tag{3–15}$$

and

$$interval = blink + fixation + saccade + CLOS, \tag{3–16}$$

where 'CLOS' denotes the duration of the eye closures.

To evaluate the efficiency of the eye tracking-based method, we compare the calculated PERCLOS index and the corresponding cognitive states with facial videos and find high correlation between them[207]. Eye tracking allows this kind of vigilance annotation to be conducted in a very natural way so that we can observe subjects' behaviors uninterruptedly. In contrast, other vigilance annotations with dual tasks can cause safety issues[208]. The proposed method can also be performed automatically without the limitations of human subjective evaluations.

3.5 Wearable Device for Vigilance Estimation

In this section, we will introduce our designed wearable EOG device with flexible dry electrodes and acquisition board [1]. The experimental setups for both laboratory simulations and real-world environments are also presented.

[1] It should be noted that the contributions of the designed dry conductive fabric electrodes and the designed EOG acquisition board are mainly from the project cooperations with the research groups of Prof. Jing-Quan Liu and Prof. Guoxing Wang, respectively, from the Department of MicroNano Electronics, Shanghai Jiao Tong University, China.

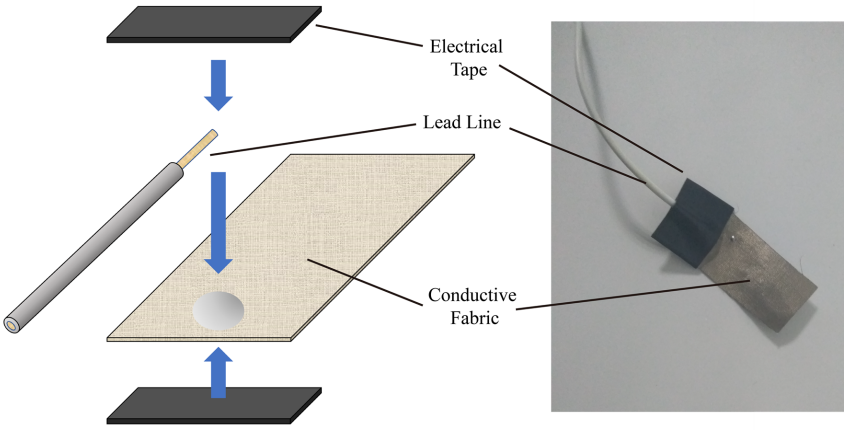


Figure 3–17 The designed dry fabric-based electrode.

3.5.1 Flexible Dry Electrodes

Wet electrodes are commonly used in EEG and EOG recordings to ensure signal quality. However, the usage of conductive gel has many limitations. We developed a novel dry fabric-based electrode for forehead EOG recording on bare skin. The dry electrode was designed using conductive fabric, which is commonly used to produce protective gear for gravida.

The designed dry fabric-based electrode is presented in Figure 3–17. Many previous studies utilize conductive polymers[209–211] and metals[212, 213] to design the dry electrodes, which makes discomfort with intense setups. For conductive polymers, they are typically not soft enough to adapt to the shrinkage on the surface of skin, and the conductivity is considerably lower than that of metals. However, metal-based dry electrodes cause discomfort due to the rigid property. In contrast, we propose to use the conductive soft material with a type of fabric woven using silver-coated nylon line. In comparison with conductive polymers, the conductive fabric has a considerably higher impedance conductivity rate (conductivity of approximately 0.07 Ohm/square) than conductive polymer materials. Moreover, it is soft enough to design flexible prototypes.

The fabric was firstly cut into pieces with 17 mm long and 10 mm wide and then pasted onto the surfaces of wet wipes. The lead lines were then welded onto the conductive fabric pieces. During welding, the wet wipes can prevent the conductive fabric from burnout. After welding, the contact of the conductive fabric and lead line was reinforced using electrical tape. The connections between the lead wire and the conductive fabric can be further processed into metal button shapes for long-term usage. In this way, the connections are strong enough and

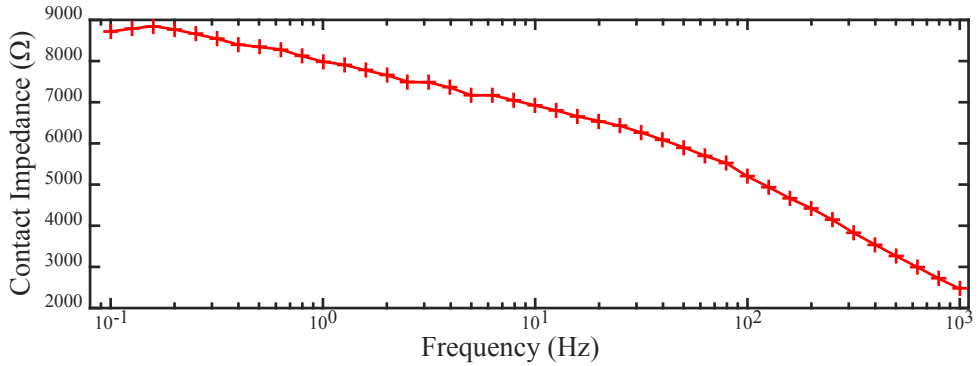


Figure 3–18 The contact impedance of different frequencies for the designed conductive fabric electrodes.

feasible for electrode replacement.

In this study, the textile surface of the conductive fabric enables the electrode to be pasted onto the surface of skin with little clearance. The textile can compensate for the shrinkage on skin and lead to a lower contact impedance. Furthermore, sweat can infiltrate into the conductive fabric and keep the surface of the skin moist, which leads to a reduction in the electrode-skin contact impedance and enables the electrode to record EOG signals similar to wet electrodes.

To compare the performance of the conductive fabric electrodes with wet electrodes, we measured the contact impedance using an electrochemical workstation (660C, Chiinstru-ments Ltd., China). The contact impedance of different frequencies is presented in Figure 3–18. From the result, the contact impedance of the dry conductive fabric electrodes is similar to that of traditional wet electrodes. Figure 3–19 shows the raw signals recorded simultaneously by using the dry fabric electrode and the commercial wet electrode on the site $FP1$ with the Neuroscan recording system. We can see that the quality of the dry electrode is comparable with that of wet electrode.

We performed experiments for measuring forehead EEG alpha rhythm to evaluate the performance of the dry conductive fabric electrodes. The alpha rhythm of EEG components is dominant during eye closure period, which is an efficient approach to evaluating the quality of EEG recordings. Figures 3–20 (a) and (b) present the forehead EEG signals and the corresponding power spectral density recorded using the conductive fabric dry electrodes, respectively. A peak corresponding to alpha rhythm can be observed between 8 and 11 Hz in Figure 3–20(b). The experimental results demonstrate the efficiency of the designed dry electrodes for EEG and EOG recordings.

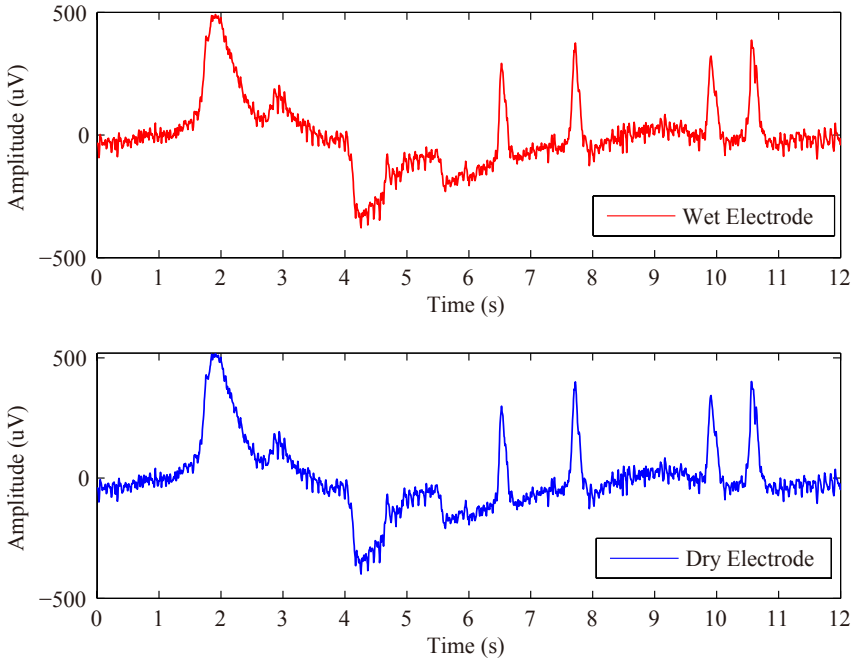
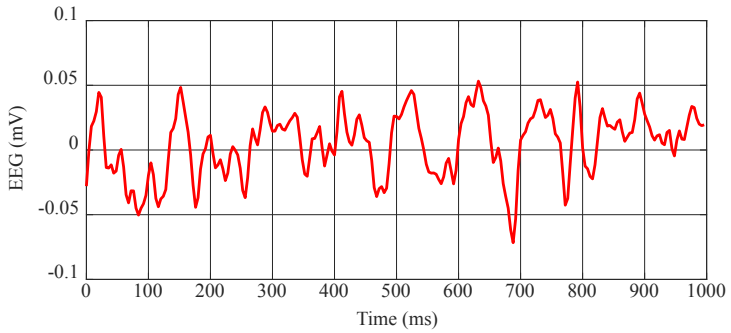


Figure 3–19 The raw signals recorded simultaneously by using the commercial wet electrode (upper) and the dry fabric electrode (lower) on the forehead site $FP1$.

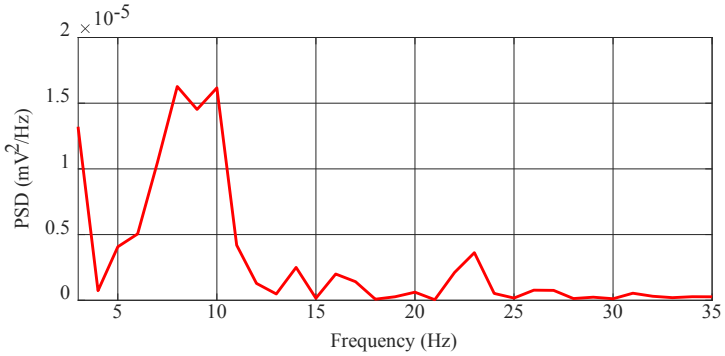
3.5.2 EOG Acquisition Board

The designed EOG acquisition board is composed of one TI ADS1298 analog front end for recording EOGs, a Nordic nRF51822 microcontroller and a Bluetooth module (BC6140, classical Bluetooth) for wireless transmission of the data to a PC or any mobile device. A gain of 12 was used on the ADS1298 differential amplifiers. The built-in 24-bit ADC had a resolution of approximately 0.4 uV. The board was battery powered, and the battery can be recharged using a micro-USB interface. The acquisition board was capable of simultaneously recording 8-channel EOG signals (4 channels are used in this study) and was approximately 4.5×6.5 cm in size. The sampling frequency was set to 250 Hz. Figure 3–21 shows the designed EOG acquisition board and its paired tablet user interface.

The "lead-off" detection of the TI ADS1298 was used to measure the electrode-scalp impedance when preparing the electrode setups. A 24 nA sinusoidal AC current at a known frequency of 30.5 Hz was injected for each electrode. The lead-off statuses of individual electrodes were denoted with a blue LED on the board. In this way, a low-impedance conductive path between drivers and our acquisition board can be guaranteed for high quality EOG recordings. Table 3–3 shows the summary of main components and technical parameters and Figure



(a)



(b)

Figure 3–20 EEG signals recorded by using the dry conductive fabric electrodes: (a) EEG signals recorded with eyes closed and (b) PSD of the recorded EEG signals.

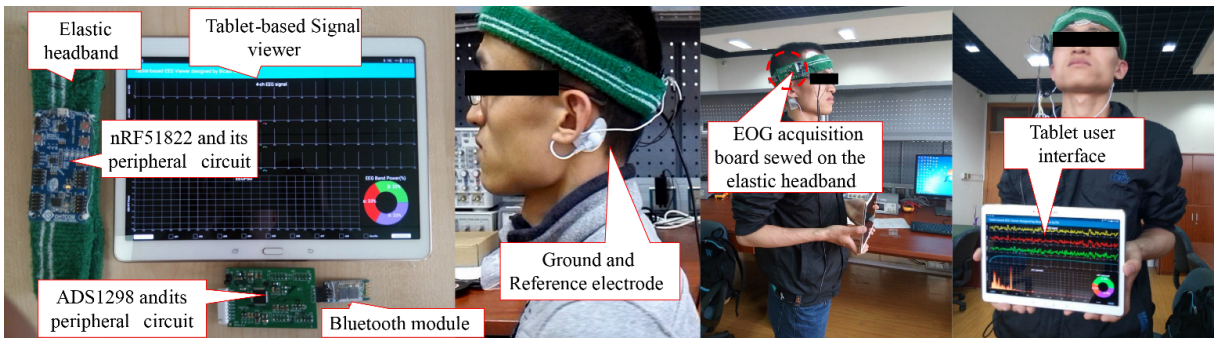


Figure 3–21 The designed EOG acquisition board and its paired tablet user interface.

3–22 presents the system diagram and signal flow of the designed EOG acquisition board. The size of the EOG acquisition board allows the entire board to be embedded in a a light-weight elastic headband for real-world applications.

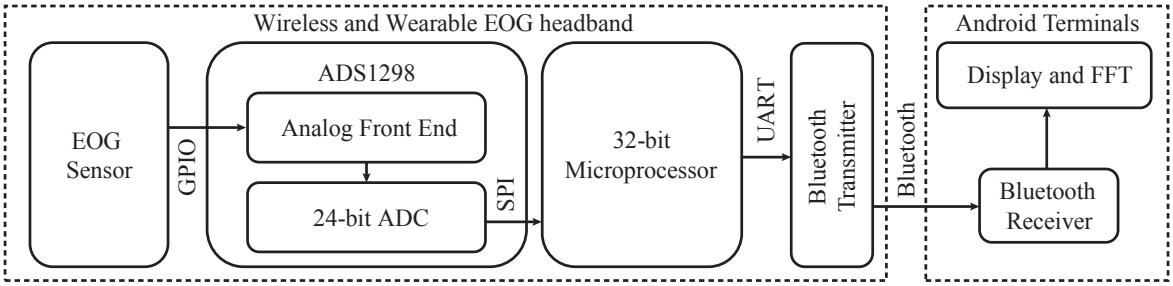


Figure 3–22 System diagram and signal flow of the designed EOG acquisition board.

Table 3–3 Summary of Main Components and Technical Parameters of the Proposed Forehead EOG Acquisition System.

Sub-system	Component	Parameter
EOG	Sensors	Ag/AgCl
	Sampling rate	250 Hz
	Bandpass Filter	4-30 Hz
	Gain	12X
	Montage	Frontal unipolar four-channel
	Frontal End	ADS1298 (built-in 24-bit ADC)
	MCU	Nordic nRF51822
Bluetooth	Protocol	Bluetooth EDR
	Baud-rate	57600 bps
Terminals	Operating System	Android 4.0+
	Processor	1.2 GHz+

3.5.3 Laboratory Driving Simulations

The laboratory driving simulations of the wearable device for vigilance estimation were similar to those of the wet electrodes. The subjects were asked to sit inside a real vehicle without engine with the simulating environments showing on the front large LCD screen. The simulated driving environment was programmed with OGRE3D graphics rendering engine and OpenAL 3D audio API. The scenarios included a four-lane highway, various cars, buses, traffic signs, buildings, and tunnels. To enhance virtual reality, the highway scenes were simultaneously updated according

to subjects' operations using a Logitech steering wheel controller that consists of a steering wheel and a gas pedal.

Fourteen subjects (mean age: 21.5 years, STD: 1.0 year, 1 female and 13 males) participated in the laboratory simulated driving experiments. The setups were similar to SEED-VIG in Section 3.4. The designed routes were primarily straight and monotonous to induce fatigue more easily. The experiments were performed in the early afternoon (approximately 13:30) after lunch and at nightfall (approximately 19:00) after dinner according to the circadian rhythm of sleepiness. The whole data recording process last for about 2 hours.

The forehead EOG and eye movements were simultaneously recorded using our designed wearable device with dry electrodes and the SMI ETG eye tracking glasses, respectively. As feedbacks of the experiments, the facial videos were also recorded with a camera mounted in front of the subjects. The vigilance annotations were the same with that in Section 3.4.

3.5.4 Real-World Driving Experiments

In this section, we introduce the real-world driving experiments to evaluate the performance of the vigilance estimation system. Ten subjects (mean age: 24.2 years, STD: 2.7 years, 1 female and 9 males) participated in the real-world driving experiments. These subjects were all different from those of the laboratory driving simulations. To ensure safety when the subjects were drowsy, the subjects were asked to sit in the front passenger seat beside the drivers during the experiments. The driving route was inside Shanghai Jiao Tong University, Minhang Campus, in Shanghai, China, as shown in Figure 3–23. The route was planned to take subjects through situations where different vigilance levels were likely to occur. The route contained crowded areas with many pedestrians and sparse areas with monotonous stretches. At the beginning of the experiment, both drivers and subjects were shown a map of the driving route to keep the drives consistent. Instructions explaining the complete experimental procedure were given to each subject. The experiments were performed in electric vehicles. One lap of the route was approximately 5.5 km. The driving speed limit was kept at approximately 30 km/h, and the duration of the driving was approximately 1.5 hours with several laps.

The experimental setups and time were similar to the laboratory driving simulations. The forehead EOG, eye tracking and facial videos were recorded during the whole experiments. During the experiments, an observer accompanied the drivers in the car to monitor the signal recording. The observer sat in the rear seat behind the subjects to avoid interfering with the drivers and the subjects. Figure 3–24 shows the sample frames recorded from the scene camera

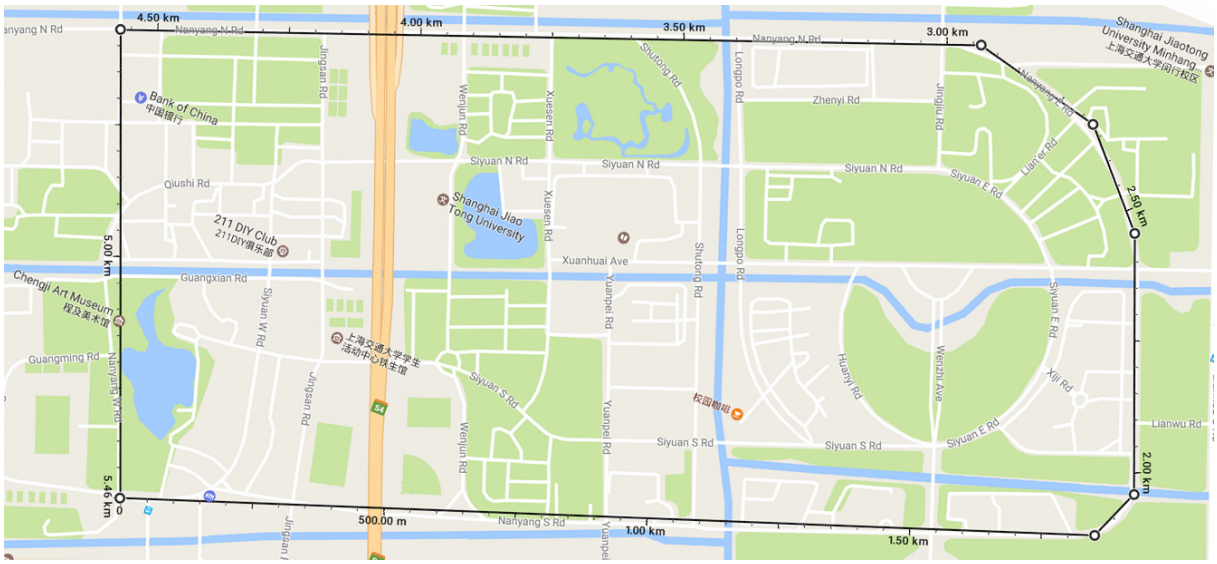


Figure 3–23 The map of the real-world driving experiments.

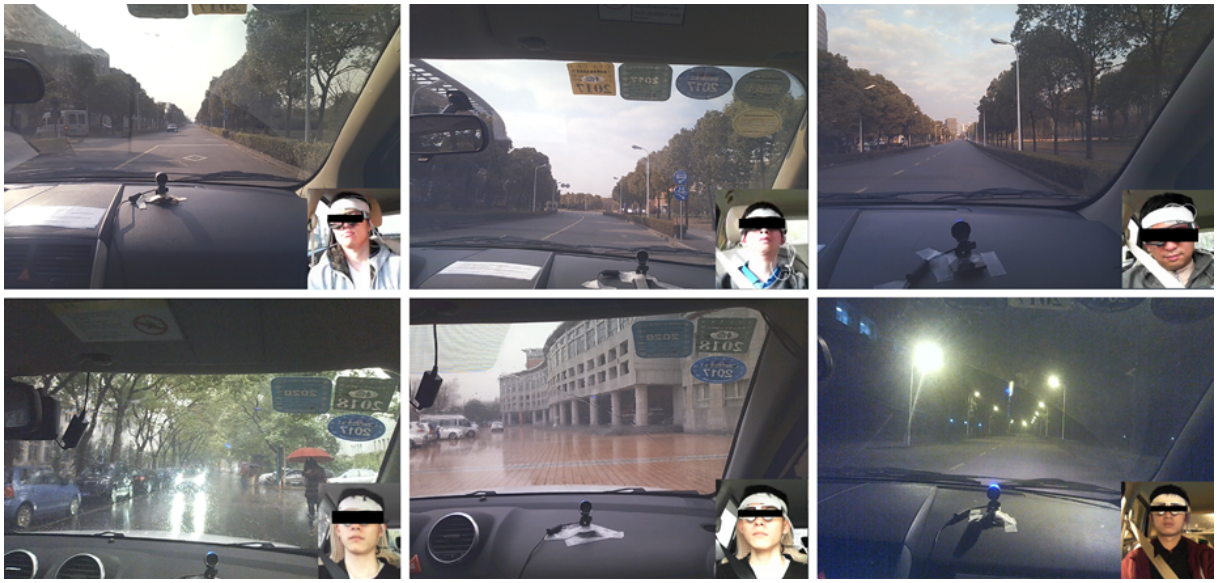


Figure 3–24 The sample frames recorded from the scene camera of eye-tracking glasses as the first perspective. The real-world driving experiments were performed under various illumination and weather conditions, including sunny, cloudy, windy, rainy and night time.

of eye-tracking glasses as the first perspective. The vigilance annotations were the same with that in Section 3.4.

In order to evaluate the robustness of the system, we performed the real-world driving experiments under various illumination and weather conditions, including sunny, cloudy, windy, rainy and night time. Some conditions are presented in Figure 3–24. The weather and illumination conditions of the ten experiments were rainy, windy, cloudy, rainy, sunny, sunny, cloudy, cloudy, sunny, and night time, respectively. Some severe illumination changes and blurred vision can highly influence the performance of computer vision based systems. In contrast, the quality of EOG signals was not sensitive to these factors.

3.6 Summary

The Chapter presents the relevant detailed experimental setups of emotion recognition and vigilance estimation. We firstly developed two multimodal emotion dataset for three and four emotions (SEED and SEED-IV), respectively. SEED was released in April, 2015 and has received more than 310 applications until June, 2018. SEED-VIG was released in April, 2017 and has received more than 30 applications until June, 2018. Then the data processing for multimodal emotion recognition was described, including feature extraction and feature smoothing for EEG and eye movements. The further analysis and model training in the rest of the thesis are mainly based on these EEG and eye movement features. Therefore, we combined these common descriptions together in the Chapter. Another multimodal vigilance estimation dataset

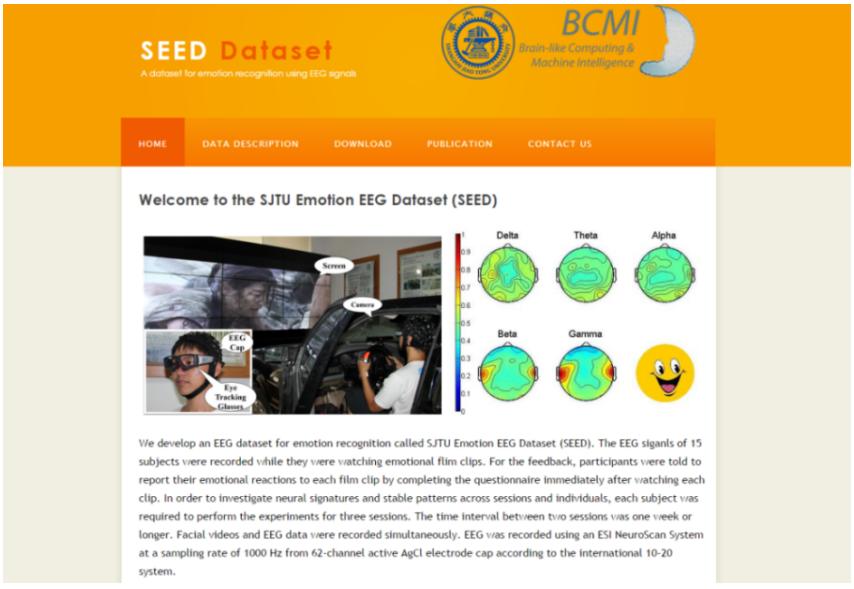


Figure 3–25 The snapshot of the project website SJTU Emotion EEG Dataset (SEED).

called SEED-VIG was introduced with EEG and EOG data. To enhance the wearability and feasibility of vigilance estimation system in real-world scenarios, we further develop a wearable device with our designed flexible dry electrodes and an acquisition board for recording forehead EOG signals. We performed both laboratory driving simulations and real-world driving experiments to evaluate the performance of the proposed system. All these datasets are freely available to the research community via the project website SJTU Emotion EEG Dataset (SEED) with the link: `http://bcmi.sjtu.edu.cn/%7Eseed/`.

Chapter 4

EEG-based Emotion Recognition

In this Chapter, we present EEG-based emotion recognition. Section 4.1 introduces the application of deep neural networks to EEG-based emotion recognition. Deep learning improves the classification performance in comparison with the conventional shallow models. Moreover, we investigate the critical frequency bands and channels with the weights of the trained DNNs and evaluate the performance of different electrode reduction profiles based on the observations. In Section 4.2, we aim to identify stable EEG patterns over time for emotion recognition. We systematically evaluate the performance of various feature extraction, feature smoothing, feature reduction, and classification methods on two public EEG datasets: DEAP and SEED. The neural patterns of happy, sad, and neutral emotions are identified. The model stability over time is evaluated across sessions. The work in this Chapter has been published at IEEE Transactions on Affective Computing[184] and IEEE Transactions on Autonomous Mental Development[36].

4.1 EEG-based Emotion Classification Using Deep Neural Networks

4.1.1 Introduction

Since 2006, deep learning has emerged and achieved the state-of-the-arts in various research areas, including computer vision, speech recognition, natural language processing etc. Various deep architectures have been proposed, such as deep autoencoder, convolution neural networks and deep belief networks. In recent years, deep learning has been successful applied to physiological signal processing. Martinez *et al.* presented the superiority of deep convolution neural networks for classifying four cognitive states using skin conductance and blood volume pulse signals[214]. The four affective states were relaxation, anxiety, excitement and fun. Their experimental results showed the proposed deep models performed better than the conventional feature extraction and feature selection methods. Martin and colleagues combined deep belief networks and hidden Markov models for sleep stage classification with multimodal clinical sleep datasets and achieved comparable performance with the hand-crafted features[215]. Li *et al.* proposed to use deep belief networks to tackle the challenges of small sample problem and

irrelevant features in EEG-based affect classification[216]. Schirrmeister *et al.* trained convolutional neural networks (ConvNets) for end-to-end EEG decoding and visualized the learned EEG features from the networks[217]. Bashivan *et al.* proposed deep recurrent-convolutional neural networks to learn efficient EEG representations with the spatial, spectral, and temporal information[218]. They achieved the state-of-the-art on the mental load classification task. Biswal and colleagues proposed SLEEPNET with deep recurrent neural networks for automatic sleep stage classification on the largest sleep physiology database and achieved human-level annotation performance[84]. Li et al. proposed to combine Gaussian process adapter and CNNs to deal with insufficient labeled data in EEG analysis[219]. The proposed approach achieved competitive performance on several datasets while yielding interpretable features.

Electrode reduction has been studied to filter the irrelevant features and reduce the calibration cost. However, most approaches to selecting optimal electrode subsets are based on statistical parameters such as correlation coefficient, F-score, and accuracy[110, 118]. Li *et al.* proposed to select critical channels for affective state classification using DBNs based on the observation that irrelevant channels randomly updated the DBN weights[216].

In this Section, we aim to investigate the critical frequency bands and channels using deep neural networks. We show the superior performance of deep models over shallow models. By analyzing the weight distributions from the trained neural networks instead of statistical parameters, we investigate different electrode set reductions and define the optimal electrode placement which outperforms original full pool of electrodes. We reveal the neural patterns in terms of critical frequency bands and brain regions for happy, sad, and neutral emotions.

4.1.2 Deep Belief Networks

Deep belief network (DBN) is a probabilistic generative model with deep architecture, which was firstly proposed by Hinton *et al.* in 2006[220] as a major milestone of nowadays deep learning era[2, 221]. DBN has multiple neural layers and each layer is a restricted Boltzmann machine (RBM) with visible units and hidden units. DBN is constructed by stacking a predefined number of RBMs on top of each other, where the output of a lower-layer RBM is the input of a higher-layer RBM. The structure of DBN is illustrated in Figure 4–1. An efficient greedy layer-wise algorithm is used to pre-train each layer of networks.

Restricted Boltzmann Machine (RBM), originally known as Harmonium has no connections either between visible units or between hidden units. The visible and hidden units have a bias vector, c and b, respectively. The joint distribution $P(v, h; \theta)$ over the visible units v and hidden

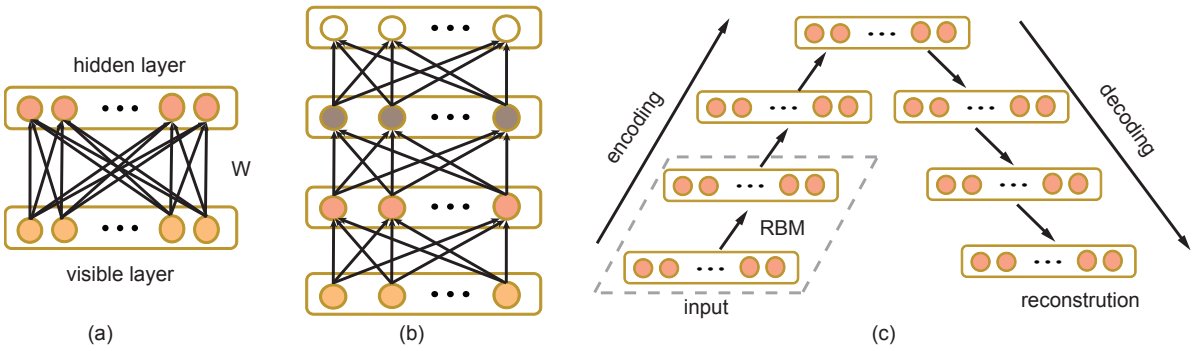


Figure 4–1 (a) A RBM contains the hidden layer neurons connected to the visible layer neurons with weights W. (b) A DBN using supervised fine-tuning of all layers with backpropagation. (c) The graphical depiction of unrolled DBN using unsupervised fine-tuning of all layers with backpropagation.

units h, given the model parameters θ, is defined in terms of an energy function $E(v,h;\theta)$ as

$$P(v,h;\theta) = \frac{exp(-E(v,h;\theta))}{Z}, \tag{4–1}$$

where $Z = \sum_v \sum_h exp(-E(v,h;\theta))$ is a normalization factor, and the marginal probability that the model assigns to a visible vector v is

$$P(v;\theta) = \frac{\sum_h exp(-E(v,h;\theta))}{Z}. \tag{4–2}$$

For a Gaussian (visible)-Bernoulli (hidden) RBM, the energy function is defined as

$$E(v,h;\theta) = -\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}v_i h_j - \frac{1}{2}\sum_{i=1}^{I}(v_i - b_i)^2 - \sum_{j=1}^{J} a_j h_j, \tag{4–3}$$

where w_{ij} is the symmetric interaction term between visible unit v_i and hidden unit h_j, b_i and a_j are the bias term, and I and J are the numbers of visible and hidden units. The conditional probabilities can be efficiently calculated as

$$P(h_j = 1|v;\theta) = \sigma(\sum_{i=1}^{I} w_{ij}v_i + a_j), \tag{4–4}$$

$$P(v_i = 1|h;\theta) = N(\sum_{j=1}^{J} w_{ij}h_j + b_i, 1), \tag{4–5}$$

where $\sigma(x) = 1/(1+exp(x))$, and v_i takes real values and follows a Gaussian distribution with mean $\sum_{j=1}^{J} w_{ij}h_j + b_i$ and variance one.

Taking the gradient of the log likelihood $\log p(v;\theta)$, we can derive the update rule for adjusting RBM weights as

$$\Delta w_{ij} = E_{data}(v_i h_j) - E_{model}(v_i h_j), \tag{4–6}$$

where $E_{data}(v_i h_j)$ denotes the expectation of data distribution and $E_{model}(v_i h_j)$ denotes the expectation of model distribution. But $E_{model}(v_i h_j)$ is intractable to compute since it requires infinitely sampling steps to approximate the model distribution.

A method called Contrastive Divergence (CD) is proposed to overcome this issue. According to the CD approach, a finite k steps of sampling is enough to converge to the model distribution. And in practise, $k = 1$ is usually sufficient for the training. In the CD method, $E_{model}(v_i h_j)$ is replaced by running the Gibbs sampler initialized at the data for one full step. Sometimes momentum in weight update is used for preventing getting stuck in local minima and regularization prevents the weights from getting too large[222]. RBM has become one of the popular fundamental building blocks of deep neural networks with its powerful approximation property.

As shown in Figure 4–1, there are three steps for the DBN training: 1) unsupervised pre-training of each layer, 2) unsupervised fine-tuning of all layers with backpropagation, and 3) supervised fine-tuning of all layers with backpropagation. Layerwise pre-training trains the network bottom up with RBM training for each layer. Unsupervised fine-tuning unrolls n RBMs to form a $2n-1$ directed encoder and decoder network. The aim is to minimize the error between the input and the reconstruction. An additional label layer is added to the top of the pre-trained DBN and the weights are updated using error backpropagation with labeled data.

4.1.3 Classifier Training

In this study, we compared the performance of DBN with several conventional shallow models for EEG-based emotion recognition, including K nearest neighbor (kNN), logistic regression (LR), and support vector machine (SVM). We used the differential entropy features as the inputs of these classifiers. We performed the experiments on the SJTU Emotion EEG Dataset (SEED) for three emotions: happy, sad, and neutral emotions. The first 9 trials were treated as the training data and the rest 6 trials of the same experiments were treated as the test data.

Table 4–1 shows the parameter details of different classifiers for training. For kNN, we used $k = 5$ for baseline in comparison with other classifiers. For LR, we employed $L2$-regularized LR and we tuned the regularization parameter in [1.5:10] with a step of 0.5. SVM with linear kernel was implemented with the LIBSVM toolbox[223]. We searched the parameter space $2^{[-10:10]}$ with a step of one for C to find the optimal value.

Table 4–1 The details of parameters used in different classifiers

Classifiers	Parameter Details
*k*NN	*k*=5
LR $L2$	$L2$-regularized, tune the regularization in [1.5:10] with a step of 0.5
SVM	Linear kernel, search space $2^{[-10:10]}$ with a step of one for C
DBN	Structure with 2 hidden layers: search optimal numbers of neurons at the first and the second hidden layers in the ranges of [200:500] and [150:500], respectively, with step of 50.
	Mini-batch size: 201
	Unsupervised and supervised Learning rate: 0.5, 0.6
	Momentum parameter: 0.1
	Activation function: sigmoid function

For training deep belief networks, the structure had two hidden layers. The optimal numbers of neurons in the first and second hidden layers were tuned in the ranges of [200:500] and [150:500], respectively. The learning rates of unsupervised learning phase and supervised learning phase were set to 0.5 and 0.6, respectively. Momentum in the weight update with the parameter of 0.1 was used to prevent getting stuck in local minima. The DE features were normalized by subtracting the mean and divided by the standard deviation before feeding them into DBNs. DBN was implemented with the DBNToolbox[215].

4.1.4 Classification Performance

The mean accuracies and the standard deviations of DBN and SVM with the DE features from different frequency bands are shown in Table 4–2. 'Total' denotes the direct concatenation of the five frequency bands. Beta and gamma frequency bands achieve better performance than the lower frequency bands, which indicates high frequency band oscillations are more associated with the recognition of three emotions. Moreover, we compare the performance of various EEG features including PSD, DE, DASM, RASM, and DCAU features. The DE features from the total frequency bands achieve the best classification accuracy of 86.08% and the lowest standard deviation of 8.34% among these features. Therefore, the DE features are superior to the other features. The asymmetric features (DASM, RASM and DCAU) obtain comparable performance

Table 4–2 The mean accuracies and standard deviations (%) of SVM and DNN for different kinds of features

Feature	Model	Delta	Theta	Alpha	Beta	Gamma	Total
PSD	SVM	58.03/15.39	57.26/15.09	59.04/15.75	**73.34/15.20**	71.24/16.38	59.60/15.93
	DNN	60.05/16.66	55.03/13.88	52.79/15.38	60.68/21.31	**63.42/19.66**	61.90/16.65
DE	SVM	60.50/14.14	60.95/10.20	66.64/14.41	80.76/11.56	79.56/11.38	**83.99/09.72**
	DNN	64.32/12.45	60.77/10.42	64.01/15.97	78.92/12.48	79.19/14.58	**86.08/08.34**
DASM	SVM	48.87/10.49	53.02/12.76	59.81/14.67	**75.03/15.72**	73.59/16.57	72.81/16.57
	DNN	48.79/09.62	51.59/13.98	54.03/17.05	69.51/15.22	70.06/18.14	**72.73/15.93**
RASM	SVM	47.75/10.59	51.40/12.53	60.71/14.57	74.59/16.18	74.61/15.57	**74.74/14.79**
	DNN	48.05/10.37	50.62/14.02	56.15/15.28	70.31/15.62	68.22/18.09	**71.30/16.16**
DCAU	SVM	55.92/14.62	57.16/10.77	61.37/15.97	75.17/15.58	76.44/15.41	**77.38/11.98**
	DNN	54.58/12.81	56.94/12.54	57.62/13.58	70.70/16.33	72.27/16.12	**77.20/14.24**

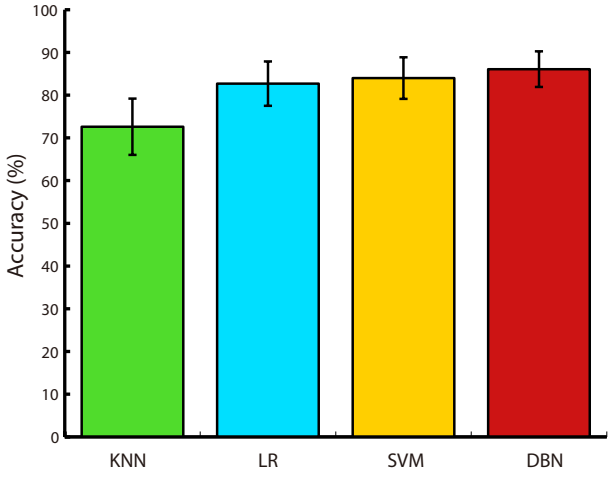


Figure 4–2 The average accuracies of different classifiers using the DE features.

with much fewer dimensions in comparison with the PSD and DE features. These experimental results show the discriminative information of the asymmetrical brain activity (lateralization in left-right direction and caudality in frontal-posterior direction) in emotion recognition.

Figure 4–2 presents the average accuracies of different classifiers using the same DE features of total frequency bands. The means and standard deviations of accuracies of kNN, LR, SVM and DBN are 72.60%/13.16%, 82.70%/10.38%, 83.9%/9.72%, 86.08%/8.34%, respectively. DBN outperforms the other conventional shallow models with higher mean accuracy

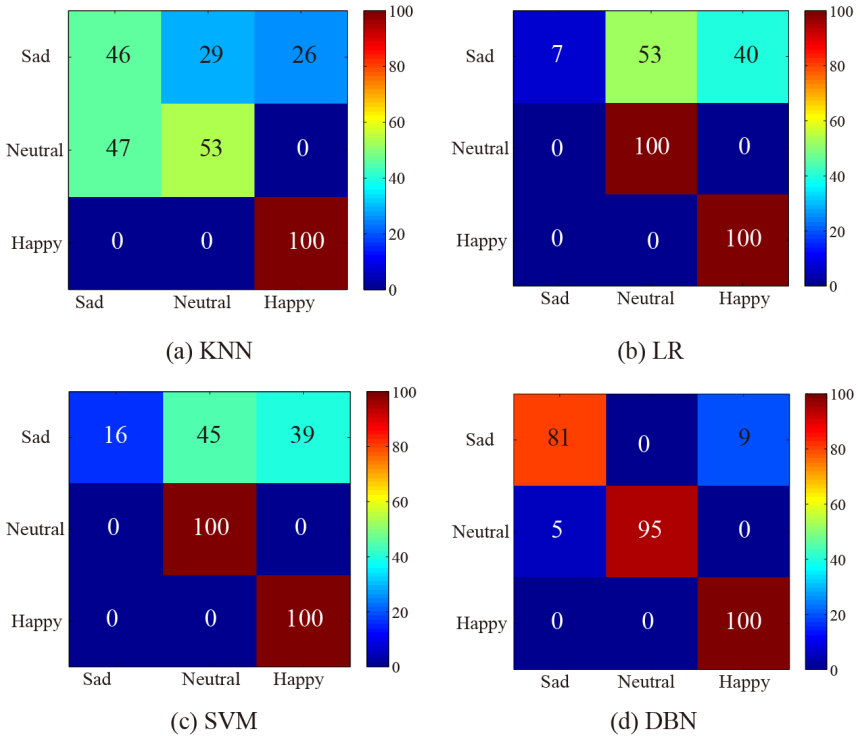


Figure 4–3 The confusion matrix of different classifiers on one experiment for one subject. Here the number inside the figures denotes the recognition accuracy in percentage.

and lower standard deviation. DBN obtains 2.09% higher accuracy and 1.38% lower standard deviation in comparison with SVM. It is usually expensive to collect high quality labeled data, especially for EEG. The unsupervised learning property of DBN in the pre-training step allows to learning efficient feature representations with the large amounts of unlabeled EEG data. In our previous work[224], we also combine DBN and hidden Markov models to capture the dynamic and sequential patterns and bridge the gap between static and sequence pattern recognition.

The confusion matrix of different classifiers on one experiment for one subject is shown in Figure 4–3, which shows the details of strength and weakness of different classifiers. Each row of the confusion matrix represents the target class and each column represents the predicted class that a classifier outputs. The element (i, j) is the percentage of samples in class i that was classified as class j. From the results in Figure 4–3, we can see that in general, happy emotion can be recognized with relatively high accuracies, while sad emotion is the most difficult to recognize. For *k*NN, LR and SVM, they confuse sad emotion with neutral and happy emotion, and cannot classify sad emotion very well. However, DBN can significantly improve the classification accuracies for sad emotion. SVM performs slightly better than LR and can predict more

sad emotion samples accurately. These results show that the deep learning method using DBN has an ability to perform feature selection task to filter out the unrelated features and achieves a better classification accuracy. Feature extraction and feature selection are crucial in the process of emotion modeling. The efficiency of DBN can combine feature extraction and feature selection with both unsupervised and supervised learning. We will further analyze the powerful representations learned from deep belief networks and how it can select the critical channels and critical frequency bands through weight distributions learned from the deep models in the next session.

4.1.5 Critical Frequency Bands and Channels

The higher accuracies of the beta and gamma frequency bands suggest that neural patterns associated with happy, sad, and neutral emotions exist. We visualize the DE features of one experiment in Figure 4–4. The neural responses of the beta and gamma frequency bands enhance for happy emotions in comparison with neutral and sad emotions. While the neural patterns of neutral and sad emotions have similar patterns in the beta and gamma bands, neutral emotions have higher responses of alpha oscillations. These findings provide fundamental evidences about critical frequency bands for emotion recognition, which helps understand the neural mechanism of emotion processing in the human brain.

Previous studies support the relations between different frequency oscillations and mental states. Ray and Cole revealed that EEG alpha bands reflect attentional processing and beta bands reflect emotional and cognitive processing in the brain[86]. Li and Lu showed that gamma bands of EEG are important for emotion classification[119]. Our findings are consistent with the existing results. Neutral emotion with less attention evokes higher alpha response and happy emotion has increased beta and gamma responses.

The above observations are based on the time-frequency analysis. We further try to interpret the connection weights from the trained deep neural networks and investigate the critical frequency bands and channels in the emotion recognition task. According to the rules of knowledge representation, if a particular feature is important, there should be a larger number of neurons involved in representing it in the network[225]. Following this knowledge representing rule in neural network, we assume that the weights of critical channels tend to be updated to certain high values, which can represent how important they are for emotion recognition models.

Figure 4–5 presents the mean absolute weight distribution of the trained DBNs in the first layers with the DE features from the total frequency bands. High peaks are mostly located

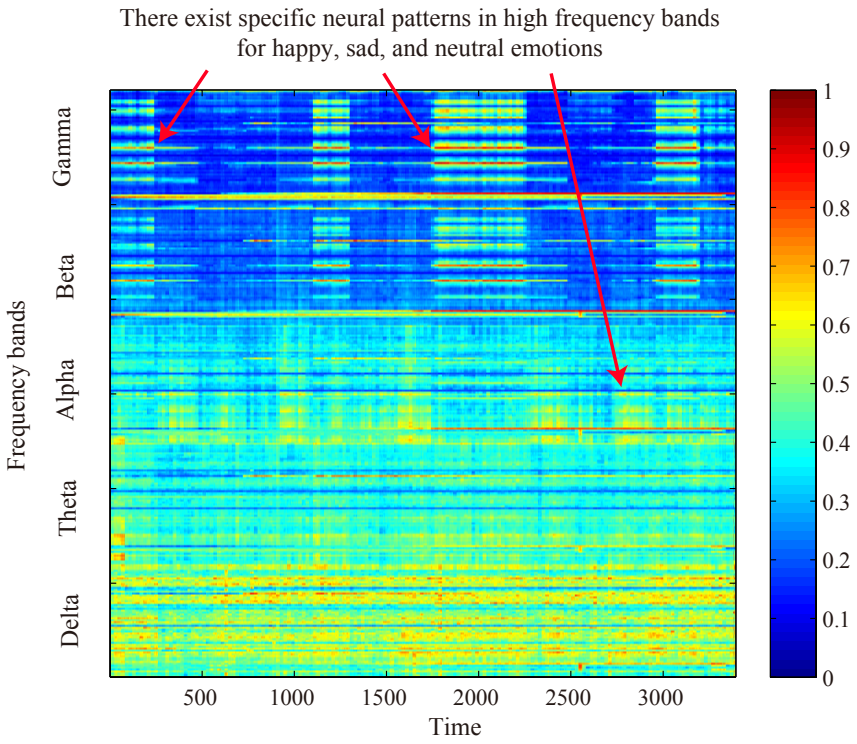


Figure 4–4 The DE feature map in one experiment, where the time frames are on the horizontal axis, and the DE features are on the vertical axis.

at the beta and gamma bands. Since the larger weights of corresponding feature dimensions contribute more to the output of the neurons in neural networks, this phenomena indicates that the feature components of beta and gamma bands contain more important discriminative information for the tasks. In other words, the critical frequency bands for emotion recognition are beta and gamma bands. Previous findings about critical frequency bands are verified with the experimental results.

To explore the critical brain areas, the topography plot of the mean weight distribution is shown in Figure 4–6. From the figure, we can see that the lateral temporal and prefrontal brain areas activate more than other brain areas in beta and gamma frequency bands. Therefore, emotion recognition of happy, sad, and neutral categories has specific neural patterns in terms of critical frequency bands and bran regions.

There is often an interference of muscle artifacts caused by facial expression when recording EEG signals. These artifacts might influence the neural patterns shown in Figure 4–6. But the influence is limited with the following reasons: 1) the dominant EMG activities often happen in the higher frequency bands up to 350 Hz, while the raw EEG signals are preprocessed with

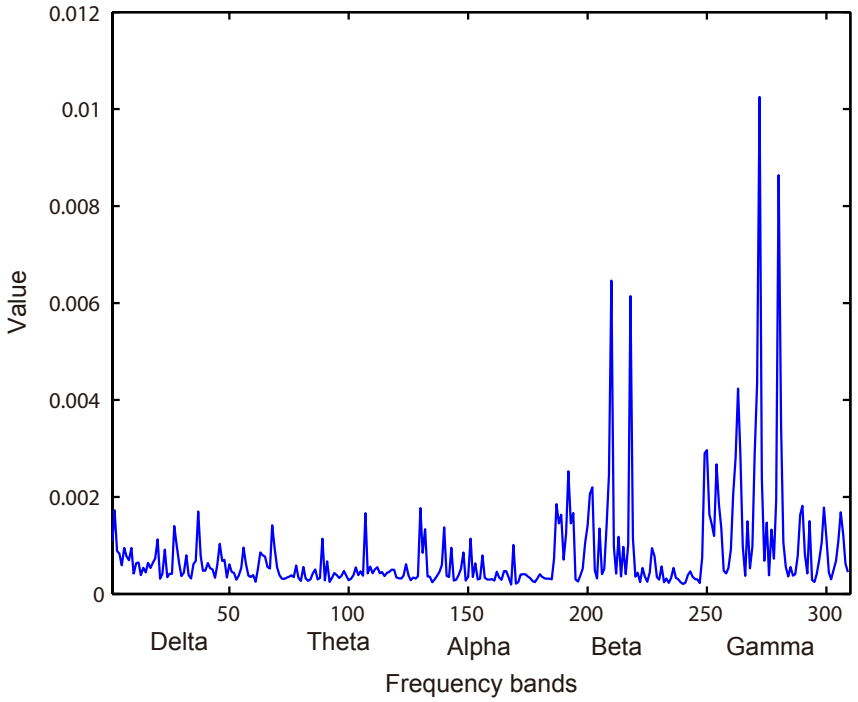


Figure 4–5 The mean absolute weight distribution of the trained DBNs in the first layers with the DE features from the total frequency band.

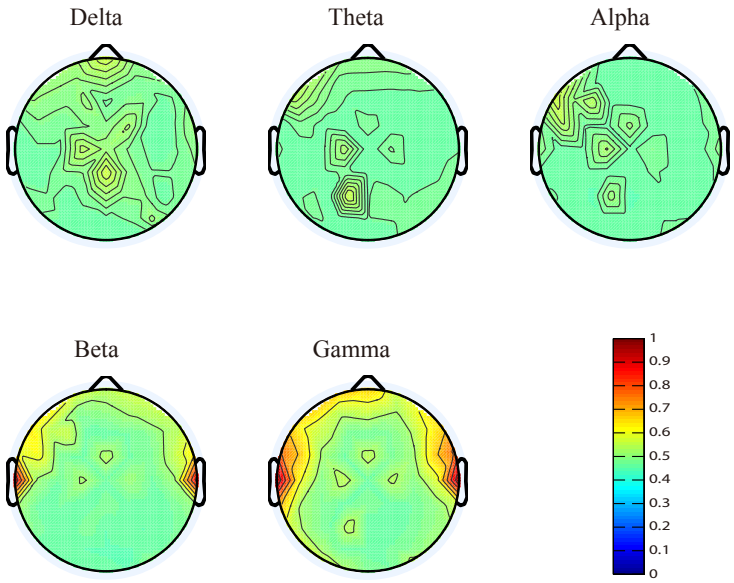


Figure 4–6 The weight distribution of different brain regions in the five frequency bands

a bandpass filter between 0.3 Hz to 50 Hz and the recordings seriously contaminated by EMG are removed manually in our study; 2) the subjects are not asked to show their facial expressions explicitly, the similar patterns are observed for those without explicit facial expressions; and 3) the findings of these neural patterns are consistent with previous emotion studies with EEG.

4.1.6 Electrode Reduction

Based on the above findings about critical frequency bands and brain regions, we further examine the performance of electrode reduction. Electrode reduction can not only reduce the computational complexity, but also filter out the irrelative noise for emotion recognition. Moreover, it provides guidelines for designing wearable brain-computer interfaces in real-world applications. The optimal electrode reduction should be the tradeoff between the detection accuracy and the setup feasibility. Various studies propose different approaches to determining optimal electrode sets for different tasks. We examine whether the findings of the neural patterns could help reduce electrode numbers to a small pool of critical channels while the performance could still be relatively high. We design four different profiles of electrode placements according to the high peaks in the weight distribution and electrode symmetry. Figure 4–7 presents the four evaluated different profiles according to the 10-20 international system: (a) 4 channels [1]; (b) 6 channels [2]; (c) 9 channels [3]; and (d) 12 channels [4]. The electrodes of the profiles (a), (b), and (d) are located in the lateral temporal regions and the profile (c) contains three additional prefrontal electrodes.

We exact the five EEG features (PSD, DE, DASM, RASM, and DCAU) of these designed four electrode profiles and compare the performance with those of the original full 62 channels. Table 4–3 shows the mean accuracies and standard deviations of different electrode profiles. From the Table, we can see that the profile with only four electrodes can achieve comparably high and stable performance with the mean accuracy of 82.88%. It is slightly lower than that of the full 62 channels with the accuracy of 83.99%. Moreover, these four electrodes are located at the lateral temporal area, which can be easily embedded or mounted in wearable devices in real-world scenarios. These experimental results indicate the possibility of developing a wearable brain-computer interface for emotion recognition in human-computer interaction applications.

The DE feature obtains the best performance among the five EEG features for all profiles,

[1] FT7, FT8, T7 and T8

[2] FT7, FT8, T7, T8, TP7 and TP8

[3] FP1, FPZ, FP2, FT7, FT8, T7, T8, TP7 and TP8

[4] FT7, FT8, T7, T8, C5, C6, TP7, TP8, CP5, CP6, P7 and P8

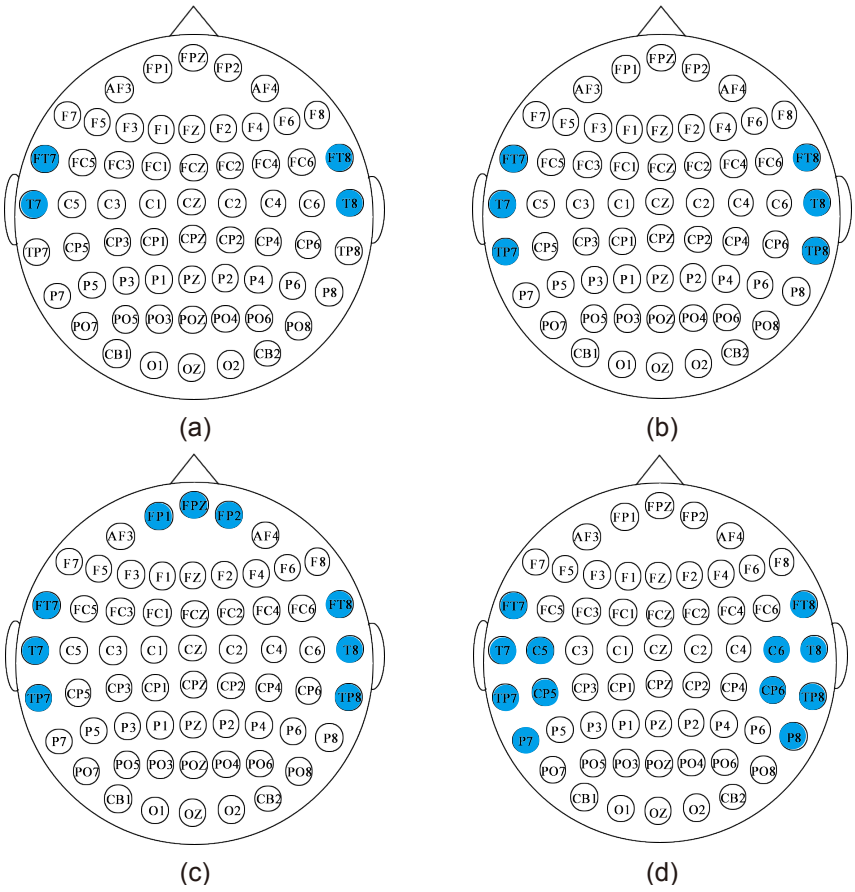


Figure 4–7 Four different profiles of selected electrode placements according to the high peaks in the weight distribution and electrode symmetry: (a) 4 channels; (b) 6 channels; (c) 9 channels; (d) 12 channels.

which indicates the superiority of the DE feature over the others. The best mean accuracies of the 4 channels, the 6 channels, the 9 channels, the 12 channels, and the original full 62 channels are 82.88%, 85.03%, 84.02%, 86.65%, and 83.99%, respectively. In comparison with the 6 channels, the profile of 9 channels has three additional frontal electrodes: FP1, FPZ, and FP2, but has slightly one percentage accuracy lower. However, the profile of 9 channels can still obtain higher accuracies for some participants and the highest accuracies in the beta and gamma bands compared with other electrode reduction. This implies that the three frontal electrodes might be still useful for some conditions, especially for the beta and gamma bands. The profiles of the 6 channels, the 9 channels, and the 12 channels can achieve higher performance than that of the 62 channels. Among these setups, the profile of 12 channels obtains the highest mean accuracy and the lowest standard deviation (86.65%/8.62%), which is better than the original full 62 channels with SVM (83.99%/9.72%) and DBN (86.08%/8.34%). From these experimental results,

Table 4–3 The mean accuracies and standard deviations (%) of SVM for different profiles of electrodes sets

(a) 4 channels

Feature	Delta	Theta	Alpha	Beta	Gamma	Total
PSD	51.38/14.22	48.39/9.04	57.97/15.86	64.28/14.31	66.60/15.90	74.09/15.73
DE	47.84/11.47	48.52/9.19	57.62/16.63	69.89/15.28	69.20/14.87	**82.88/10.92**
DASM	42.24/5.97	40.55/7.15	45.72/9.15	46.35/12.41	47.41/12.48	70.00/14.84
RASM	41.73/5.89	40.14/6.98	45.54/9.03	45.94/12.13	47.86/12.62	69.21/14.76

(b) 6 channels

Feature	Delta	Theta	Alpha	Beta	Gamma	Total
PSD	54.99/14.48	56.27/10.59	63.12/14.77	72.59/15.07	74.36/15.37	70.53/14.69
DE	49.99/12.77	57.55/12.56	66.02/12.57	75.82/13.94	75.28/14.28	**85.03/9.63**
DASM	42.50/9.86	44.11/9.43	53.84/12.62	60.26/12.79	61.23/14.04	74.31/11.90
RASM	41.24/9.37	44.17/9.62	53.82/9.44	61.19/14.21	62.58/14.31	73.81/12.34
DCAU	37.56/7.72	41.14/7.22	43.56/9.73	45.41/9.99	45.65/11.08	71.82/14.28

(c) 9 channels

Feature	Delta	Theta	Alpha	Beta	Gamma	Total
PSD	59.95/15.43	59.03/11.51	70.11/13.28	78.81/11.87	79.03/13.97	69.69/14.43
DE	55.68/14.75	60.65/12.64	71.28/14.20	80.19/10.21	81.33/11.27	**84.02/10.34**
DASM	44.43/10.94	47.25/10.53	58.34/10.29	68.97/15.62	66.38/15.83	75.56/10.23
RASM	43.44/11.64	45.94/10.37	58.17/10.90	65.89/16.23	65.83/14.64	74.80/10.81
DCAU	37.56/7.72	41.14/7.22	43.56/9.73	45.41/9.99	45.65/11.08	71.82/14.28

(d) 12 channels

Feature	Delta	Theta	Alpha	Beta	Gamma	Total
PSD	57.55/14.68	62.73/13.80	65.87/15.81	75.80/12.72	75.68/12.79	62.92/15.64
DE	54.70/13.45	62.13/13.69	68.18/14.90	77.60/13.58	77.86/14.35	**86.65/8.62**
DASM	46.94/10.42	46.85/9.92	59.45/12.37	69.04/15.19	70.61/14.42	75.86/14.06
RASM	45.88/12.21	45.97 /10.28	58.97/11.14	68.80/15.29	70.58/13.66	75.70/13.74
DCAU	37.56/7.72	41.14/7.22	43.56/9.73	45.41/9.99	45.65/11.08	71.82/14.28

we can see that these selected critical channels can achieve relatively high performance across all the experiments of different subjects. Electrode reduction can improve the performance of emotion recognition with less computational and setup cost. According to various requirements of different applications, it is feasible to develop wearable emotion recognition systems with the balance of accurate detection and easy setups.

It should be noted that although the profile of 12 channels outperforms the 62 channels in term of accuracy, the rest 50 electrodes are not 'uninformative' for emotion recognition. We aim to find the minimum numbers of electrode with comparable performance. The neighboring

electrodes usually contain similar redundant information and only the critical electrodes are kept in the optimal electrode sets. Different participants might have different optimal electrode sets due to the individual differences across participants. The optimal electrode setups can be personalized to specific target users with more target data for training.

4.2 Stable EEG Patterns over Time for Emotion Recognition

4.2.1 Introduction

In this section, we investigate the stable EEG patterns over time for emotion recognition using a machine learning approach. Many previous studies propose the activated patterns associated with different emotions and perform subject-dependent and subject-independent evaluations. However, most previous studies report activated patterns across individual but do not consider the time factor. the stability of these neural patterns and performance of computational models over time has not been fully investigated yet. Here, the stable EEG patterns are considered as neural activities in terms of critical frequency oscillations and brain regions that share commonality across individuals and sessions. EEG signals are non-stationary and might change with the differences of cognitive states and environmental factors. However, the EEG patterns of specific tasks exhibit consistency across individuals and sessions. In this study, we try to address the following three issues in EEG-based emotion recognition: a) what is the capability of EEG signals for discriminating between different emotions? (b) are there any stable EEG patterns of neural oscillations or brain regions for representing emotions? and (c) what is the day-to-day performance of the models based on machine learning approaches?

The studies on the internal consistency and test-retest stability of EEG can date back to many years ago[125, 226, 227], especially in clinical applications. However, these studies consider different cognitive states, instead of emotional states. Some researchers have conducted pilot studies on the stability of EEG features for emotion recognition[127, 128]. The stability of neural patterns over time still lacks systematical evaluations. Moreover, the performance of emotion recognition over time is important for developing robust real-world applications.

We perform systematic comparisons and qualitative evaluations of various feature extraction, feature selection, feature smoothing and classification methods on two public emotion EEG dataset: DEAP and SEED. The discriminative Graph regularized Extreme Learning Machine (GELM) is introduced and achieves the best classification performance. The stable neural patterns of happy, sad, and neutral emotions over time and the stability of models within and

between sessions are evaluated.

4.2.2 Discriminative Graph Regularized Extreme Learning Machine

We extracted the six efficient EEG features including PSD, DE, DASM, RASM, ASM, and DCAU that are introduced in Chapter 3 as the inputs of classifiers. Besides the conventional classifiers, e.g., kNN, logistic regression (LR) and SVM, we introduced a newly developed neural network called discriminative Graph regularized Extreme Learning Machine (GELM)[228]. For kNN, we used the Euclidean distance as the distance metric and set $k = 5$ after cross-validation. LIBLINEAR toolbox[229] was used to implement SVM with linear kernel and the soft margin parameter C was tuned using cross-validation.

Extreme Learning Machine (ELM)[230] is a feed-forward neural network with a single hidden layer, where the weight connections between inputs and hidden neurons are randomly assigned and the output weights can be learned with the least square method. ELM can usually achieve comparable performance with much less training time, which has been adopted in various applications. Inspired by learning with local consistency of data, Peng *et al.* proposed the discriminative Graph regularized Extreme Learning Machine (GELM) for face recognition[228]. The basic idea is that the local consistency property can be used to enhance the performance by incorporating the graph regularization term into conventional ELM model. The local consistency assumes that nearby samples are embedded in a manifold space and share similar properties. The constraint of the local consistency imposed on output weights enforces the outputs of the same class to be similar in GELM.

Given a training data set,

$$L = \{(x_i, t_i) | x_i \in R^d, t_i \in R^m\}, \tag{4–7}$$

where $x_i = (x_{i1}, x_{i2}, \cdots, x_{id})^T$ and $t_i = (t_{i1}, t_{i2}, \cdots, x_{im})^T$. In GELM, the adjacent W is defined as follows,

$$x_i = \begin{cases} 1/N_t, & \text{if } h_i \text{ and } h_j \text{ belong to the } t\text{th class} \\ 0, & \text{otherwise,} \end{cases} \tag{4–8}$$

where $h_i = (g_1(x_i), \cdots, g_K(x_i))^T$ and $h_j = (g_1(x_j), \cdots, g_K(x_j))^T$ are hidden layer outputs for two input samples x_i and x_j. We can then compute the graph Laplacian $L = D - W$, where D is a diagonal matrix and each of the entries in D contains the column sums of W. Therefore, GELM can incorporate two regularization terms into the conventional ELM model.

The objective function of GELM is defined as follows,

$$\min_{\beta} ||H\beta - T||_2^2 + \lambda_1 Tr(H\beta L\beta^T H^T) + \lambda_2 ||\beta||_2^2, \tag{4–9}$$

where $Tr(H\beta L\beta^T H^T)$ is the graph regularization term, $||\beta||^2$ is the l_2-norm regularization term, and λ_1 and λ_2 are regularization parameters to balance the two terms.

By setting the derivative of the objective function (4–9) with respect to β as zero, we have

$$\beta = (HH^T + \lambda_1 HLH^T + \lambda_2 I)^{-1} HT. \tag{4–10}$$

In GELM, the graph regularization term in the objective function imposed on the output weights enforces the outputs of samples from the same class to be similar.

Peng and colleagues[228] presented the motivation of GELM with superiority over other models in two folds. The Local Consistency property can be used as side information for improving the performance of learning models. GELM preserves the local structure during feature learning, which make the learned mapping function vary smoothly along the geodesics of the data manifold. Moreover, the distance information among samples are destroyed by nonlinear mapping in conventional ELM. The nonlinear mapping enhances the representative capacity of neural networks while destroying the local consistency. Distance information based local consistency is not proper to construct nearest neighbor graph. In GELM, label information is used to construct the adjacent matrix.

4.2.3 Experiment Results on DEAP Data

This section presents the experimental results on the publicly available emotion EEG dataset called DEAP [1][97]. We first extracted various features and classifiers and compared the performance with the state-of-the-art approaches on the same dataset.

The DEAP dataset consists of EEG and peripheral physiological signals of 32 participants. For each experiment, the participants were asked to watch 40 trials of one-minute duration music videos and elicit their own emotions. There are 32-channel EEG recordings and 8-channel peripheral physiological signals, including galvanic skin response, skin temperature, blood volume pressure, respiration rate, electromyogram and electrooculogram (horizontal and vertical). After each trial, the participants are asked to rate their affective states using the SAM questionnaires with valence, arousal, and dominance dimensions. More details about the DEAP dataset can refer to the reference[97].

[1] http://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html

In this experiment, we used the valence-arousal model as the emotion representation. In the SAM feedback, each dimension has values ranging from 1 to 9. We segmented the four quadrants of the valence-arousal (VA) space according to the ratings from the participants. LALV, HALV, LAHV, and HAHV denote low arousal/low valence, high arousal/low valence, low arousal/high valence, and high arousal/high valence, respectively. Considering the fuzzy boundary of emotions and the variations of participants' ratings possibly associated with individual differences on the rating scale, we added a gap to segment the quadrants of VA space to ensure the correct ratings of participants' true self-elicitation emotion. The EEG data with ratings of arousal and valence between 4.8 and 5.2 were discarded. The ratings of arousal and valence and the corresponding instance numbers for the quadrants of VA space are shown in Table 4–4. The rating distribution of DEAP on the arousal-valence plane (VA plane) for the four conditions is shown in Figure 4–8. We can see that the ratings are distributed approximately uniformly[97]. We labeled the EEG data according to the corresponding ratings of valence and arousal from the participants.

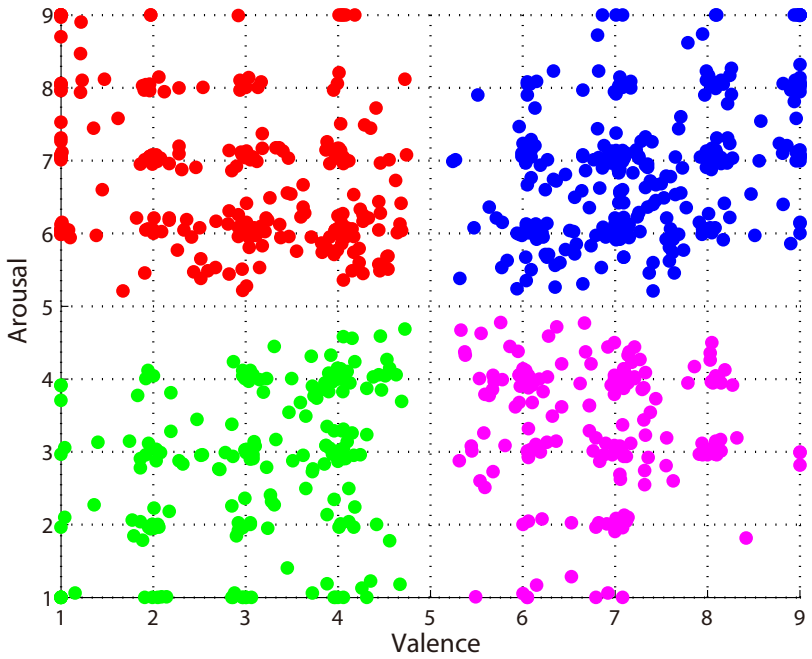


Figure 4–8 The rating distribution of DEAP on the arousal-valence plane (VA plane) for the four conditions (LALV, HALV, LAHV, and HAHV)

The raw EEG data were first down-sampled to 128 Hz and preprocessed with a bandpass filter between 4 Hz and 45 Hz. Six EEG features including the PSD, DE, DASM, RASM, ASM,

Table 4–4 The ratings of arousal and valence for the quadrants of VA space.

Quadrants	LALV	HALV	LAHV	HAHV
Arousal	<4.8	>5.2	<4.8	>5.2
Valence	<4.8	<4.8	>5.2	>5.2
#instances	12474	16128	10962	21420

Table 4–5 The mean accuracy rates (%) of SVM and GELM classifiers for different features obtained from separate and total frequency bands in DEAP.

Feature	Classifier	Theta	Alpha	Beta	Gamma	Total
PSD	SVM	32.86	33.49	33.73	31.99	**36.19**
	GELM	**61.78**	61.14	61.77	61.56	61.46
DE	SVM	44.31	41.59	43.54	42.74	**47.57**
	GELM	61.45	61.65	62.17	61.45	**69.67**
DASM	SVM	**43.18**	42.72	42.07	41.24	40.70
	GELM	**57.86**	57.02	56.08	56.48	52.54
RASM	SVM	**54.34**	52.54	53.12	52.76	51.83
	GELM	46.66	44.52	44.88	45.56	**52.70**
ASM	SVM	**45.03**	44.03	43.59	44.03	40.76
	GELM	**56.16**	54.30	54.40	54.73	51.82
DCAU	SVM	**41.51**	41.05	41.27	40.00	40.85
	GELM	57.47	56.58	**58.25**	**58.25**	55.26

and DCAU features were extracted in the four frequency bands: theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (31-45 Hz). Since the EEG data of DEAP were preprocessed with the bandpass filter, the delta frequency band was not included for analysis. The linear dynamic system algorithm was used for feature smoothing. The neuron number of the hidden layer for GELM was set to 10 times the input dimensions. We adopted the five-fold cross-validation evaluation for each experiment. The classifiers kNN, logistic regression and SVM with linear and RBF kernels were used as the baseline method.

Table 4–5 shows the mean accuracies of SVM and GELM with different EEG features. 'To-

Table 4–6 Comparison of the various studies using EEG in the DEAP dataset.

Study	Results
Chung *et al.*[231]	66.6%, 66.4% for valence and arousal (2 classes), 53.4%, 51.0% for valence and arousal (3 classes) with all 32 participants.
Koelstra *et al.*[97]	62.0%, 57.6% for valence and arousal (2 classes) with all 32 participants.
Liu *et al.*[232]	63.04% for arousal-dominance recognition (4 classes) with the selected 10 participants.
Zhang *et al.*[233]	75.19 % and 81.74 % on valence and arousal (2 classes) with the selected eight participants.
Our method	69.67% for quadrants of VA space (4 classes) with all 32 participants.

tal' denotes the direct concatenation of all features from the four frequency bands. The DE features outperform the other EEG features. The best accuracies of SVM and GELM were 54.34% and 69.67%, respectively. We also evaluate the performance of kNN, logistic regression and SVM with the RBF kernel on DEAP, which achieve the respective accuracies (%) and standard deviations (%) of 35.50/14.50, 40.86/16.87, and 39.21/15.87, respectively, using the DE features of the total frequency bands. We perform one-way analysis of variance (ANOVA) to study the statistical significance. The DE features outperform the PSD features significantly ($p < 0.01$), and for classifiers, the performance of GELM is better than that of SVM ($p < 0.01$). As we can see from Table 4–5, the diversity of classification accuracy for different frequency bands is not significant for the DEAP dataset ($p > 0.95$). The results here do not demonstrate any critical frequency bands for the quadrants of the VA space. We can also see that the asymmetric features of the DCAU features achieve comparable accuracies. These results indicate that there exists some asymmetry properties with discriminative information for the four affect conditions (LALV, HALV, LAHV, and HAHV).

We compare the performance of our proposed method with the state-of-the-art approaches on the same dataset, as presented in Table 4–6. Here, we only analyze the related results with single EEG modality. Chung *et al.*[231] defined a weighted-log-posterior function for the Bayes classifier and performed the evaluations on the DEAP dataset. The accuracies for valence and arousal classification are 66.6% and 66.4% for two classes and 53.4% and 51.0% for three classes, re-

spectively. Koelstra *et al.*[97] developed the DEAP dataset and obtained an average accuracy of 62.0% and 57.6% for the valence and arousal (2 classes), respectively. Liu *et al.*[232] proposed a real-time fractal dimension-based valence level recognition algorithm from EEG signals and obtained a mean accuracy of 63.04% for arousal-dominance recognition (4 classes) with the selected 10 participants. Zhang *et al.*[233] described an ontological model for representation and integration of the EEG data, and their model yielded an average recognition ratio of 75.19% for valence and 81.74% for arousal for the eight participants. Although their accuracies were relatively high, there are only two categories for each dimension, and these results were achieved using a subset of the original dataset. In contrast, from Table 4–5, we can see that our method GELM achieves an average accuracy of 69.67% on the same data set for quadrants of the VA space (LALV, HALV, LAHV, and HAHV) with DE features of total frequency bands for all 32 participants. These results shows the superiority of our proposed method over the existing method on EEG-based emotion recognition. It should be noted that more good results on DEAP have been published with recent advanced machine learning methods, especially deep neural networks, after our published results[234, 235].

4.2.4 Experiment Results on SEED Data

We will present the experimental results on the other public emotion EEG dataset called SEED developed in this study. SEED contains three emotion categories: happy, sad, and neutral emotions. Compared with DEAP, SEED contains the EEG data of three sessions that were preformed on different days from the same participants. The time interval between two sessions was about one week. Therefore, we can study the stability of neural patterns over time. The detailed introduction of SEED is included in Chapter 3.

Table 4–7 The means and standard deviations of accuracies (%) for the PSD, DE, DASM, RASM, ASM and DCAU features from the total frequency bands on SEED.

Feature	PSD	DE	DASM	RASM	ASM	DCAU
Mean	72.75	**91.07**	86.76	86.98	85.27	89.95
Std.	11.85	**7.54**	8.43	9.09	9.16	6.87

Similar to the setups in DEAP, we adopted the five-fold cross-validation in SEED. Table 4–7 presents the mean accuracies of various EEG features. The DE feature obtains a higher accuracy

and lower standard deviation among the features. Although the asymmetry features have much fewer feature dimensions, they achieve significantly better performance than the PSD features. Asymmetrical characteristics are helpful for classifying happy, sad, and neutral emotions. We compare the performance of two feature smoothing methods: linear dynamic system (LDS) approach and moving average algorithm in Table 4–8. The size of the moving window is set to 5 in the experiments. The means and standard deviations of the accuracies in percentages (%) for without smoothing, moving average, and the LDS approach are 70.82/9.17, 76.07/8.86 and 91.07/7.54, respectively. We can see that the LDS approach significantly outperforms the moving average method ($p < 0.01$), which achieves 14.41% higher accuracy. The results also demonstrate that feature smoothing plays a significant role in EEG-based emotion recognition.

Table 4–8 The Performance (%) of Different Feature Smoothing Methods on SEED.

Stats	Without Smooth	Moving Average	LDS
Mean	70.82	76.07	**90.48**
STD.	9.17	8.86	**7.89**

We compare the performance of four different classifiers, kNN, Logistic Regression, SVM and GELM. The DE features from the total frequency bands were used as the inputs of the classifiers. The parameter k of kNN was fixed to be the constant value of five. For LR and linear SVM, grid search with cross-validation was used to tune the parameters. The mean accuracies and standard deviations of the five classifiers are demonstrated in Table 4–9. From the above results, we can see that GELM outperforms other classifiers with higher accuracies and lower standard deviations, which imply the superiority of GELM for EEG-based emotion recognition.

Table 4–9 The Mean Accuracies (%) of the Five Classifiers (KNN, LR, SVM and GELM) on SEED.

Stats	Classifiers				
	KNN	LR	SVM (RBF)	SVM (Linear)	GELM
Mean	70.43	84.08	78.21	83.26	**90.48**
Std.	12.73	8.77	9.72	9.08	**7.89**

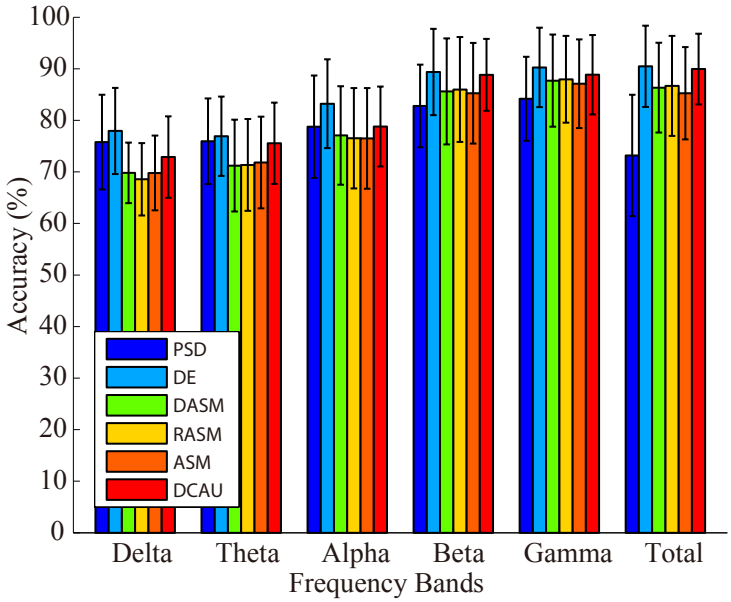


Figure 4–9 The average accuracies of GELM using different features obtained from five frequency bands and using a fusion method on SEED.

4.2.5 Neural Signatures and Stable Patterns

Figure 4–9 presents the performance of different features from the individual five frequency bands and the total frequency band on SEED. From this figure, we can see that the beta and gamma frequency bands perform better than the lower frequency bands, which verify the previous findings about the critical frequency bands. The beta and gamma oscillations are more related to the emotion recognition of these three emotional states than the other frequency oscillations.

We further visualize the EEG features to study the specific patterns of different emotions. Figure 4–10 shows the spectrogram of the electrode T7 in one experiment and Figure 4–11 presents the average spectrogram over participants for each sessions at some electrodes (FPZ, FT7, F7, FT8, T7, C3, CZ, C4, T8, P7, PZ, P8 and OZ). The spectrograms have specific patterns for different emotions. The higher frequency oscillations of the temporal areas are related to the happy and sad emotions. The same participants performed the experiments three times on different days. The neural patterns over time are relatively stable across sessions from Figure 4–11.

We project the DE features to the scalp to observe their temporal dynamics and the stable patterns in terms of frequency oscillations and brain regions. Figure 4–12 depicts the aver-

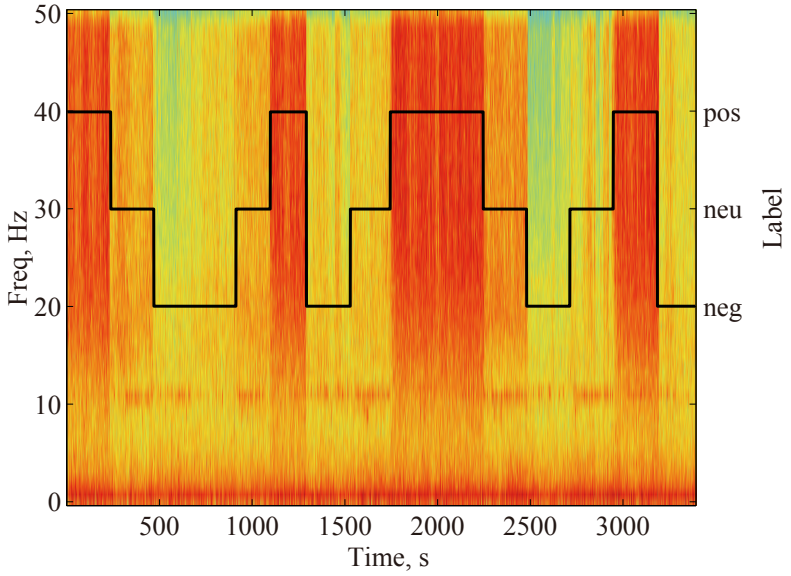
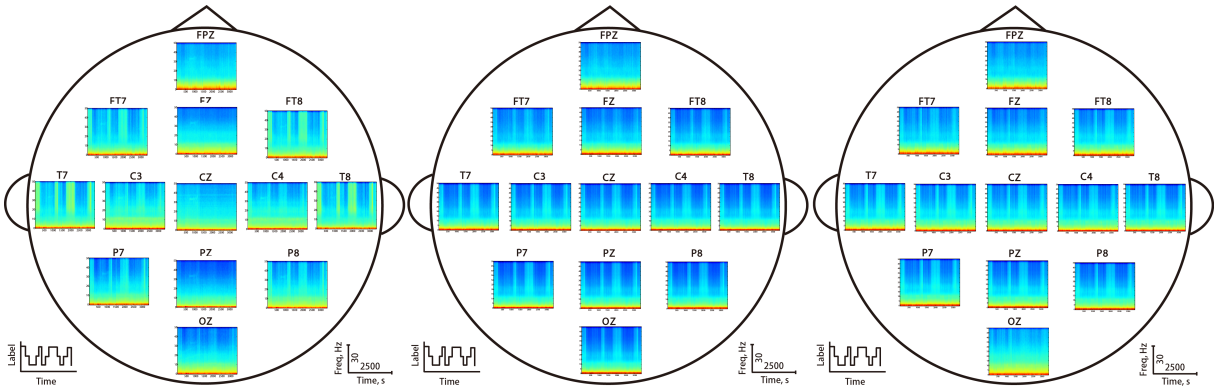


Figure 4–10 The spectrogram of the electrode position T7 in one experiment. As different emotions elicited, we can see that the spectrogram has different patterns.

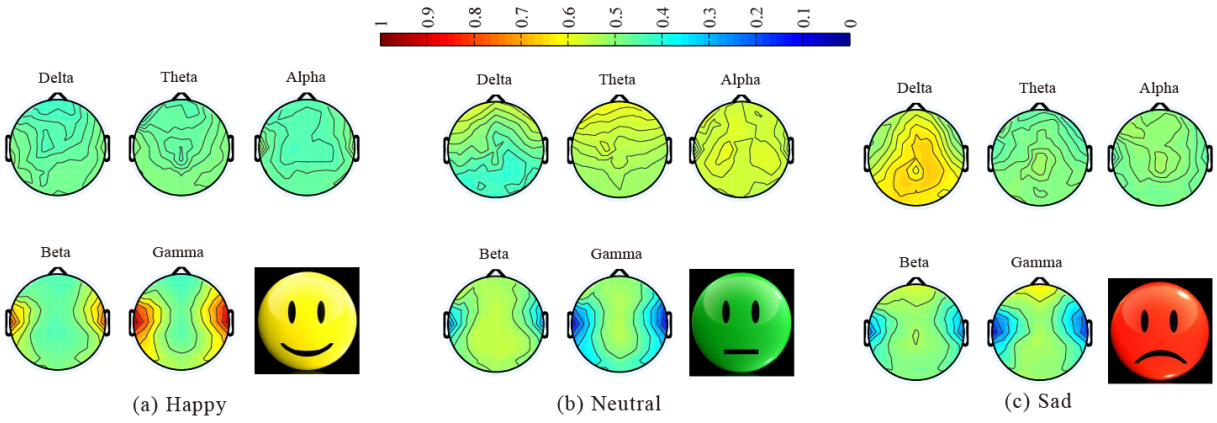
age neural patterns for happy, sad, and neutral emotions. The neural signatures associated with happy, sad, and neutral emotions do exist. The responses of the lateral temporal area enhance for happy emotion while those of the prefrontal area increase for sad emotion in the beta and gamma bands. The neural patterns of neutral and sad emotions are similar to each other with common less activation in the temporal area. The neural patterns of neutral emotion have higher alpha responses at parietal and occipital sites. For sad emotion, the neural patterns have significant higher delta responses at parietal and occipital sites and significantly higher gamma responses at prefrontal sites. The existing studies[86],[236] have shown that EEG alpha activity reflects attentional processing and that beta activity reflects emotional and cognitive processes. When the participants watched the neutral stimuli, they tended to be more relaxed and less attentional, which reflect enhanced alpha responses. For happy emotion processing, the energy of the beta and gamma responses was increased. The findings of these neural patterns are consistent with previous emotion studies[86, 105, 113, 124].

Based on the findings of the neural patterns in terms of critical frequency bands and brain regions, we investigate the performance of different dimensionality reduction. There exists a low-dimension manifold structure for emotion related EEG features. We compared two kinds of dimensionality reduction methods, i.e., the principle component analysis (PCA) algorithm and the minimal redundancy maximal relevance (MRMR) algorithm, with DE features of 310



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

Figure 4–11 The average spectrogram of the participants for each session at some electrodes, which shows the stable neural patterns over time in the temporal lobes and high-frequency bands (a red color indicates a high amplitude).



(a) Happy (b) Neutral (c) Sad

Figure 4–12 The average neural patterns for all participants and sessions for different emotions, which shows that neural signatures associated with happy, neutral and sad emotions do exist.

dimensions as inputs and GELM as a classifier.

Figure 4–13 shows the results of the dimensionality reduction. We find that dimensionality reduction does not affect the performance of our model greatly. For the PCA algorithm, when the dimension is reduced to 210, the accuracy drops from 91.07% to 88.46% and then reaches a local maximum value of 89.57% at the dimension of 160. For the MRMR algorithm, the accuracies vary slightly with lower dimension features. In comparing with PCA, MRMR algorithm is more suitable for EEG-based emotion recognition. Because the MRMR algorithm yields the best emotion-relevant and minimal redundancy features, it also preserves the original domain information such as channel and frequency bands, which have interpretable features for

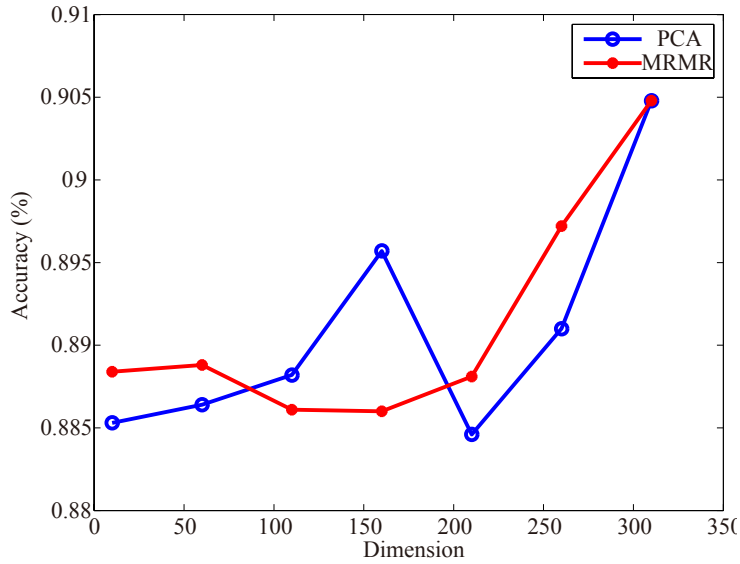
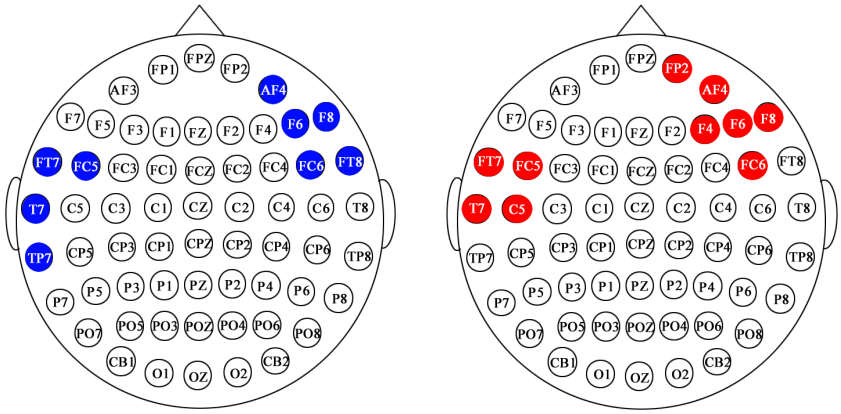


Figure 4–13 The results of dimensionality reduction using PCA and MRMR



(a) Beta Band (b) Gamma Band

Figure 4–14 Distribution of the top 20 participant-independent features selected using the correlation coefficients. One electrode location "FT8" that is not shown is from the alpha frequency band.

understanding the underlying mechanisms. We select the 20 top participant-independent features using the correlation coefficients between emotion annotations and individual features as shown in Figure 4–14. Most of these selected features are located in the temporal and frontal brain regions in the beta and gamma bands, which are consistent with our previous analysis about the neural patterns.

Table 4–10 The average accuracies (%) of our emotion model across sessions.

Stats.	Train	Test		
		First	Second	Third
Mean	First	**90.83**	72.55	67.22
	Second	75.86	**88.22**	76.62
	Third	76.28	78.17	**87.80**
Std.	First	**8.64**	10.29	10.42
	Second	**7.71**	8.59	15.34
	Third	11.47	13.41	**10.97**

4.2.6 Stability of the Emotion Recognition Model over Time

The feature of SEED is that every participant performed the experiments in three sessions from different days. Therefore, we can evaluate the stability of the computational models over time. Different sessions of the same subjects are split into training and test datasets. The input features are the DE features from the total frequency bands and GELM is used as the state-of-the-art model.

Tables 4–10 and 4–11 show the mean accuracies and the individual accuracies of the cross-session experiments. The experiments with the training set and test set from the same sessions obtain higher performance than those from different session. And the accuracies of the model with the training and test sets from the nearer sessions usually are higher. Our proposed emotion recognition model with GELM achieve a mean accuracy of 79.28% across sessions for three emotions. These experimental results indicate that the cross-session differences in EEG-based emotion recognition, even for the same participants. The neural patterns of different emotions are relatively stable for one subject over a period time. However, With the passage of time, the performance of the model may become worse. Therefore, the adaption of the computational model should be further studied in the future.

So far, we only analyze the participant-dependent situations. Here, we would like to examine the cross-participant performance. We adopt a leave-one-out cross validation method to investigate the performance of the participant-independent models. Here, we use the DE features of total frequency bands and SVM with linear kernels for comparison. The average accuracy and standard deviation with participant-independent features are 60.93% and 13.95%, respec-

Table 4–11 The classification accuracies (%) of the training and test data from different sessions using GELM.

Participant	Train	Test			Participant	Train	Test		
		1st	2nd	3rd			1st	2nd	3rd
#1	1st	**91.62**	60.26	60.26	#2	1st	**100.00**	68.28	60.84
	2nd	68.28	**80.2**	68.28		2nd	**85.12**	71.68	81.65
	3rd	68.28	52.53	**92.56**		3rd	85.12	80.42	**90.82**
#3	1st	95.95	**100.00**	75.51	#4	1st	**100.00**	68.28	68.28
	2nd	76.95	**97.04**	82.95		2nd	83.02	**100.00**	91.69
	3rd	80.20	**88.08**	68.93		3rd	76.81	**100.00**	**100.00**
#5	1st	**75.94**	59.61	61.05	#6	1st	**79.70**	71.60	53.83
	2nd	**80.78**	75.00	69.65		2nd	67.92	**80.78**	55.71
	3rd	56.50	54.48	**98.12**		3rd	70.30	71.75	**86.27**
#7	1st	**90.39**	66.69	66.47	#8	1st	**75.07**	67.77	51.01
	2nd	67.49	**95.81**	59.83		2nd	72.83	**91.33**	45.95
	3rd	**83.60**	81.79	74.93		3rd	59.61	**75.94**	73.05
#9	1st	**91.98**	80.56	78.47	#10	1st	**85.12**	70.38	69.29
	2nd	81.36	**100.00**	95.95		2nd	60.12	86.05	**87.14**
	3rd	93.42	**95.52**	93.42		3rd	87.07	83.02	**95.74**
#11	1st	**96.24**	67.99	76.01	#12	1st	**86.78**	74.86	63.29
	2nd	77.89	85.33	**95.59**		2nd	84.39	**91.62**	68.06
	3rd	65.39	66.04	**100.00**		3rd	75.58	**83.45**	73.48
#13	1st	**93.71**	76.88	92.63	#14	1st	**100.00**	86.27	64.02
	2nd	70.38	90.75	**94.00**		2nd	76.23	**86.42**	77.67
	3rd	84.10	87.21	**100.00**		3rd	**91.69**	82.15	89.31
#15	1st	**100.00**	68.79	67.34					
	2nd	85.12	**91.26**	75.14					
	3rd	66.47	70.16	**80.35**					

'1st', '2nd', and '3rd' denote the data obtained from the first, second, and third experiments, respectively, for a participant.

tively. The classification performance decrease moderately due to the challenges of individual differences. Therefore, building robust affective brain-computer interactions should consider the personalized training in the future.

The experimental results demonstrate that the performance of cross-session and cross-participant are worse than those of single experiment. Although different emotions share some commodities of neural patterns, there are still some individual differences across sessions and

participants. These variations from participants to participants and from sessions to sessions cause the discrepancy of feature distributions. This is why the average accuracies of the classifiers trained and tested on each individual participant or session are much higher than those of a classifier trained on a set of participants or sessions and tested on other participants or sessions. How to tackle these differences across participants and environmental changes is one of the key challenges in affective brain-computer interactions. In this way, we can recycle the labeled data from a set of participants in the database for model training and make inference on the new data from target participants. We will introduce and discuss how we deal with these problems with advanced transfer learning in the following Chapter of the thesis.

4.3 Summary

In this Chaper, we have presented the experimental results of EEG-based emotion recognition. Section 4.1 introduces the application of deep neural networks to investigate the critical frequency bands and brain regions while enhancing the classification performance. The electrode numbers can be reduced to a limited subset with comparable or even higher performance. These experimental results provide the guidelines for developing wearable devices in real-world scenarios. In Section 4.2, we focus on identifying the stable neural patterns over time for happy, sad, and neutral emotions. The proposed approaches including feature extraction, feature smoothing, feature reduction, and classification methods are systematically evaluation on two public datasets: DEAP and SEED, which achieve the state-of-the-art results. We do not simply treat the problem as a pattern classification, instead we aim to interpret the underlying neural pattern of different emotions and whether these neural patterns are relatively stable cross participants and sessions. We find that these stable neural patterns share commonality but also have specific individual differences, which poses technical challenging for building robust emotion recognition models. We will present how we apply the transfer learning algorithms to tackle these problems in the rest of the thesis.

Chapter 5

Multimodal Emotion Recognition with EEG and Eye Movements

In this Chapter, we propose a multimodal emotion recognition framework called *EmotionMeter* with the combination of EEG and eye movements. Section 5.1 introduces the motivation of fusing the EEG and eye movements for emotion recognition. In the Section 5.2, we present a bimodal deep auto-encoder to extract the shared representations of EEG and eye movements. Section 5.3 demonstrates the conventional fusion strategies, including the feature-level fusion and the decision-level fusion. We compare the performance of various modality fusion methods on two multimodal datasets for emotion recognition. Section 5.4 and Section 5.5 show the experimental results on SEED for three emotions and SEED-IV for four emotions. We evaluate the performance of multimodal approaches with single modality and analyze the complementary characteristics between EEG and eye movements for emotion recognition. In Section 5.6, we summarize the work of multimodal emotion recognition in this Chapter. The work in this Chapter has been published at IEEE Transactions on Cybernetics[185] and International Joint Conference on Artificial Intelligence 2015[237].

5.1 Introduction

Emotion is associated with various external and internal activities. Different modalities describe different aspects of emotions. It is technically difficult to build robust emotion recognition systems with single modality. Therefore, researchers proposed various fusion approaches to integrating different modalities with complementary information for multimodal emotion recognition. Many studies utilize auditory and visual modalities or different physiological signals. In this study, we proposed an efficient multimodal framework called *EmotionMeter* to combining EEG and eye movements for emotion recognition. EEG and eye movements contains different clues from internal cognitive states and external subconscious behaviors, respectively. Moreover, recent studies proposed the importance of contextual analysis in emotion recognition[144]. Eye movements provide the context-related clues in a natural way among various sensors, which

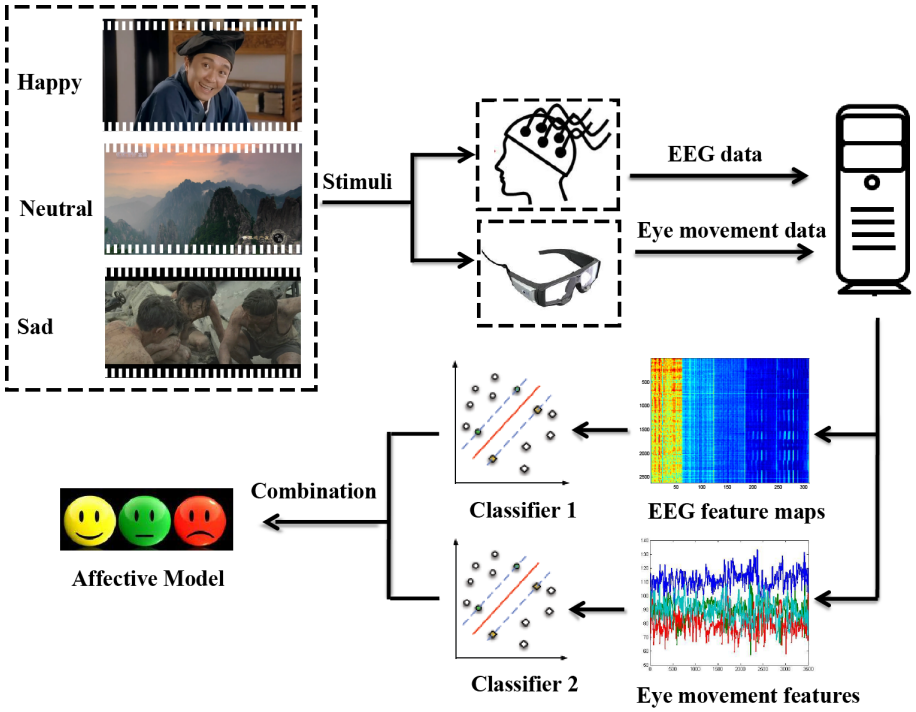


Figure 5–1 The multimodal framework with EEG and eye movements

are promising for affective computing.

Various wearable prototypes for affective sensing and feedbacks have been developed. MacLean *et al.* developed a wearable butterfly called *MoodWings*, which can respond to users' affective states with wing actuation[238]. Williams and colleagues designed a wearable scarf with modular actuation components called *SWARM* that can detect and interpret emotional states[239]. Another affective valence detection system called *EnviroPulse* has been proposed to evaluate the surrounding environments[143]. Hassib *et al.* recorded one-channel EEG from the frontal cortex and computed the engagement index with their proposed system *EngageMeter*[240]. In order to increase the feasibility and wearability of affective brain-computer interfaces, we design a multimodal framework called *EmotionMeter* with six-electrode above the ears for EEG recordings and wearable eye tracking glasses. The efficiency, complementary characteristics, and stability across sessions of the proposed *EngageMeter* are systematically evaluated. Figure 5–1 shows the multimodal framework adopted in this work.

Facial expression is one of the popular modalities used for emotion recognition with the advantages of nonintrusive and low-cost setups as well as reasonable accuracy. With much progress in computer vision, there are some bottlenecks for emotion recognition using facial

expression: 1) facial expression can be subjectively controlled and most systems are trained and tested with posed instead of naturalistic facial expressions; 2) illumination variations and heavy occlusion influence the detection performance in real-world scenarios; and 3) facial expression does not consider the contextual factors in different situations. In comparison with facial expression, eye movements also have the wearable and nonintrusive properties. Moreover, they provide physiological signals, e.g., pupil response, and contextual clues, e.g., regions of interests for emotion recognition. One of the core technologies in recent popular virtual reality devices is the eye tracking. Eye movement-based emotion recognition can enhance the affective interactions between human and virtual agents or environments.

5.2 Multimodal Deep Learning

To efficiently combine EEG and eye movements, we adopt a multimodal deep neural network called bimodal deep auto-encoder (BDAE)[141] to extract the high-level shared representations across these two modalities. It is not efficient to feed the direct concatenation of both feature vectors from different modalities to the neural networks as the conventional approaches. Figure 5–2 depicts the architecture of the designed multmodal deep neural network. There are mainly three steps for training the neural networks: (a) two restricted Boltzmann machines (RBMs) are constructed using EEG and eye movement features as input; (b) the two hidden layers of EEG RBM and eye RBM are concatenated, and an upper RBM is trained above them; and (c) The stacked RBMs are unfolded into a bimodal deep auto-encoder (BDAE). The weights are fine tuned to minimize the error between inputs and reconstructions using the unsupervised back-propagation algorithm. In this way, the high-level shared representations of both modalities are learned from the neural networks. We further train linear SVMs as classifiers using the new representations as inputs.

Here, we describe the details of parameters in neural network training. The input features of EEG and eye movements were firstly normalized to the range between zero and one before modal training. There were three RBMs and the neuron numbers of the hidden layers were tuned in the range of [200, 150, 100, 90, 70, 50, 30, 20, 15, 10] using cross-validation. The learning rate and mini-batch size were set to the constants of 0.001 and 100, respectively. We implemented the multimodal neural networks in Python using the open source deep learning libraries Keras[1] and Tensorflow[2].

[1] https://keras.io/

[2] https://www.tensorflow.org/

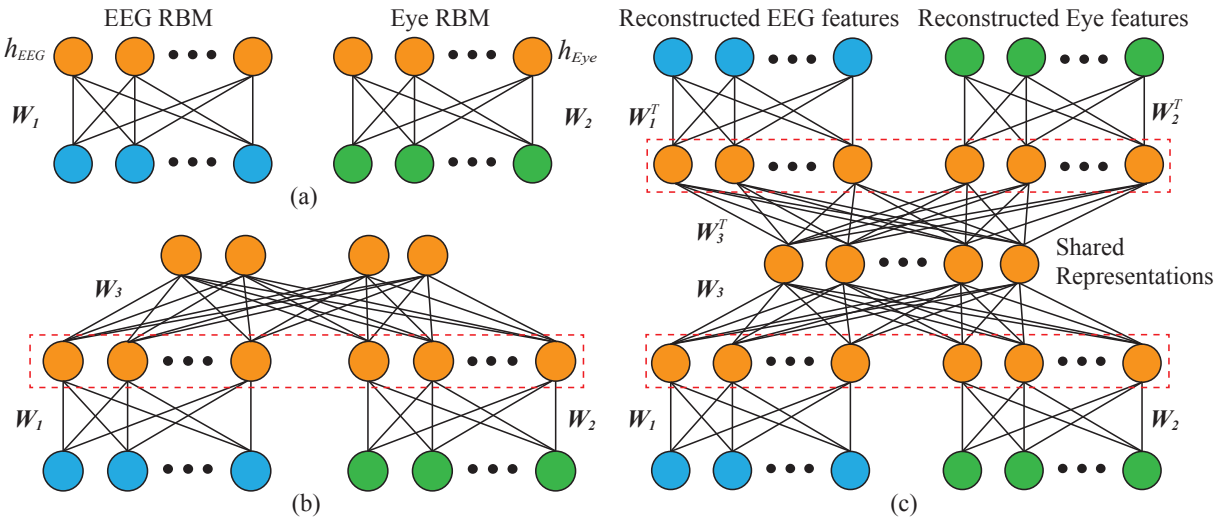


Figure 5–2 Multimodal deep neural network architecture for combining EEG and eye movements.

5.3 Modality Fusion Methods

Besides multimodal deep learning, we use different modality fusion strategies to fuse EEG and eye movements. There are mainly two types of fusion strategies: feature level fusion (FLF) and decision level fusion (DLF). The feature vectors of EEG and eye movements are directly concatenated into a larger feature vector as the inputs of classifiers for feature level fusion. For decision level fusion, individual classifiers are constructed with different modalities and the outputs of these classifiers are combined to obtain the final predictions. Maximal rule and sum rule are usually used in decision level fusion without training. Given the outputs of each classifier, the maximal (sum) rule is to compute the maximal (sum) values of all the probabilities of every category from all classifiers and predict the corresponding class labels that have the highest probability. Here, we used the linear SVMs as basic classifiers for comparison. The hyper-parameter C was tuned in the range of $2^{[}-10:10]$ with cross-validation.

The maximal rule and sum rule rely on the assumption that the individual classifiers from different modalities are mutually independent, which might not be satisfied in some situations. The maximal rule and sum rule can be regraded as the fixed weighted output of individual classifiers. We adopt more feasible metrics to determine the weights using advanced fusion strategy called fuzzy integral[241]. The fuzzy integral is integrals of a real function with regard to fuzzy measures.

Definition 1. A fuzzy measure μ defined on a finite index set $X = \{x_1, x_2, \ldots, x_n\}$ is a set

function $\mu : \mathcal{P}(X) \to [0,1]$ ($\mathcal{P}(X)$ is the power set of X) satisfying:

1. $\mu(\emptyset) = 0, \mu(X) = 1$,
2. $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$.

In this paper, we adopt the discrete Choquet integral[241].

Definition 2. Let μ be a fuzzy measure on X. The discrete Choquet integral of a function $f : X \to \mathbb{R}^+$ with respect to μ, is

$$\begin{aligned} &\mathcal{C}_\mu(f(x_1), f(x_2), \ldots, f(x_n)) \\ &:= \sum_{i=1}^{n} [f(x_{(i)}) - f(x_{(i-1)})]\mu(A_{(i)}), \end{aligned} \tag{5–1}$$

where $\cdot_{(i)}$ presents the permuted indices to satisfy $0 \leq f(x_{(1)}) \leq f(x_{(2)}) \leq \cdots \leq f(x_{(n)}) \leq 1$. Also $f(x_{(0)}) = 0$ and $A_{(i)} := \{x_{(i)}, x_{(i+1)}, \ldots, x_{(n)}\}$.

Let $C_1, C_2, \ldots, C_m$ be m classes and $X^T = [x_1 \ldots x_n]$ be a n-dimensional vector. There are n classifiers, one for each attribute x_i, which provide a confidence value denoted by $\Phi_i^j(X^\circ)$ for an unknown sample X° in the statement "X° belongs to class C_j", for all C_j.

To integrate all the confidence values of n classifiers, a fuzzy integral is used. The global confidence value in the statement "X° belongs to class C_j" is given by

$$\Phi_{\mu^j}(C_j; X^\circ) := \mathcal{C}_{\mu^j}(\Phi_1^j, \Phi_2^j, \ldots, \Phi_n^j), \tag{5–2}$$

where μ^j ($j \in \{1, 2, \ldots, m\}$) are defined on the set of attributes (or classifiers) and represent the importance of the classifiers. Ultimately, X° is predicted to be in the class with the highest confidence value.

The goal is to learn the fuzzy measure μ, which has $m(2^n - 2)$ coefficients. Suppose the number of classes is 2 (i.e. $m = 2$) for the sake of simplicity. Then there are $l = l_1 + l_2$ training examples labelled $X_1^j, X_2^j, \ldots, X_{l_j}^j$, $j = 1, 2$. We can compute μ by minimizing error J,

$$\begin{aligned} J = &\sum_{k=1}^{l_1} (\Phi_{\mu^1}(C_1; X_k^1) - \Phi_{\mu^2}(C_2; X_k^1) - 1)^2 \\ &+ \sum_{k=1}^{l_2} (\Phi_{\mu^2}(C_2; X_k^2) - \Phi_{\mu^1}(C_1; X_k^2) - 1)^2. \end{aligned} \tag{5–3}$$

This reduces to a quadratic optimization problem with $2(2^n - 2)$ variables and $2n(2^{n-1} - 1)$ constraints which can be written in the following form:

$$\begin{aligned} &\text{minimize } \frac{1}{2} u^T D u + \Gamma^T u \\ &\text{under the constraint } Au + b \geq 0, \end{aligned} \tag{5–4}$$

where u is a $2(2^n - 2)$ dimensional vector including all of the fuzzy measures μ^1, μ^2, i.e. $u := [u_1^T u_2^T]^T$, with

$$\begin{aligned} u_j := [&\mu^j(\{x_1\})\mu^j(\{x_2\}) \dots \mu^j(\{x_n\}) \\ &\mu^j(\{x_1, x_2\}) \dots \mu^j(\{x_{n-1}, x_n\}) \dots \\ &\mu^j(\{x_2, x_3, \dots, x_n\})]^T. \end{aligned} \tag{5–5}$$

We use the appropriate set of fuzzy measures to represent the importance of each classifier and the relative importance between any subset of the classifiers. In this way, the maximal and sum rules can be considered as fuzzy integral with certain fuzzy measures. Moreover, the fusion method with fuzzy integral can learn an optimized set of fuzzy measures according to different modalities.

5.4 Experimental Results on SEED for Three Emotions

In this Section, we present the experimental results on the subset of SEED dataset for recognizing three emotions: happy, sad, and neutral emotions. In SEED, there were only nine participants (5 females and 4 males) with both EEG and eye movement recordings. Therefore, we present the experimental results of these nine participants for multimodal emotion recognition. Every participant performed the experiments for three times in different days. There are in total 27 experiments evaluated in this study. The details about preprocessing and feature extraction of EEG and eye movements are demonstrated in Chapter 3. The performance are evaluated with the training data from the first 9 trials and the test data from the rest 6 trials of the same experiments.

5.4.1 Eye Movement-Based Emotion Recognition

We first study the relation between pupil responses and emotion. We extract the PSD and DE features in the four frequency bands of pupil diameter. The PSD and DE features of pupil diameter obtain the mean accuracies and standard deviations of 52.75%/21.59% and 57.40%/17.92%, respectively, for three emotions. The comparable performance indicate the discriminative information of pupil diameter for emotion recognition. We further explore the changes of pupil diameter under individual emotional conditions. Figure 5–3 shows the individual and average pupil diameters of each of nine participants for three emotion states. We find that sad emotion has the largest pupil diameter and follows by happy emotion. The neutral emotion has the smallest pupil diameters among these three emotions. These findings are consistent with previous studies in the literature[196]. The study of statistical significance with one way analysis

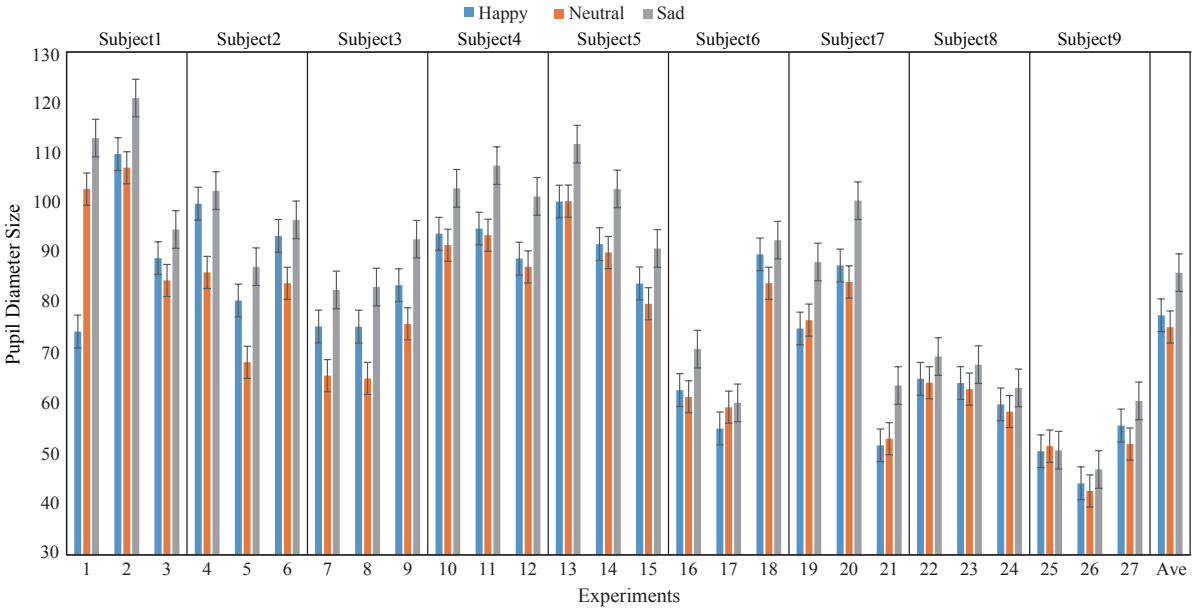


Figure 5–3 Individual and average pupil diameters of each of nine participants for three emotion states.

variance (ANOVA) indicates that the differences of pupil diameters are significant ($p < 0.05$). These results demonstrate that pupil diameter is a promising indicator for emotion recognition.

The other eye movement parameters such as dispersion (X and Y), fixation duration, blink duration, saccade and event statistics are also evaluated under individual emotional conditions. Except for blink duration, the differences between these parameters are significant ($p < 0.05$). Figure 5–4 presents the box plots of the four eye movement parameters under different emotions. The dispersion value is smaller for neutral emotion, while saccade amplitude and duration are larger in neutral emotion. Fixation duration is the lowest and the highest in happy and sad emotions, respectively.

We do not find the significant differences of blink duration under different emotions and the inclusion of blink duration features do not increase the average emotion recognition accuracies. Therefore, blink durations are not utilized in the further data analysis. The classification accuracies of individual eye movements and the concatenation of all features are shown in Table 5–1. All the individual eye movement features achieve accuracies higher than 45%, much above the random level 33.3% for three-class classification. These eye movement features contain discriminative properties for emotion recognition. The classification accuracies of event statistics, saccade and pupil diameter are the top three high among the features. The best average clas-

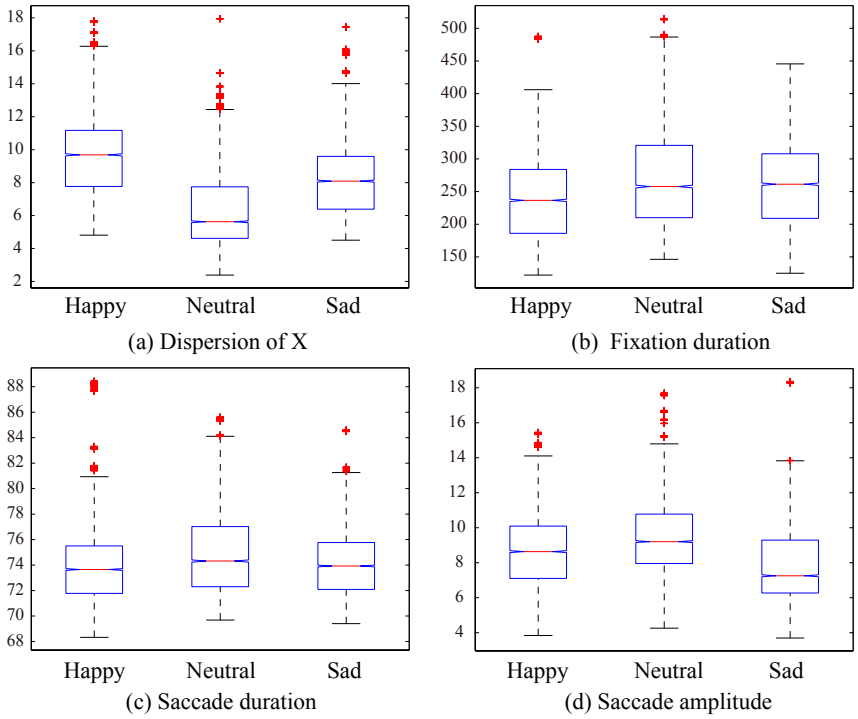


Figure 5–4 Box plots of four eye movement parameters (dispersion, fixation duration, saccade duration, and saccade amplitude) for three emotional states.

	PD	Dispersion	Fixation	Saccade	Event	FLF
Ave.	57.40	50.17	47.32	57.95	62.26	**77.80**
Std.	17.92	16.99	16.79	21.51	23.82	**14.61**

Table 5–1 Classification accuracies (%) of different eye movement features. ('FLF' means feature level fusion by combining all the eye movement features)

sification accuracy of the total 31-dimension feature achieves 77.80% (SD = 14.61%). These promising results suggest that these eye movement features can be used to discriminate different emotions efficiently.

5.4.2 Performance of Modality Fusion

The performance of EEG-based emotion recognition are shown in Table 5–2. The differential entropy features from the beta and gamma bands outperform the other EEG features. For the feature level fusion (FLF), the EEG features of the five individual frequency bands are concatenated directly. The DE features from the total frequency bands achieve the best performance with the mean accuracy of 78.51%. We further use the DE features and eye movement features

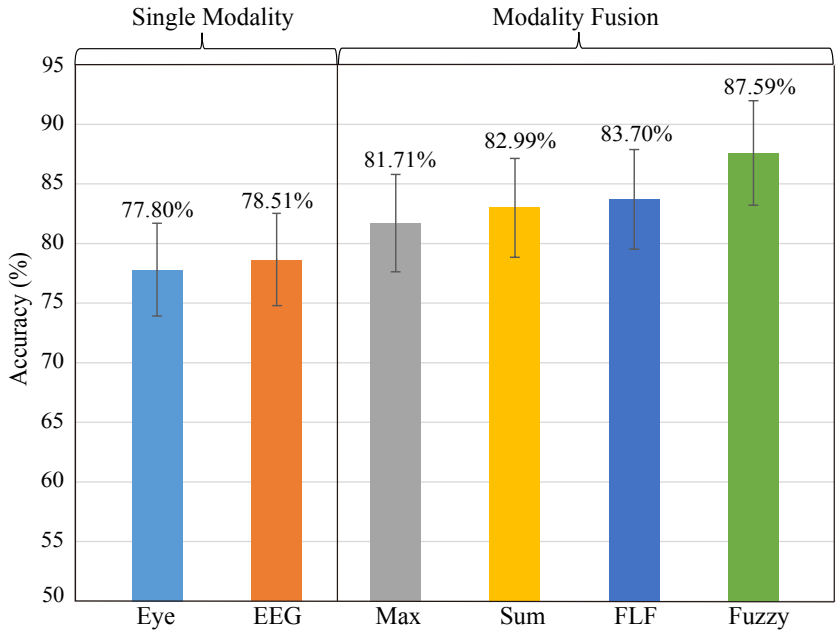


Figure 5–5 Performance of single modality and different modality fusion strategies on SEED.

to build multimodal emotion recognition models.

Feature		Delta	Theta	Alpha	Beta	Gamma	FLF
PSD	Ave.	60.62	60.18	61.69	**72.10**	69.46	65.35
	Std.	17.23	17.72	19.04	**15.50**	18.53	18.41
DE	Ave.	69.14	63.39	68.07	78.33	77.48	**78.51**
	Std.	14.45	16.94	16.94	**13.11**	16.60	14.32

Table 5–2 Classification accuracies (%) of different features and their feature level fusion from EEG.

After training two models of emotion recognition using eye movement and EEG signals, different modality fusion strategies are applied to improve the recognition performance. Modality fusion can combine important information in single modalities and improve the performance of emotion recognition system. Figure 5–5 presents the performance of single modality and different modality fusion methods. Our results show that modality fusion could significantly enhance the recognition accuracy compared with single modality. These two modalities contain complementary information and the combination can efficiently enhance the performance for emotion recognition. The best accuracy achieved by a fuzzy integral fusion strategy is 87.59%, whereas the accuracies of eye movement and EEG data are 77.80% and 78.51%, respectively. An about ten percents performance improvement is achieved. The average accuracy of FLF is ranked sec-

	Happy	Neutral	Sad
Happy	**0.81**	0.04	0.15
Neutral	0.09	**0.83**	0.08
Sad	0.17	0.15	**0.68**

(a) Eye Movements

	Happy	Neutral	Sad
Happy	**0.94**	0.02	0.04
Neutral	0.08	**0.77**	0.15
Sad	0.09	0.34	**0.57**

(b) EEG

	Happy	Neutral	Sad
Happy	**0.90**	0.02	0.08
Neutral	0.07	**0.84**	0.09
Sad	0.10	0.15	**0.75**

(c) FLF

	Happy	Neutral	Sad
Happy	**0.84**	0.03	0.13
Neutral	0.06	**0.87**	0.07
Sad	0.13	0.16	**0.71**

(d) Max Rule

	Happy	Neutral	Sad
Happy	**0.95**	0.02	0.03
Neutral	0.05	**0.88**	0.07
Sad	0.10	0.26	**0.64**

(e) Sum Rule

	Happy	Neutral	Sad
Happy	**0.96**	0.00	0.04
Neutral	0.10	**0.81**	0.09
Sad	0.02	0.13	**0.85**

(f) Fuzzy Integral

Table 5–3 Confusion matrices of single modality and different modality fusion strategies.

ond with 83.70%, followed by sum and max rules. This study paves the way for further investing potential applications of wearable devices utilizing eye movement and EEG signals to detect and recognize users' emotions for more excellent interactive experience.

5.4.3 Analysis of Complementary Characteristics

Confusion matrix is used in order to analyze the detailed complementary characteristics of EEG and eye movements for emotion recognition. Table 5–3 shows the confusion matrices of single modality (EEG and eye movements) and different modality fusion strategies. Each row of the confusion matrix represents the target class and each column represents the predicted class. The element (i, j) represents the percentage of samples in class i that is classified as class j. From the results, we can see that happy emotion can be well recognized with relatively high performance, while in contrast, sad emotion is the most difficult category with the lowest accuracy. Figure 5–6 depicts the confusion graph of eye movements and EEG, which shows their complementary characteristics for emotion recognition. For example, EEG modality achieves much higher accuracy than eye movements for happy emotion (94% vs. 81%), while it is difficult to classify neutral and sad emotions with the accuracies of 77% and 57%, respectively. For EEG, sad emotion is often confused with neutral emotion (34%) and vice versa (15%). In contrast, neutral and sad emotions can be well recognized with higher accuracies for eye movements (83% and 68%) in comparison with EEG (77% and 57%). Therefore, EEG and eye movements are complementary for recognizing happy, neutral, and sad emotions.

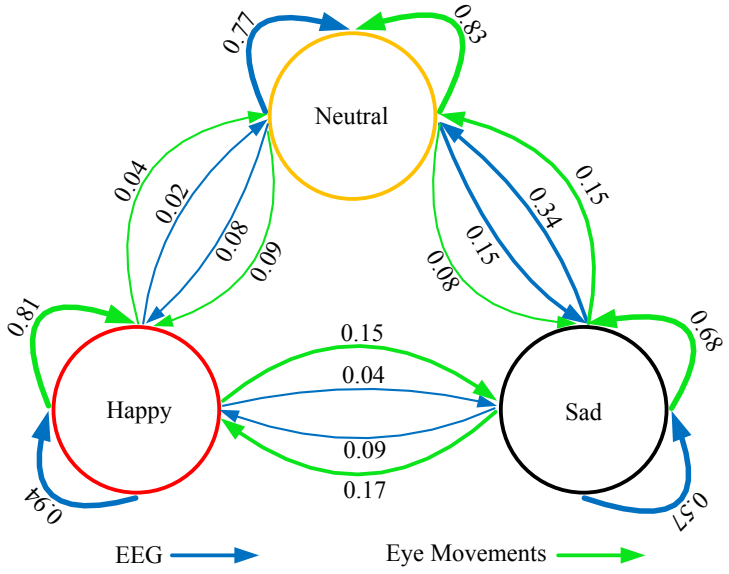


Figure 5–6 Confusion graph of eye movements and EEG, which shows their complementary characteristics for emotion recognition. (The numbers is the percentage of samples in class (arrow tail) that is classified as class (arrow head). Bolder lines mean higher values.)

Comparing the confusion matrices of single modality and different fusion methods in Table 5–3. The performance of multimodal approaches are usually better than unimodal approaches. The max and sum rules obtain moderate improvements for neutral emotion (87% and 88%, respectively), but still low accuracies for negative emotion (71% and 64%, respectively). FLF method has the advantage of recognizing sad emotion with 75% accuracy. Fuzzy integral method achieves the overall improvements for each emotion category, which are the best performance among these approaches. There are still 13% samples of sad emotion that are misclassified as neutral emotion.

In conclusion, multimodal approaches obtain an improvement of about 10 percent accuracy over unimodal method. Modality fusion is efficient for emotion recognition. EEG has the advantage of recognizing happy emotion compared with eye movements, while inferior performance for classifying neutral and sad emotions. EEG and eye movements leverage the favor of complementary information for enhancing the overall performance for emotion recognition. It should be noted that more progresses of multimodal emotion recognition on SEED have been achieved with recent developed methods, such as deep neural networks and bimodal LSTM[242, 243]. Our study suggests HCI designers to develop new applications in wearable devices utilizing eye movement and EEG signals to detect and recognize users' emotion states to make HCI systems

much more adaptable to emotion changes and the interactive experience much friendlier.

5.5 Experimental Results on SEED-IV for Four Emotions

In comparison with SEED, SEED-IV contains one more category of emotion and moreover, we focus on investigating how the performance varies with the electrode numbers in wearable requirements inspired by the previous findings about electrode reduction. SEED-IV consists of 45 experiments in total with 8 females and 7 males. We present modality fusion with multimodal deep neural networks significantly enhances the performance for classifying four emotions: happy, sad, fear, and neutral. We analyze the complementary characteristics of EEG and eye movements and the stability of emotion recognition models across sessions. Linear SVMs are used as basic classifiers for comparison. There are in total 24 trials for each experiment. For evaluations of single modality and modality fusion, the first 16 trials are used as the training data and the rest 8 trials of the same experiments (each emotion has two trials) are used as the test data. The fusion framework of EEG and eye movements using deep neural networks on SEED-IV is presented in Figure 5–7.

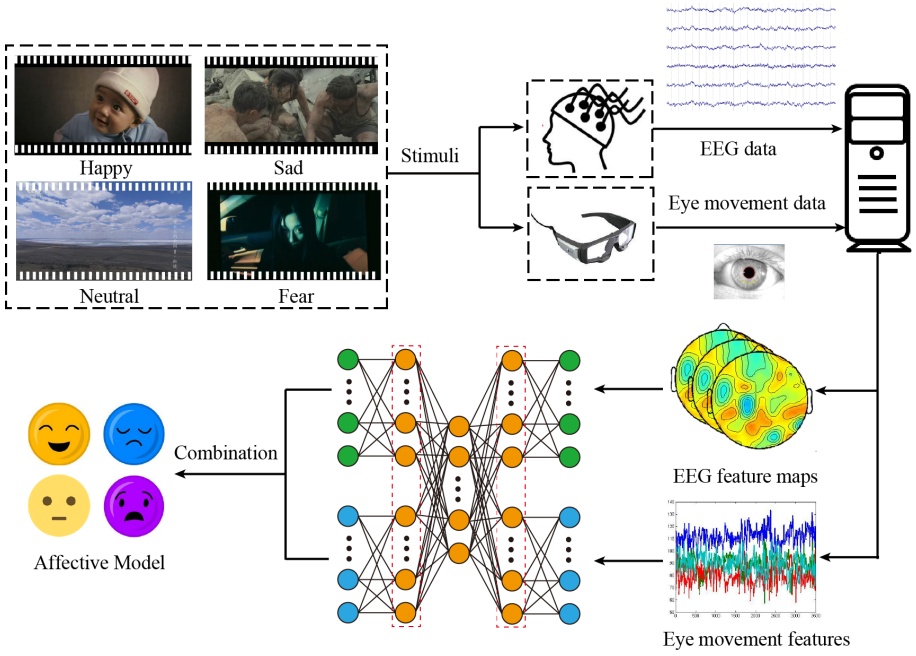


Figure 5–7 The fusion framework of EEG and eye movements using deep neural networks on SEED-IV.

5.5.1 EEG-Based Emotion Recognition

In this section, we evaluate the performance of different setups of EEG recordings. Inspired by previous findings of critical channels, we design three setups in the *EmotionMeter* framework: 1) T7 and T8; 2) T7, T8, FT7, and FT8; and 3) T7, T8, FT7, FT8, TP7, and TP8. Table 5–4 demonstrates the performance of different features with different electrode numbers. Linear SVM is used as a basic classifier for comparison. For four emotions, similar results are obtained in comparison with three emotions. The DE features are better than the PSD features and the higher frequency bands in the beta and gamma bands perform slightly better than the other bands in general. The visualization of the DE features and the corresponding emotion labels is presented in Figure 5–8. There are some dynamic neural patterns in high frequency bands with changes of emotion. The delta response does not show specific patterns, whereas the beta and gamma responses have consistent changes with the evolution of emotional states. The neural patterns of neutral and happy emotions have significantly higher beta and gamma activations than those of sad and fear emotions. The neutral emotion has higher alpha response among these emotions. Therefore, the high frequency bands in the alpha, beta, and gamma bands contain the most discriminative information. These neural patterns provide fundamental supports for developing affective brain-computer interactions using machine learning approaches.

For electrode reduction, we can see that the electrode setups with two, four, and six electrodes in the temporal areas can achieve relatively good performance for the four emotions. The mean accuracy of using only 6 electrodes is slightly lower than that of the original 62 electrodes (70.33% vs. 70.58%). The significant reduction of electrodes to two and four can still obtain comparable performance with the accuracies of 64.24% and 67.02%, respectively. As expected, the overall classification performance is better with more electrodes. But the computational complexity and calibration time are considerably increased as the cost and the improvement is limited. In real-world applications, it is attractive to reduce system setup and improve the comfortability with enough recognition accuracy and less device cost. Therefore, we propose to fuse EEG of only six electrodes and eye tracking to improve the feasibility and wearability of *EmotionMeter*.

5.5.2 Analysis of Modality Fusion and Complementary Characteristics

The mean accuracy and standard deviation of only eye movements are 67.82% and 18.04%, respectively, which are slightly worse than those of single EEG data (70.33%/14.45%). For modality fusion methods, we compare feature level fusion and multimodal deep neural networks.

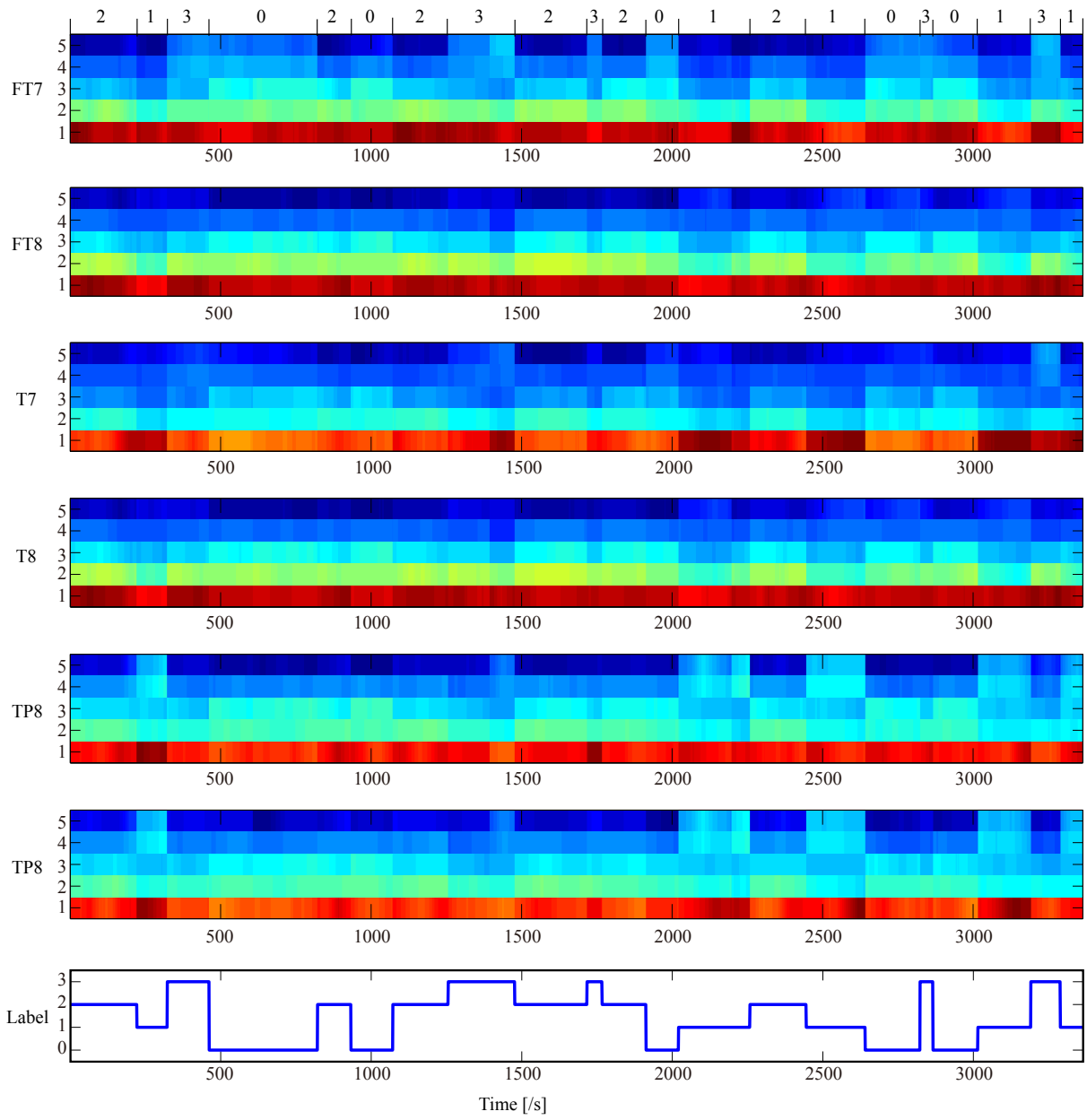


Figure 5–8 Visualization of the DE features and the corresponding labels in one experiment. Here, labels with 0, 1, 2, and 3 denote the ground truth, neutral, sad, fear and happy emotions, respectively. The numbers 1, 2, 3, 4, and 5 in the vertical coordinates respectively denote the five frequency bands: δ, θ, α, β, and γ. The dynamic neural patterns in high frequency bands (α, β, and γ) have consistent changes with the emotional labels during the whole experiment.

Table 5–4 The mean accuracy rates (%) of different electrode setups for the two different features obtained from the separate and total frequency bands.

#	Feature	Statistics	δ	θ	α	β	γ	Total
2	PSD	Mean	39.23	39.19	38.77	39.90	41.22	**60.22**
		Std.	8.36	8.00	7.35	9.15	10.77	16.18
	DE	Mean	36.84	37.18	39.25	39.23	41.21	**64.24**
		Std.	6.76	7.01	9.63	9.61	9.59	15.39
4	PSD	Mean	47.30	47.57	50.73	52.83	51.26	**58.98**
		Std.	12.29	12.22	15.65	13.26	15.04	17.85
	DE	Mean	45.21	47.99	50.65	54.41	57.19	**67.02**
		Std.	14.64	12.69	13.68	14.44	17.59	15.87
6	PSD	Mean	49.74	49.52	48.95	57.58	54.84	**57.70**
		Std.	17.48	13.47	15.66	20.92	18.46	16.94
	DE	Mean	50.47	48.88	51.99	60.70	60.57	**70.33**
		Std.	17.62	14.61	15.10	19.88	16.89	14.45
62	PSD	Mean	53.68	57.13	60.57	**63.60**	58.47	56.34
		Std.	13.40	14.23	15.54	18.78	18.60	14.54
	DE	Mean	57.58	57.98	61.22	66.66	66.34	**70.58**
		Std.	12.64	12.30	16.46	18.80	17.49	17.01

Figure 5–9 shows the performance of single modality with only EEG and eye movements and the two modality fusion methods. The performance with modality fusion is significantly greater than that with only a single modality ($p < 0.01$, one-way analysis of variance, ANOVA). The mean accuracies and standard deviations of feature fusion and multimodal deep neural networks are 75.88%/16.44% and 85.11%/11.79%, respectively. The multimodal approach with DNN achieves significant improvements compared with unimodal approaches (85.11% vs. 67.82% for eye movements and 85.11% vs. 70.33% for EEG).

Multimodal deep neural networks consist of three restricted Boltzmann machines. Two basic restricted Boltzmann machines are constructed based on individual modality and one restricted Boltzmann machines in the higher layer is built to combine the outputs of these two models. In this way, multimodal deep neural networks can learn the high-level shared representations between EEG and eye movements in comparison with direct feature fusion with shallow models. Efficient shared features are extracted with the processing of multiple layers. Feature fu-

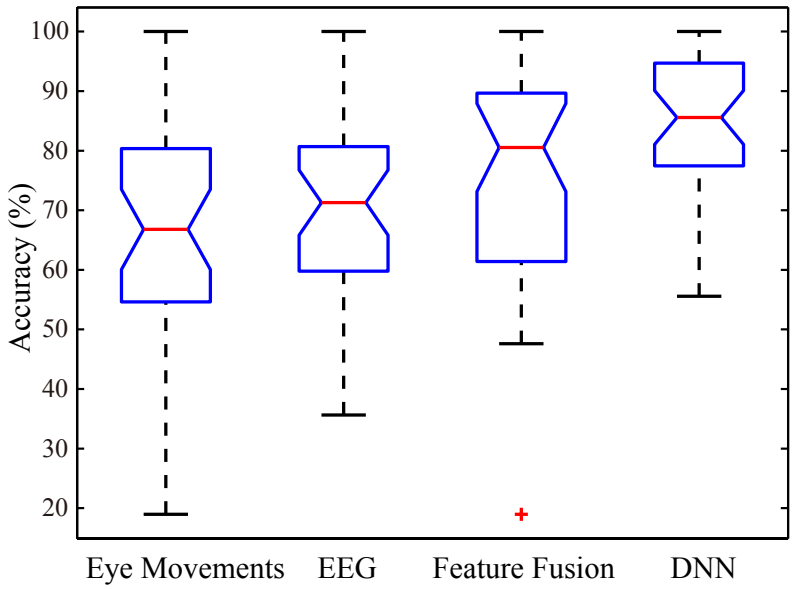
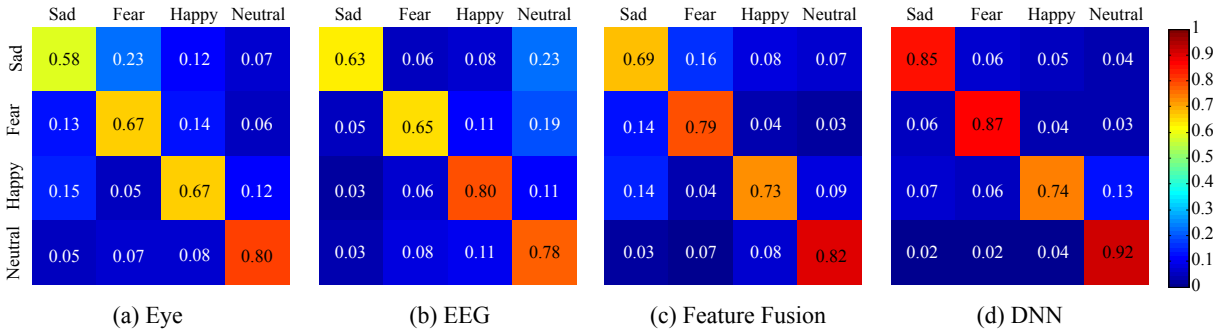


Figure 5–9 The box plot of the accuracies with each single modality (eye movements and EEG) and the two modality fusion approaches on SEED-IV. The red lines indicate the median accuracies.

sion approach lacks sufficient properties to capture the complexity and relations across various modalities. This method tends to learn unimodal features, rather than multimodal features. In general, the relations across modalities for the same targets are deep instead of shallow. With deep architectures, multimodal deep learning can capture these complex relations across various modalities and enhance the performance.

To analyze the complementary characteristics of EEG and eye movements for the four emotions (happy, sad, fear, and neutral), we summarize the confusion matrices individually to show their strength and weakness. Figures 5–10 and 5–11 present the confusion matrices and confusion graph of single modality and different modality fusion methods. In Figure 5–10, each row of the confusion matrices represents the target class and each column represents the predicted class. The element (i, j) is the percentage of samples in class i that is classified as class j. We still observe that EEG is superior to recognizing happy emotion in comparison with eye movements (80% vs. 67%). In contrast, eye movements have the advantage of recognizing fear emotion (67% vs. 65%). It is difficult to classify sad emotion for both modalities with the lowest accuracy among the four emotions. However, the misclassifications of sad emotion are different. More samples of sad emotion are classified as neutral emotion for EEG (23%), while as fear emotion for eye movements (23%). Both modalities obtain moderately high performance for neural emotion with the mean accuracies of 78% and 80%, respectively. From the above



(a) Eye (b) EEG (c) Feature Fusion (d) DNN

Figure 5–10 The confusion matrices of single modality and multimodal fusion methods on SEED-IV.

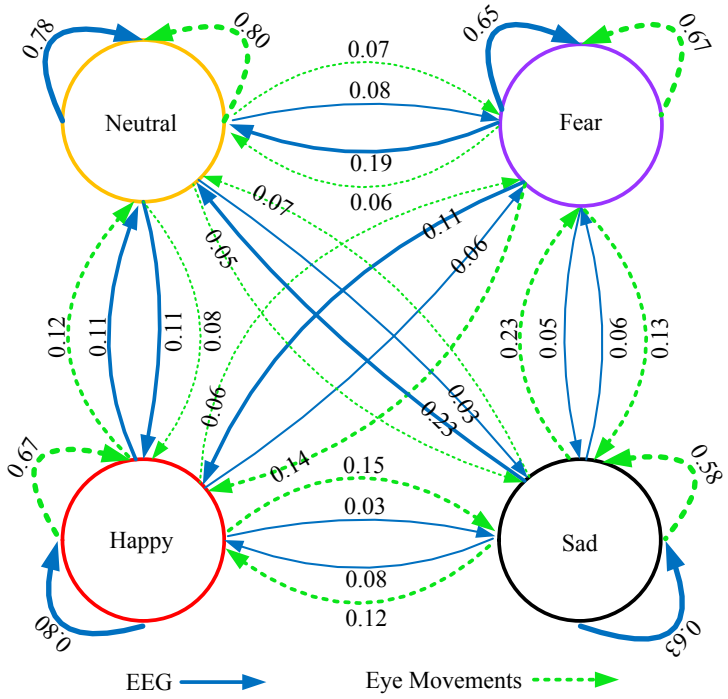


Figure 5–11 Confusion graph of EEG and eye movements on SEED-IV, which shows their complementary characteristics for emotion recognition. (The numbers denote the percentage values of samples in the class (arrow tail) classified as the class (arrow head). Bolder lines indicate higher values.)

experimental results, EEG and eye movements have different discriminative powers for emotion recognition. Therefore, the efficient combination of these two modalities can significantly improve the overall performance.

Comparing the confusion matrices of single modality and modality fusion in Figure 5–10, the feature level fusion achieves significantly increased accuracies of 6% and 12% for sad and

fear emotions, respectively. And multimodal deep neural networks obtain the best performance with much higher improvements of 22%, 20%, and 12% accuracies for sad, fear, and neutral emotions, respectively. The performance improvement of classifying sad emotion is significant for modality fusion in comparison with single modality. Both fusion methods have decreased performance for happy emotion, and the approach with sole EEG performs better. From the experimental results, the modality fusion can integrate the strength of EEG for recognizing happy emotion and the strength of eye movements for recognizing fear emotion. Meanwhile, the interactions between EEG and eye movements help improve the overall classification performance of sad and neutral emotion.

Human convey and interpret emotion through various modalities jointly and the brain combines these information together in parallel approaches for different tasks. Researchers have reached a consensus for constructing multimodal emotion recognition while concerning the fusion mechanisms of these multimodal information. The direct concatenations of available modalities to the training models are verified inefficient. The underlying fusion mechanisms of various input modalities and selective attentions for specific tasks are still an unsolved scientific problem. In the future, more modalities including audio-visual, physiological and contextual information can be efficiently combined with advanced fusion mechanisms.

5.5.3 Analysis of Stability Across Sessions

SEED-IV dataset also consists of three sessions in different days for the same participants for studying the stability across sessions. For evaluations, EEG and eye movement features from different sessions are split as the training and test datasets. Table presents the performance of cross sessions experiments with the two, four, and six electrodes used for EEG. With only six-electrode EEG data, multimodal DNN and feature fusion methods can achieve the mean classification accuracies of 72.39% and 59.52%, respectively, in the cross sessions experiments. Multimodal DNN significantly outperforms the feature fusion. Another interesting observation is that the performance is better with the training and test data from the nearer sessions. The relatively good performance across sessions indicate the stability of the proposed multimodal approach *EmotionMeter*.

5.6 Summary

In this Chapter, we introduce multimodal emotion recognition with EEG and eye movements using various modality fusion methods. Emotion is conveyed and interpreted via various modal-

ities including physiological signals and behaviors. Robust emotion recognition systems should consider how to efficiently combine different clues. We first study the relations between different eye movement parameters, such as pupil diameter, saccade duration, fixation duration and dispersion, and propose their individual patterns under different emotion conditions. We investigate the detailed complementary characteristics of EEG and eye movements for the three emotions in SEED and the four emotions in SEED-IV. The multimodal framework called *EmotionMeter* with deep neural networks is utilized to extract the high level shared representations across modalities. Modality fusion is superior to single modality in both datasets. The designed setups with the six electrodes above the ears improve the wearability and feasibility in real scenarios. The experimental results demonstrate the efficiency of the proposed *EmotionMeter*.

Moreover, we also observe the cross subject and cross session variations due to the nonstationary characteristics of EEG and changing environments such as noise, impedance variability, and the relative position of the electrodes. How to deal with these differences in feature distributions from different subjects and sessions is one of the popular topics in related research fields. In the next Chapter, we will introduce transfer learning or domain adaptation algorithms to overcome subject transfer problems. Therefore, we can recycle the previously labeled data in exiting datasets and reduce the calibration cost for target subjects.

Chapter 6

Personalizing Affective Models with Transfer Learning

In this Chapter, the application of transfer learning to affective brain-computer interactions is introduced. Individual differences across subjects and non-stationary characteristics of EEG limit the generalization of systems in real-world applications. A subfield in machine learning called transfer learning or domain adaption is shown as an promising approach for knowledge transfer between different domains. Section 6.1 illustrates the background and motivation of this study. Section 6.2 introduces several popular transfer learning algorithms. In Section 6.3, we present the detailed experimental setups for evaluations. And Section 6.4 shows the experimental results and discussions about performance. In Section 6.5, we focus on another topic about heterogeneous knowledge transfer between eye tracking and EEG. Spatiotemporal scanpath analysis is incorporated to help align EEG feature distributions from different subjects in the calibration phase. Section 6.6 summarizes the work in this Chapter. Some of the work in this Chapter have been published at International Joint Conference on Artificial Intelligence 2016[244].

6.1 Introduction

Most previous studies focus on constructing subject-specific models with efficient features and classifiers[105, 245]. Although the performance can be relatively high, these approaches usually need to collect a large number of labeled data from target subjects in the calibration phase. Therefore, is it possible to use the previously recorded labeled data for training the models of target subjects? A simple and straightforward approach is to directly train a general classifier with available labeled data from a group of subjects and make inference on the new data from target subjects. However, the performance of such general classifiers are usually degraded a lot due to individual differences across subjects and non-stationary characteristics of EEG[131, 246]. Even for the same subjects, there also exist the day-to-day variability across sessions. This issue refers to the covariate-shift problem in machine leaning field[247]. How to reduce the differences of feature distributions from different subjects with knowledge transfer is the main concern for personalizing EEG-based affective models across subjects.

To deal with this issue, various transfer learning or domain adaptation algorithms are pro-

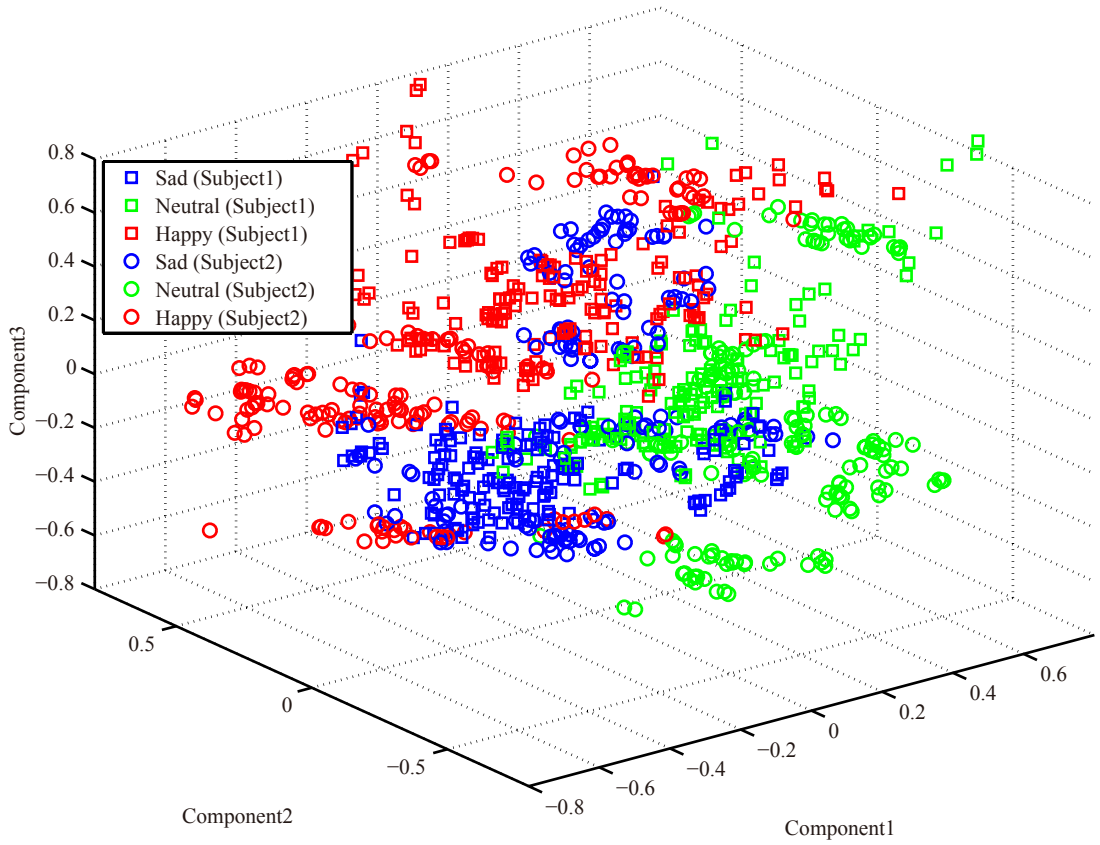


Figure 6–1 Illustration of the covariate-shift challenges of constructing EEG-based affective models.

posed[129]. Transfer learning aims to transfer knowledge between two different but relevant domains called source domain and target domain. Source domain usually has a large amount of labeled data, while target domain has few or no label data in inductive or transductive setups. In this work, source domain and target domain refer to the data from available subjects in the database and new subjects, respectively. Conventional machine learning methods are based on one prior assumption that the feature distributions of training and test data are in independently and identically distributed (i.i.d.). However, data from different domain might not satisfy this assumption, for example, the subject-to-subject variability in aBCIs. Figure 6–1 illustrates the covariate-shift challenges of constructing EEG-based affective models. The features of the same emotions from different subjects are distributed quite differently.

In this work, we adopt transfer learning algorithms for subject transfer in EEG-based emotion recognition, as shown in Figure 6–2. We assume that target subjects do not have any label samples in this case (the transductive setup). Let $X \in \mathcal{X}$ be the EEG recording of a sample (X, y), here $y \in \mathcal{Y}$ represents the corresponding emotion labels. In this case, $\mathcal{X} = \mathbb{R}^{C \times d}$,

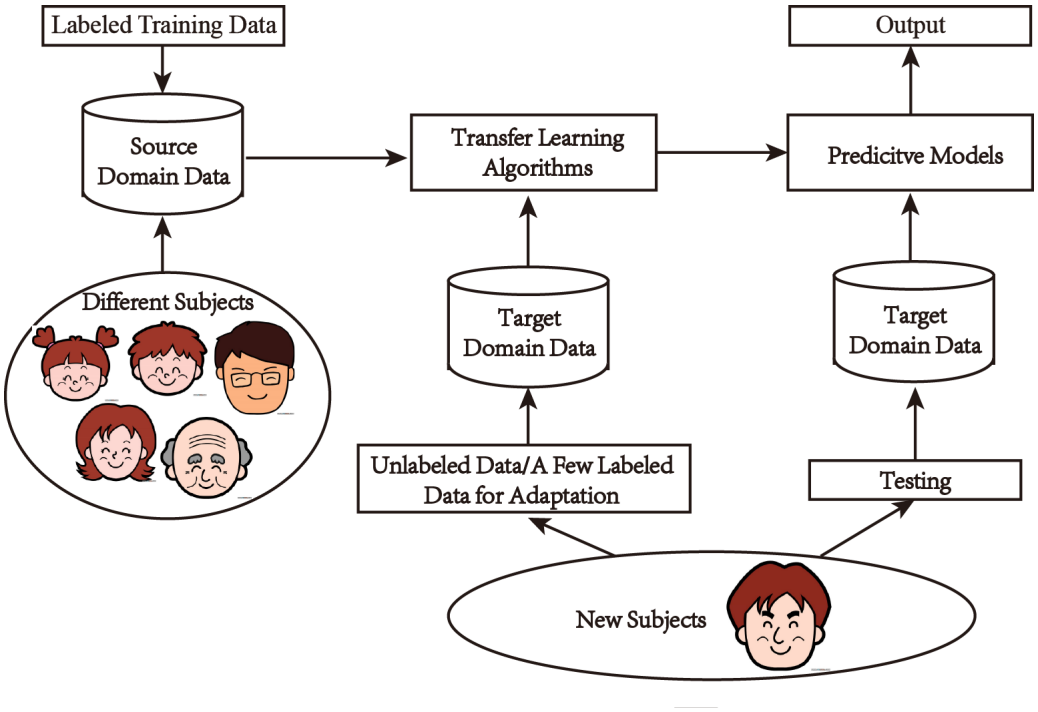


Figure 6–2 The subject transfer framework of EEG-based emotion recognition.

C is the number of channels, and d is the number of time series samples. Let $P(X)$ be the marginal probability distribution of X. According to the work of Pan and colleagues[129], $\mathcal{D} = \{\mathcal{X}, P(X)\}$ is a domain, which is EEG feature space and its marginal distribution from a given subject. In our case, the source and target domains contain the same feature space, $\mathcal{X}_S = \mathcal{X}_T$, but the respective marginal probability distributions are different, that is, $P(X_S) \neq P(X_T)$. In general, the conditional probability distributions of both domains are assumed the same, $P(Y_S|X_S) = P(Y_T|X_T)$.

The applications of transfer learning in brain-computer interfaces have drawn attention for tackling domain dependency challenges in recent years. For example, Jayaram and colleagues presented a survey about transfer learning in BCIs[130]. Many studies focus on developing the variants of common spacial patterns (CSP) for motor imagery tasks[246, 248]. Few studies have been performed in EEG-based emotion recognition.

In this study, we adopt two kinds of transfer learning methods for subject transfer in EEG-based emotion recognition. One is to find the shared common feature space between source and target domains. Transfer component analysis and kernel principle component analysis based methods are proposed by Pan and colleagues[249]. The basic idea is to reduce the differences between source and target domains and preserve the discriminative structures in the new feature

space. The other is called transductive parameter transfer[250], which is to measure the similarity of different feature distributions and learn the mapping from the feature distributions to classifier parameters. We evaluate the performance of different transfer learning algorithms on the public dataset, SEED, for three emotions (happy, neutral, and sad).

6.2 Transfer Learning

6.2.1 Transfer Component Analysis

Transfer Component Analysis (TCA) was firstly proposed by Pan and colleagues[249], which is the feature-based transfer learning algorithm. The main idea of TCA is to map the original feature space of source and target domains to the latent feature space spanned by a set of common transfer components. The domain differences are reduced and the original discriminative structure are preserved in the latent feature space. The feature dimension of the latent space is usually lower than that of original space. The formulation of TCA is to learn a transformation $\phi(\cdot)$ such that $P(\phi(X_S)) \approx P(\phi(X_T))$ and $P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$ without any labeled data in target domain (target subjects). An intuitive approach to find the mapping $\phi(\cdot)$ is to minimize the Maximum Mean Discrepancy (MMD)[251] between the empirical means of the two domains,

$$MMD(X_S', X_T') = ||\frac{1}{n_1}\sum_{i=1}^{n_1}\phi(x_i^s) - \frac{1}{n_2}\sum_{i=1}^{n_2}\phi(x_i^t)||_{\mathcal{H}}^2, \tag{6–1}$$

where n_1 and n_2 represent the sample numbers of source domain and target domain, respectively. However, $\phi(\cdot)$ is still very difficult to learn due to its highly nonlinear and many local minima[249].

The mapping $\phi(\cdot)$ reduces the original source and target data to a common low-dimensional latent space. TCA learns the transformation with the kernel matrix. Specially, let the Gram matrices defined on the source domain, target domain and cross-domain data in the embedded space be $K_{S,S}$, $K_{T,T}$, $K_{S,T}$, respectively. The kernel matrix K is defined on all the data as

$$K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}. \tag{6–2}$$

By virtue of kernel trick, the MMD distance can be rewritten as $tr(KL)$, where $K = [\phi(x_i)^\top\phi(x_j)]$, and $L_{ij} = 1/n_1^2$ if $x_i, x_j \in X_S$, else $L_{ij} = 1/n_2^2$ if $x_i, x_j \in X_T$, otherwise, $L_{ij} = -(1/n_1n_2)$. A matrix $\widetilde{W} \in \mathbb{R}^{(n_1+n_2)\times m}$ transforms the empirical kernel map K to an m-dimension space (where $m \ll n_1 + n_2$). The resultant kernel matrix is

$$\widetilde{K} = (KK^{-1/2}\widetilde{W}(\widetilde{W}^\top K^{-1/2}K)) = KWW^\top K, \tag{6–3}$$

where $W = K^{-1/2}\widetilde{W}$. With the definition of $\widetilde{K}$ in Eq.(6–3), the MMD distance between the empirical means of the two domain X'_S and X'_T can be rewritten as

$$Dist(X'_S, X'_T) = tr((KWW^\top K)L) = tr(W^\top KLKW). \tag{6–4}$$

A regularization term $tr(W^\top W)$ is usually added to control the complexity of W, while minimizing Eq.(6–4).

While reducing the domain differences, the mapping $\phi(\cdot)$ should preserve the discriminative structures for different categories, which can be modeled with data variance or scatter matrix. From Eq.(6–3), the variance of the projected samples is $W^\top KHKW$, where $H = I_{n_1+n_2} - (1/(n_1+n_2))\mathbf{1}\mathbf{1}^\top$ is the centering matrix, $\mathbf{1} \in \mathbb{R}^{n_1+n_2}$ is the column vector with all 1's, and $I_{n_1+n_2} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ is the identity matrix.

Therefore, the objective function of TCA can be formulated as

$$\begin{aligned} \min_{W} \quad & tr(W^\top KLKW) + \mu tr(W^\top W) \\ s.t. \quad & W^\top KHKW = I_m \end{aligned} \tag{6–5}$$

where $\mu > 0$ is a regularization parameter, and $I_m \in \mathbb{R}^{m\times n}$ is the identity matrix. The solutions W are the m leading eigenvectors of $(KLK+\mu I)^{-1}KHK$, where $m \leq n_1+n_2-1$. The algorithm of TCA for subject transfer is summarized in Algorithm 6–1. We recommend the readers to refer to the original reference[249] for the detailed descriptions of TCA. After obtaining the transformation matrix W, standard machine learning methods can be used in the new common feature subspace of both domains.

Algorithm 6–1 TCA-based Subject Transfer

input : Source domain data set $\mathcal{D}_S = \{(x_i^s, y_i^s)\}_{i=1}^{n_1}$, and target domain data set $\mathcal{D}_T = \{x_j^t\}_{j=1}^{n_2}$.
output : Transformation matrix W.

1: Compute kernel matrix K from $\{x_i^s\}_{i=1}^{n_1}$ and $\{x_j^t\}_{j=1}^{n_2}$.
2: Compute matrix L, and the centering matrix H.
3: Eigendecompose the matrix $(KLK+\mu I)^{-1}KLK$ and select the m leading eigenvectors to construct the transformation matrix W.
4: **return** tranformation matrix W.

6.2.2 Kernel Principle Component Analysis

In order to examine whether the performance improvements are made by dimensionality reduction, we compare the performance with KPCA-based subject transfer method. Kernel PCA is

the extended version of PCA with nonlinear dimensionality reduction using kernel methods[252]. A nonlinear transformation $\phi(x)$ transforms the original D-dimensional feature space to an M-dimensional feature space, where $M \gg D$. In the new space, the sample x_i is projected to a point $\phi(x_i)$.

The covariance matrix of the projected features can be calculated by $C = \frac{1}{N}\sum_{i=1}^{N}\phi(x_i)\phi(x_i)^\top$. Its eigenvalues and eigenvectors are given by $Cv_k = \lambda_k v_k$, where $k = 1, 2, \cdots, M$. Combining these two formulas, we have

$$\frac{1}{N}\sum_{i=1}^{N}\phi(x_i)\phi(x_i)^\top v_k = \lambda_k v_k, \tag{6–6}$$

All solutions v_k lie in the span of $\phi(x_1), \cdots, \phi(x_N)$. This implies that there exist coefficients $\alpha_1, \cdots, \alpha_N$ such that

$$v_k = \sum_{i=1}^{N}\alpha_{ki}\phi(x_i). \tag{6–7}$$

If we define the kernel function $\kappa(x_i, x_j) = \phi(x_i)^\top\phi(x_j)$, and combine Eq.(6–6) and Eq.(6–7), we have

$$\mathbf{K}^2\mathbf{a}_k = \lambda_k N\mathbf{K}\mathbf{a}_k, \tag{6–8}$$

where $\mathbf{K}_{i,j} = \kappa(x_i, x_j)$, and $\mathbf{a}_k$ is the N-dimensional column vector of α_{ki}. α_{ki}, which can be obtained by

$$\mathbf{K}\mathbf{a}_k = \lambda_k N\mathbf{a}_k. \tag{6–9}$$

Thus the kernel principal components can be calculated by

$$p_k(x) = \phi(x)^\top v_k = \sum_{i=1}^{N}\alpha_{ki}\kappa(x, x_i). \tag{6–10}$$

The Gram matrix $\widetilde{\mathbf{K}}$ is used to substitute the kernel matrix $\mathbf{K}$. The Gram matrix is given by

$$\widetilde{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N\mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N\mathbf{K}\mathbf{1}_N, \tag{6–11}$$

where $\mathbf{1}_N$ is the $N \times N$ matrix with all elements equal to $1/N$. The algorithm of KPCA for subject transfer is summarized in Algorithm 6–2.

6.2.3 Transductive Parameter Transfer

Transductive Parameter Transfer (TPT) is a parameter transfer method, which was firstly proposed by Sangineto and colleges for action units detection and spontaneous pain recognition[250].

Algorithm 6–2 KPCA-based Subject Transfer

input : Source domain data set $\mathcal{D}_S = \{(x_{S_i}, y_{srci})\}_{i=1}^{n1}$, and target domain data set $\mathcal{D}_T = \{x_{T_j}\}_{j=1}^{n2}$.

output : The kernel principal components p_k.

1: Concatenate the source and target domain data sets as the training data set, $\{x_i\}_{i=1}^{n1+n2} = [\{x_{S_i}\}_{i=1}^{n1}; \{x_{T_j}\}_{j=1}^{n2}]$.
2: Construct the kernel matrix $\mathbf{K}$ from the training data set $\{x_i\}_{i=1}^{n1+n2}$.
3: Compute the Gram matrix $\widetilde{\mathbf{K}}$ using Eq. (6–11).
4: Compute the vectors $\mathbf{a}_k$ using Eq. (6–9) (substitute $\mathbf{K}$ with $\widetilde{\mathbf{K}}$).
5: Compute the kernel principal components p_k using Eq. (6–10).
6: **Return** the kernel principal components p_k.

Figure 6–3 shows the TPT framework for subject transfer in EEG-based emotion recognition. TPT consists of three main steps: 1) multiple individual classifiers are learned on each training dataset D_i^s from different subjects; 2) a regression function is trained to learn the relation between the data distributions and classifiers' parameter vectors; 3) the target classifier is obtained using the target feature distribution and the distribution-to-classifier mapping. We will present the detailed descriptions of these three steps in the next part.

In the first step, individual classifiers are constructed on each source dataset D_i^s from different subjects using linear SVM. The objective function can be formulated as follows,

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + \lambda_L \sum_{j=1}^{n_i^s} l(\boldsymbol{w}'\boldsymbol{x}_j^s + b, y_j^s), \tag{6–12}$$

where $l(\cdot)$ is the hinge loss and $\boldsymbol{\theta}_i = [\boldsymbol{w}_i', b_i]$ represents the hyperplane parameters in the feature space.

In the second step, the aim is to learn the mapping from the data distributions to the optimal hyperplane based on the motivation of their relevance. A regression function f is learned for the mapping: $D \rightarrow \Theta$ with the source data distributions and their corresponding hyperplane parameters. Using the mapping, the hyperplane parameters of classifiers for new subjects can be predicted without any labeled data. There are many metrics for measuring the similarity between different data distributions. A kernel function $\kappa(X_i, X_j)$ is usually adopted to quantify the similarity between pairs of datasets X_i and X_j. In this implementation, we use two kinds of kernels: the density estimation (DE) based kernel[253] and Earth Mover's Distance (EMD) based

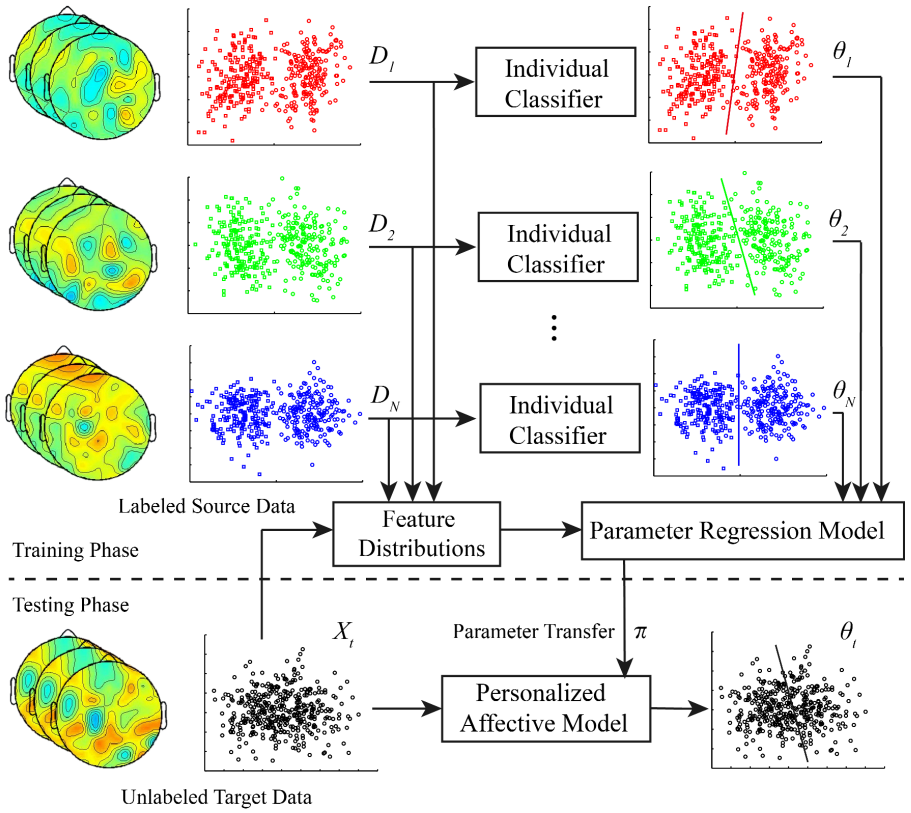


Figure 6–3 The framework of the transductive parameter transfer (TPT) approach for subject transfer in EEG-based emotion recognition.

kernel[254]. The density estimation based kernel is defined as follows,

$$\kappa(X_i, X_j) = \frac{1}{nm} \sum_{p=1}^{n} \sum_{q=1}^{m} \kappa_{\mathcal{X}}(\boldsymbol{x}_p, \boldsymbol{x}_q), \tag{6–13}$$

where n, m are cardinality of X_i, X_j, respectively, and $\kappa_{\mathcal{X}}(\cdot)$ is a Gaussian kernel.

To compute the EMD between X_i and X_j, the signatures of each set $\mathcal{I} = \{(\boldsymbol{v}_1^i, w_1^i), ..., (\boldsymbol{v}_Q^i, w_Q^i)\}$ and $\mathcal{J} = \{(\boldsymbol{v}_1^j, w_1^j), ..., (\boldsymbol{v}_Q^j, w_Q^j)\}$ are computed with K-means algorithm, where $\boldsymbol{v}_q^i$, $\boldsymbol{v}_q^j$ are cluster centroids from X_i and X_j, respectively, and w_q^i, w_q^j are the weights associated to each cluster. The number of clusters was set to five with cross-validation and the cardinality of each cluster was used as cluster weight. The EMD between X_i and X_j is defined as follows,

$$\begin{aligned} D_{EMD}(X_i, X_j) &= \min_{f_{pq} \geq 0} \sum_{p,q=1}^{Q} d_{pq} f_{pq}, \\ s.t. \sum_{p=1}^{Q} f_{pq} &= w_q^i, \sum_{q=1}^{Q} f_{pq} = w_p^j, \end{aligned} \tag{6–14}$$

where f_{pq} is the flow variable and d_{pq} is the ground distance defined as $d_{pq} = \| \boldsymbol{v}_p^i - \boldsymbol{v}_q^j \|$. Therefore, The EMD-based kernel is defined as follows,

$$\kappa_{EMD}(X_i, X_j) = e^{-\rho D_{EMD}(X_i, X_j)}, \tag{6–15}$$

where ρ is a tuning parameter. The detailed introduction of EMD is included in the references[250, 254]. After computing the mapping function f can be learned using the multioutput support vector regression method with the source data distributions and their corresponding hyperplane parameters.

Finally, The hyperplane parameters of target classifiers can be predicted by $\boldsymbol{\theta}_t = f(X^t)$ with the unlabeled data from target subjects and the mapping f of distribution-to-classifier. Given the new data $\boldsymbol{x}$ from target subjects and the predicted classifier parameters $\boldsymbol{\theta}$, the predictions can be obtained with the decision function: $y = sign(\boldsymbol{w}_t' \boldsymbol{x} + b_t)$. The algorithm of TPT-based subject transfer is summarized in Algorithm 6–3.

Algorithm 6–3 TPT-based Subject Transfer

input : Source domain data sets $\mathcal{D}_1^s, ..., \mathcal{D}_N^s$, target domain data set X^t, and some regularization parameters for SVMs.

output : The parameter vector of target classifier: $\boldsymbol{w}_t$, b_t.

1: Construct individual classifiers: $\{\boldsymbol{\theta_i} = (\boldsymbol{w}_i, b_i)\}_{i=1}^N$.
2: Create a training set $\mathcal{T} = \{X_i^s, \boldsymbol{\theta_i}\}_{i=1}^N$.
3: Compute the kernel matrix $\boldsymbol{K}$, $\boldsymbol{K}_{ij} = \kappa(X_i^s, X_j^s)$.
4: Given K and $\mathcal{T}$, learn $f(\cdot)$ using multioutput support vector regression.
5: Compute $(\boldsymbol{w}_t, b_t) = f(X^t)$
6: **Return** $\boldsymbol{w}_t, b_t$.

6.3 Experiment Setup

We perform the cross-subject evaluations on the SEED dataset for three emotions: happy, sad, and neutral emotions. SEED consists of 62-channel EEG data of 15 participants in total. The differential entropy (DE) features from the five frequency bands (δ: 1-3 Hz, θ: 4-7 Hz, α: 8-13 Hz, β: 14-30 Hz, and γ: 31-50 Hz) are extracted as the inputs of transfer learning models. The feature dimension is 310. Since the focus of the work is about the individual differences across subjects, we adopt a leave-one-subject-out cross validation method for evaluations. For each

experiment, EEG data from one subject are the test data (target domain) and the data from the rest 14 subjects are the training data (source domain).

As the baseline method, the direct concatenations of data from all available subjects are used to train linear SVMs as generic classifiers. One variant of SVM called Transductive SVM (T-SVM) is proposed to use both the unlabeled and labeled data to maximize the margin and learn the decision boundary[255]. Transductive SVM is implemented using the toolbox called SVM^{light}.

TCA and KPCA algorithms contain singular value decomposition, which cost large memory and time cost for calculating. Therefore, we choose to randomly select a subset of samples (about 5000) from all available subjects as the training data. We adopt the linear kernels for both TCA and KPCA because of their simplicity and relatively good performance in our relevant experiments. The performance with varying dimensionality of latent feature space are evaluated for TCA and KPCA. After transforming original features to new features in low-dimensional subspace with components learned by TCA and KPCA, standard linear SVMs are trained based on new features for subject transfer. We set the regularization parameter to 0.1 in TPT. All the algorithms are implemented in MATLAB programs. The one versus one strategy is used to tackle the multi-class classification and avoid label unbalanced problem.

6.4 Experiment Results

The performance of TCA and KPCA with varying dimensionality of the latent subspace are firstly evaluated. Figure 6–4 presents the accuracy curves of TCA and KPCA with respect to varying dimensions. TCA outperforms KPCA in the lower dimensions (smaller than 30) and achieves the best performance with the mean accuracy of 63.64% at 30 dimensions. In contrast, the classification accuracy of KPCA increases along with the increasing dimensions and it reaches the saturation at about 35 dimensions. Regarding the detection accuracy, the performance of TCA is lightly higher than KPCA (63.64% vs. 61.28%) with its explicit modeling both domain difference reduction and discriminative structure preservation. Figure 6–5 illustrates the visualizations of the EEG features of one experiment in a 3-dimension latent space using TCA and KPCA, respectively.

We compare the performance of different transfer learning methods for the task of subject transfer on SEED. Figure 6–6 presents the individual performance of the five transfer learning methods for each subject and Table 6–1 shows the corresponding mean accuracies and standard deviations. The baseline method without transfer learning performs the worst among the ap-

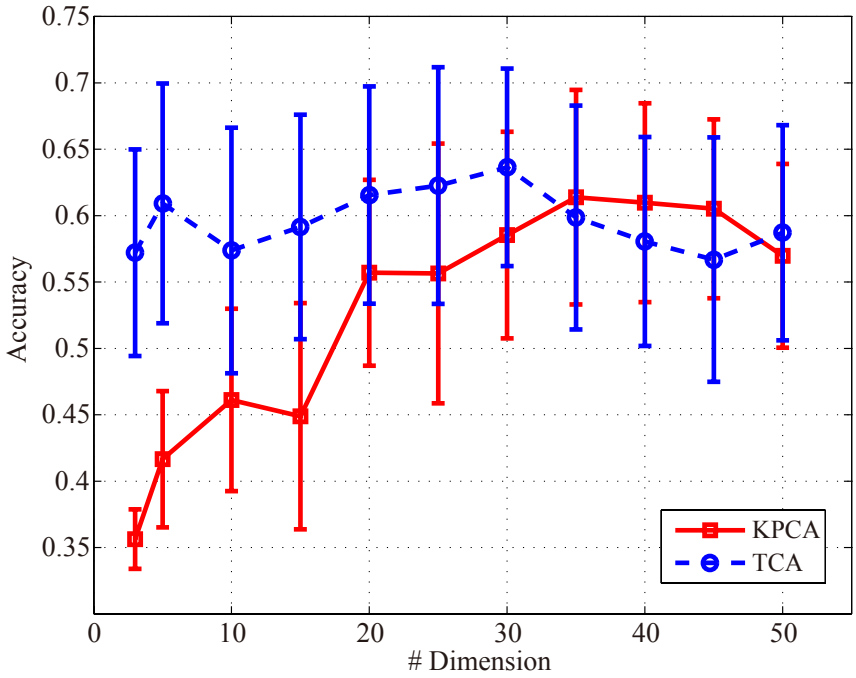
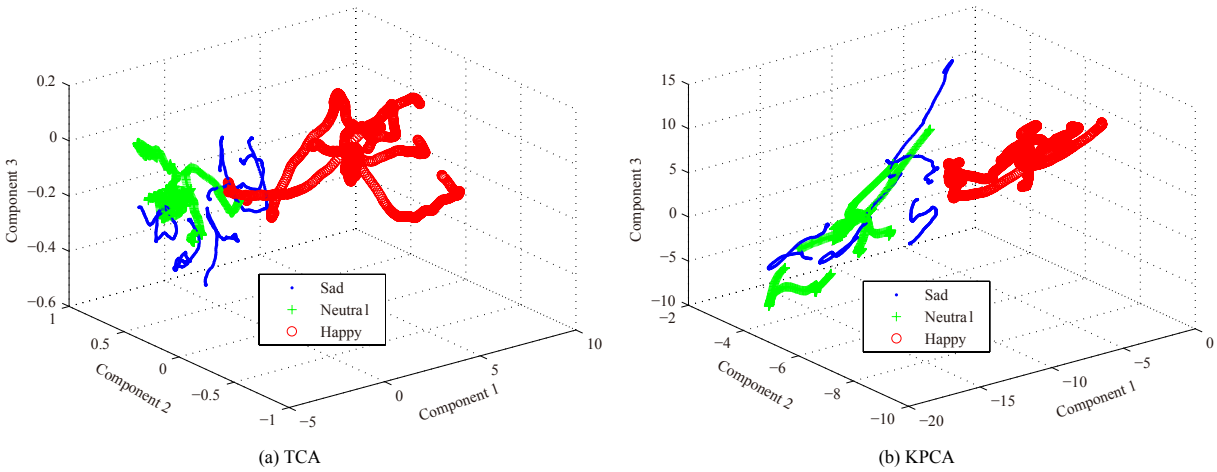


Figure 6–4 Comparison of KPCA and TCA approaches with varying dimensionality of the latent subspace.



(a) TCA (b) KPCA

Figure 6–5 Visualizations of the EEG features of one experiment in a 3-dimension latent space using TCA and KPCA, respectively.

proaches with the mean accuracy of only 56.73%. These generic classifiers do not consider the individual difference across subjects and use all the available source data for training. Some feature distributions of the source subjects might be quite different from those of target subjects. And these irrelevant factors bias the optimal hyperplane of classifiers and degrade the

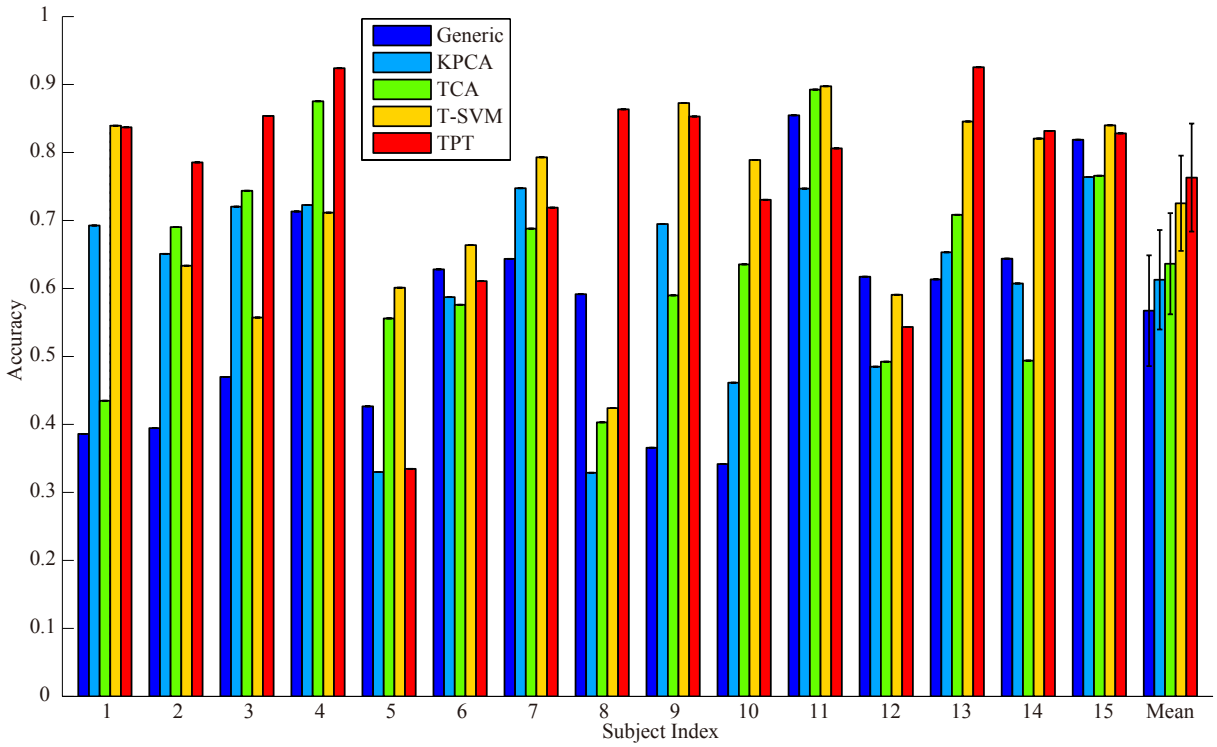


Figure 6–6 The accuracies of the five different transfer learning methods (Generic classifiers, KPCA, TCA, T-SVM, and TPT) for each subject.

Stats.	Generic	KPCA	TCA	T-SVM	TPT
Mean	0.5673	0.6128	0.6364	0.7253	**0.7631**
Std.	0.1629	0.1462	0.1488	**0.1400**	0.1589

Table 6–1 The mean accuracies and standard deviations of the five different transfer learning approaches.

performance of subject transfer dramatically. TCA and KPCA perform better than the generic method with explicit knowledge transfer. TCA and KPCA learn some transfer components underlying both the source and target domains. The domain differences are reduced while the data structures are also preserved at the same time in the latent subspaces spanned by the learned transfer components. T-SVM obtains the second best performance among all the methods with the mean accuracy of 72.53%. T-SVM extends standard SVM to learn the decision boundary with both the labeled and unlabeled data in a semi-supervised manner. Therefore, the hyperplane can be regularized adaptively to target data distributions but the iterations have high time cost for training.

TPT achieves the best performance among the transfer learning methods with the highest mean accuracy of 76.31%. The improvement in comparison with generic classifiers is significant (one way analysis of variance, $p < 0.01$). Different from the other transfer learning algorithms, TPT can deal with multiple source domains. In TPT, each source subject maintains one individual classifier. The mapping function from data distributions and classifier parameters is learned with the similarity measurements between different data distributions. Therefore, the relevant information from other subjects can be utilized to determine the decision function. Besides the detection accuracy, TPT is superior in feasibility to the other methods. KPCA and TCA learn the transfer components via singular value decomposition and find the low-dimensional subspaces underlying source and target domains. The limitation is the high memory requirements and time cost for training. With more source data inputting for training, the transfer components should be re-learned, which is not feasible for generalization. In contrast, TPT maintains individual classifiers for each subject in the source domain. It is incremental when more source data are available and the new kernel matrix is calculated. The new regression function can be learned based on the new kernels and previous classifier parameters. Therefore, TPT is more feasible in practice regarding both the classification performance across subjects and its incremental learning property.

Figure 6–7 presents the confusion matrices of the five transfer learning methods. Each row of the confusion matrix represents the target class and each column represents the predicted class. The element (i, j) is the percentage of samples in class i that is classified as class j. The generic method does not have any accuracy bias to some emotion detection. The mean accuracies of individual emotion are similar. TCA and KPCA achieve large improvements in detecting happy emotion with the mean accuracies of 75.88% and 66.23%, respectively. T-SVM improves the performance of recognizing both happy and neutral emotions with the mean accuracy of 71.52%. Among these methods, TPT obtains the highest accuracy of detecting all the three emotions. Similar to our previous findings, happy emotion can be well recognized with comparatively high accuracy of 85.01% using EEG. In contrast, sad emotion is often confused with neutral emotion (25.76%) and vice versa (10.24%). The differences of neural patterns between happy emotion and neutral or sad emotions are significant, while those of neutral and sad emotions are similar to each other. In summary, the experimental results demonstrate that TPT outperforms the other methods with an significant improvement of 19.58%. The TPT approach is efficient to deal with the individual differences across subjects and construct personalized EEG-based affective models.

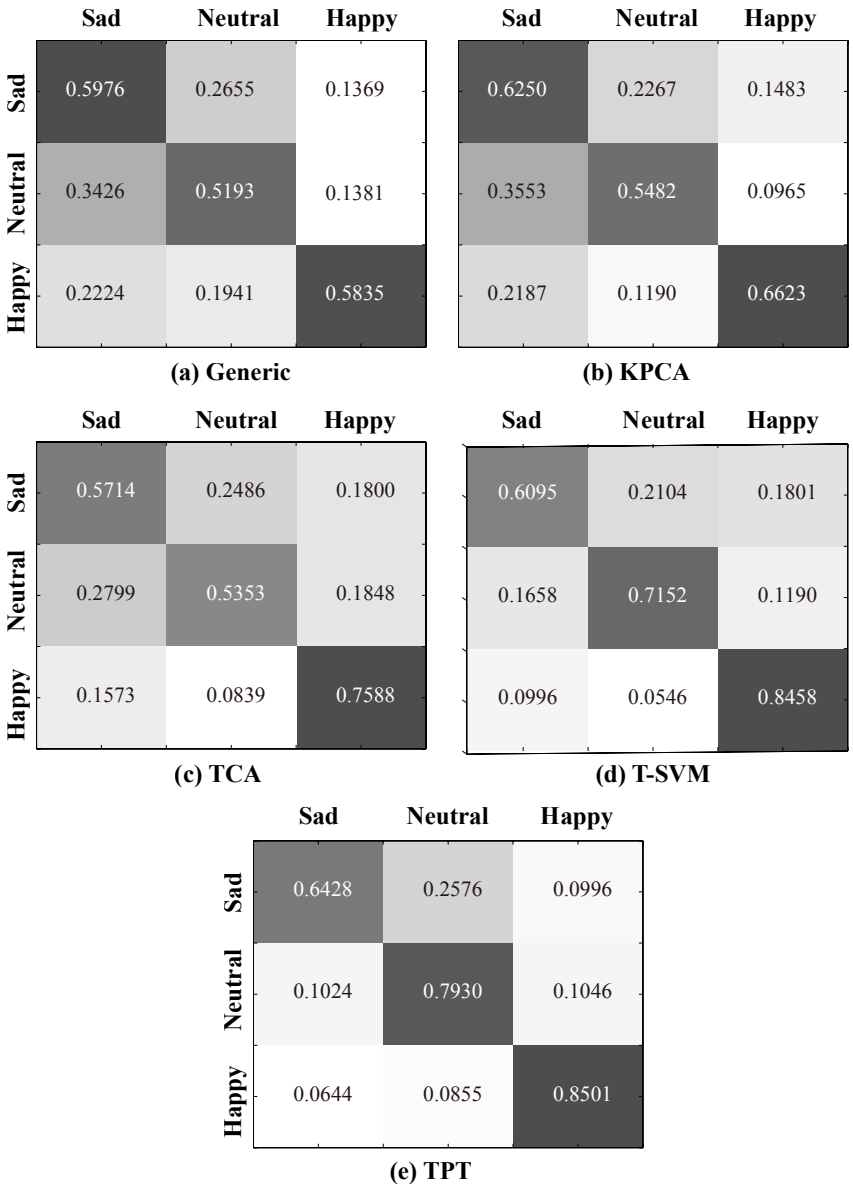


Figure 6–7 The confusion matrices of the five transfer learning methods.

6.5 Heterogeneous Knowledge Transfer From Eye Tracking To EEG

6.5.1 Introduction

The above experiments focus on homogeneous transfer learning, where the data modality is the same in source domain and target domain. In this framework, the knowledge transfer is solely based on EEG features. It is still necessary for users to acquire a moderately large amount of unlabeled EEG data from target subjects. In contrast, heterogeneous transfer learning aims

to transfers knowledge across different modalities. How to efficiently extract heterogeneous knowledge from different types of sources is still technically challenging. Most heterogeneous transfer learning methods focus on learning homogeneous feature representations based on some correspondences between domains[256].

In this section, we propose a novel approach to leveraging heterogeneous knowledge from spatiotemporal scanpath patterns to enhance subject transfer in building EEG-based affective models. The main ideas behind our approach are that what and where subjects are watching would evoke their specific neural activities in the brains and such useful information would provide important clues to emotion recognition. The key difference of our approach from the existing transfer learning paradigms is that the easily accessed eye tracking data from target subjects are utilized for subject transfer. Homogeneous transfer learning still needs time-consuming and high cost for collecting unlabeled EEG data. It is attractive for new subjects to wear an eye tracking glasses for calibration and improve the performance of affective models by reusing previous EEG from other subjects. Transferring affective knowledge from eye movements to EEG helps reduce calibration costs and obtain high recognition performance across subjects.

Figure 6–8 illustrates the framework of the proposed approach based on the transductive parameter transfer (TPT) algorithm. Here, individual classifiers on source subjects are constructed, EEG-based kernels and scanpath-based kernels are calculated, respectively, for quantifying the domain discrepancy, a regression function is trained to map the relationship between feature distributions and classifier parameters, and EEG-based and scanpath-based transfer models are combined using various decision-level fusion strategies to give the final predictions. In general aBCIs paradigms, film clips are often used as stimuli in the calibration phase. Therefore, although scanpath might be dependent on specific stimuli and impacted by perceptual processes (such as low-level, bottom-up saliency of the visual properties) and the comprehension tasks (understanding the scene), eye tracking can be utilized for qualifying individual differences across subjects and help align the EEG modality in the calibration phase with the same specific stimuli. For conventional aBCIs framework, the calibration phase requires recording moderately a large amount of labeled EEG data. In our framework (Heterogeneous Transfer shown in Figure 6–8), the calibration phase for new subjects can record only eye tracking data instead and transfer discriminative information of EEG from potential source subjects. It is feasible in some scenarios where collecting eye tracking data is much easier, while adaptive models are still able to recycle the representational capacity of EEG recorded previously for emotion recognition taking the advantage of high performance of EEG.

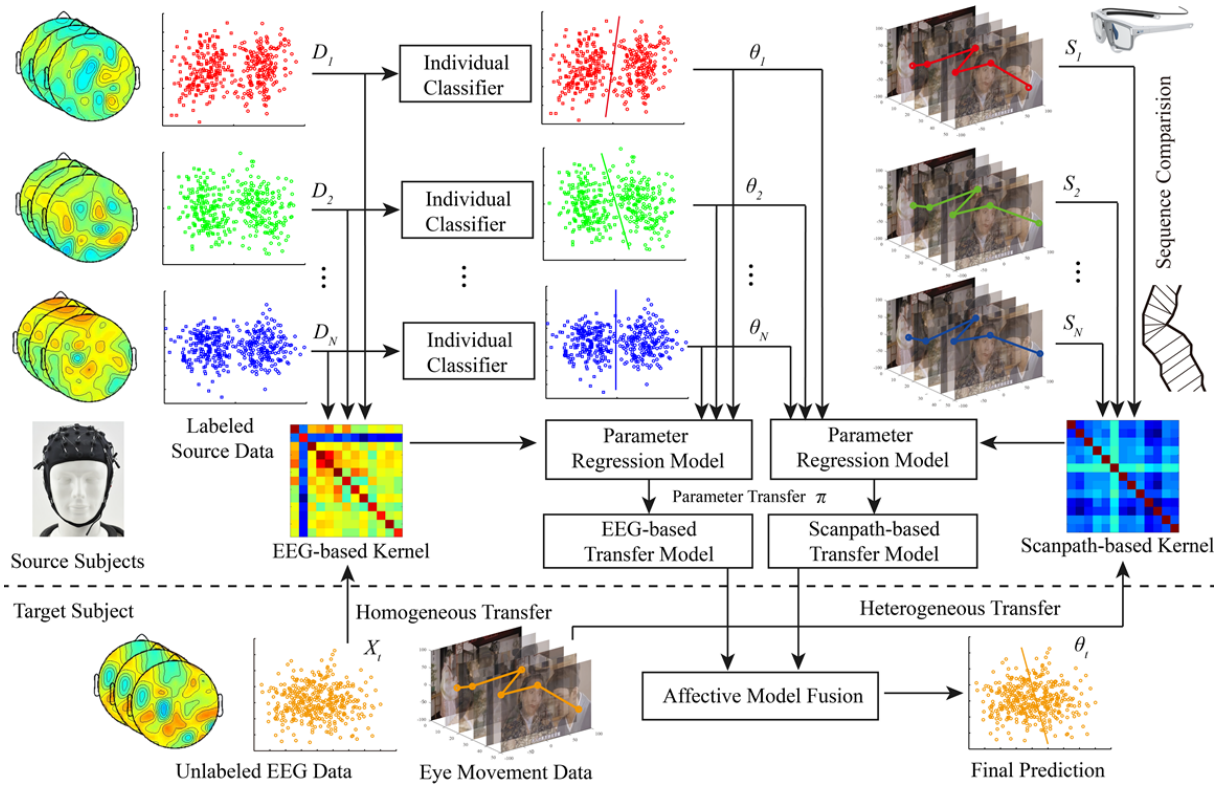


Figure 6–8 The framework of the proposed multimodal transductive parameter transfer for leveraging heterogeneous knowledge from spatiotemporal scanpath patterns to enhance subject transfer in EEG-based affective models.

Unlike the previous studies about multimodal emotion recognition using EEG and eye movements, we do not directly extract eye movement features for emotion recognition in this study. Instead, we construct and compare the spatiotemporal scanpath sequences under the same clips from different subjects. The similarity-based kernels are calculated as domain discrepancy to help align EEG feature distributions from different subjects and improve knowledge transfer in EEG feature space. Even there are only eye tracking data and not any EEG data for target subjects in calibration phase, the proposed approach can utilize the discriminative properties of EEG from other subjects. In contrast, all modalities should be available in both training and test phases. The existing transfer learning frameworks for BCIs have been conducted exclusively on only EEG data. To the best of our knowledge, this is the first study to introduce heterogeneous knowledge transfer with spatiotemporal scanpath analysis to enhance subject transfer for building EEG-based affective models.

6.5.2 Spatiotemporal Scanpath Analysis

In order to further quantitatively analyze spatiotemporal scanpath, scanpath sequences should be encoded in a computable formulation. The scanpath can be defined as temporally encoded locations, $\{(x_0, y_0, t_0), (x_1, y_1, t_1),..., (x_n, y_n, t_n)\}$, where x_i and y_i represent the fixation positions and t_i is the time stamp, $i = 0, 1, ..., n$. Figure 6–9 illustrates the encoding method for scanpath sequences. Firstly, we divided the scenes into 30×40 regions of interest (ROIs) and distribute each region with a pair of unique numbers. We encoded each fixation position x and y with the pair of numbers corresponding to that region and initialized a sequence with the order of fixation start time. To take the fixation duration into account, we introduced temporal binning into the sequences by repeating the pair of numbers corresponding to the position several times. The repeating times are proportional to the fixation duration. In our experiments, we set this temporal sampling to 50 ms, since the duration of a valid saccade is usually longer than 50 ms. Therefore, we repeated and inserted the pair of numbers into the sequences for each fixation with time windows of 50 ms. In this way, the sequence encoded from the eye movement data incorporates spatial location, sequential information, and temporal duration.

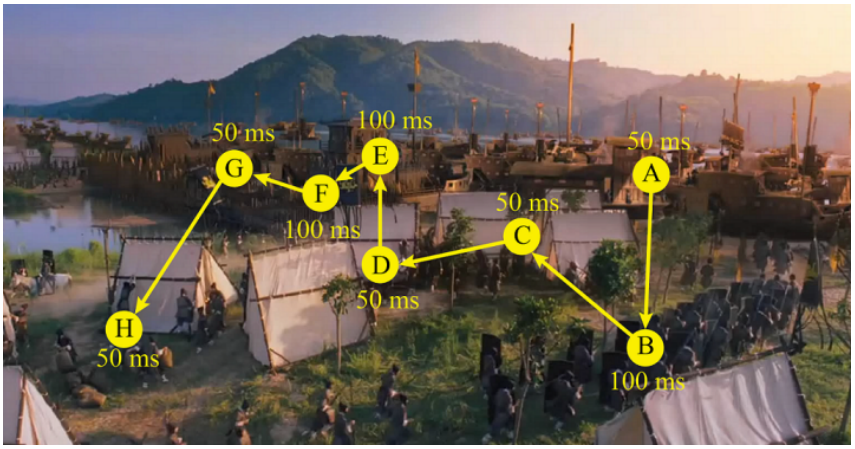


Figure 6–9 Illustration of scanpath sequence encoding. Here, the original sequence is *ABCDEFGH*. Taking fixation duration into account, the temporal binning is introduced and the original sequence is transformed into the following sequence: *ABBCDEEFFGH*.

The second step is to quantitatively measure how similar two scanpath sequences are. Scanpath sequences from different subjects are of different length even for the same clips. We adopt the Fast Dynamic Time Warping (DTW) algorithm to align different sequences with varying length. Dynamic Time Warping (DTW) can measure the similarity between two sequences with varying length and provides a quantitative measurement[257]. It calculates an optimal match

Algorithm 6–4 Scanpath-based Kernel Calculation

input : Raw eye movement data sets $F_1, ..., F_N$ from different subjects $1, 2, ...N$, with fixation positions, fixation durations and time stamps.

output : Scanpath-based kernel: $\boldsymbol{K}$.

1: Construct scanpath sequences: $S_i = Encoding(F_i)$.
2: Compare scanpath sequences: $\kappa(S_i, S_j)$ with Eq. (6–16).
3: Compute the kernel matrix $\boldsymbol{K}$: $\boldsymbol{K}_{ij} = \kappa(S_i, S_j)$.
4: **Return** $\boldsymbol{K}$.

between two given sequences with certain restrictions using dynamic programming with computational complexity of $O(N^2)$ in general. FastDTW recursively optimizes the DTW algorithm in a multilevel way from a coarser resolution to current resolution.

Based on the FastDTW, we calculate the distance for each pair of scanpath sequences from the same clips as a distance matrix for measuring the domain discrepancy. The scanpath-based kernel is defined as follows,

$$\kappa(S_i, S_j) = \frac{1}{k} \sum_{p=1}^{k} FastDTW(\boldsymbol{s}_{pi}, \boldsymbol{s}_{pj}), \tag{6–16}$$

where S_i and S_j are scanpath sequences from subject i and subject j, respectively, and k denotes the number of film clips. The algorithm of calculating the scanpath-based kernels is described in Algorithm 6–4.

6.5.3 Heterogeneous Transfer Learning

The transductive parameter transfer (TPT) algorithm is modified for heterogeneous transfer learning in this study. As our previous introduction about TPT in Section 6.2, A kernel function is defined to measure the differences of feature distributions across subjects and a regression function is trained to learn the relation between the data distributions and classifier parameters. The original TPT framework is based on homogeneous transfer learning, where only one data modality is available. We modify the TPT framework by incorporating both EEG-based kernels and scanpath-based kernels as domain discrepancy to personalize affective models. By using the scanpath-based kernel, we can transfer heterogenous information from eye tracking instead of EEG data from target subjects to construct EEG-based affective models in the calibration phase.

Algorithm 6–5 Multimodal TPT-based Subject Transfer

input : Source domain EEG data sets $\mathcal{D}_1^s, ..., \mathcal{D}_N^s$ and scanpath sequences $\mathcal{S}_1^s, ..., \mathcal{S}_N^s$, target domain EEG data set X^t and scanpath sequence S^t, and some regularization parameters for SVMs.

output : The parameter vector of target classifiers: $\boldsymbol{w}_t^1$, b_t^1,$\boldsymbol{w}_t^2$, b_t^2.

1: Construct individual classifiers: $\{\boldsymbol{\theta_i} = (\boldsymbol{w}_i, b_i)\}_{i=1}^N$.
2: Create training sets $\mathcal{T}^1 = \{X_i^s, \boldsymbol{\theta_i}\}_{i=1}^N$ and $\mathcal{T}^2 = \{S_i^s, \boldsymbol{\theta_i}\}_{i=1}^N$.
3: Compute EEG-based and scanpath-based kernel matrices $\boldsymbol{K}^1$ and $\boldsymbol{K}^2$, $\boldsymbol{K}_{ij}^1 = \kappa(X_i^s, X_j^s)$, $\boldsymbol{K}_{ij}^2 = \kappa(S_i^s, S_j^s)$.
4: Given $\boldsymbol{K}^1$,$\boldsymbol{K}^2$, $\mathcal{T}^1$ and $\mathcal{T}^2$, learn $f^1(\cdot)$ and $f^2(\cdot)$ using multioutput support vector regression.
5: Compute $(\boldsymbol{w}_t^1, b_t^1) = f^1(X^t)$ and $(\boldsymbol{w}_t^2, b_t^2) = f^2(S^t)$
6: **Return** $\boldsymbol{w}_t^1$, b_t^1, $\boldsymbol{w}_t^2$, b_t^2.

The detailed descriptions of TPT algorithm can refer to previous Section 6.2. In this framework, we use the density estimation (DE)-based kernel, Earth Mover's Distance (EMD)-based kernel with EEG modality for homogeneous transfer learning and scanpath-based kernel with eye tracking for heterogeneous transfer learning. Therefore, we can build up two subject transfer models using both EEG-based and scanpath-based kernel matrices. The algorithm of multimodal TPT-based subject transfer combining homogeneous and heterogeneous transfer learning is summarized in Algorithm 6–5.

After constructing two transfer models with EEG-based and scanpath-based kernel matrices, we adopt different decision level fusion strategies to combine them, including max rule, sum rule, and fuzzy integral algorithm. The discrete Choquet integral is used as the fuzzy measures in this study. After computing an appropriate set of fuzzy measures for different individuals in training datasets, we apply them to represent the importance of each classifier and the relative importance between any subset of the classifiers.

6.5.4 Evaluation Details

We conducted the performance evaluations on a subset of SEED. The experimental protocol was the same as that of SEED. Thirteen subjects (7 females and 6 males) with age range between 18 and 30 years participated in these experiments. The EEG and eye tracking data were simultaneously recorded and three emotions including happy, sad, and neutral emotions were included. For EEG processing, the differential entropy features were extracted from the five

frequency bands (δ: 1-3 Hz, θ: 4-7 Hz, α: 8-13 Hz, β: 14-30 Hz, and γ: 31-50 Hz) with a non-overlapping 1-s time window.

To evaluate the performance of subject transfer, we adopted a leave-one-subject-out cross validation method. Each time, we separated the data from one subject as the target domain and the resting data from other 12 subjects as the source domain. We directly concatenated the data from all available subjects as training data and train a generic classifier with linear SVM as the baseline method called *Generic*. We also selected the subset of all available subjects that have more similar scanpath patterns to that of target subject, which refers to *GenSub*. Through cross-validation, we found that the best subset number of source subjects is nine.

We compared our proposed subject transfer method to the existing transfer learning approaches that use EEG data *solely*, for example, transfer component analysis (TCA) and kernel PCA. The detailed descriptions of TCA and KPCA were included in previous Section 6.2. The main idea of TCA and KPCA is to learn transfer components across different domain and find a low-dimensional subspace, where domain differences are reduced and data structures are preserved. In this experiment, the dimensionality of the subspace in TCA and KPCA was tuned to 25 and 50 with cross-validation, respectively and both methods used linear kernels. Due to limits of memory and time cost, we randomly selected a subset of samples (5000 samples) from the rest 12 subjects as training data for TCA and KPCA.

Moreover, geodesic flow kernel (GFK) integrates an infinite number of subspaces that characterize changes in geometric and statistical properties between source and target domains[258]. Adaptation regularization based transfer learning (ARTL) optimizes the structural risk functional, the joint distribution matching between domains, and the manifold consistency[259]. The popular kernel mean matching algorithm (KMM) estimating density ratio between source and target domains is also included for comparison[260]. The one vs. one strategy was used to deal with the multi-class classification problems in the experiments.

6.5.5 Experimental Results

We compare the performance of subject transfer with the state-of-the-art transfer learning algorithms. Figure 6–10 presents the comparisons of various transfer learning for subject transfer in recognizing happy, sad, and neutral emotions. The first nine methods construct subject transfer models using EEG data solely. The baseline method, generic classifiers with SVM, achieves the mean accuracy and standard deviation of 50.46% and 17.14%, respectively. Among these methods, KMM performs the worst due to its inaccurate estimation of density ratio between source

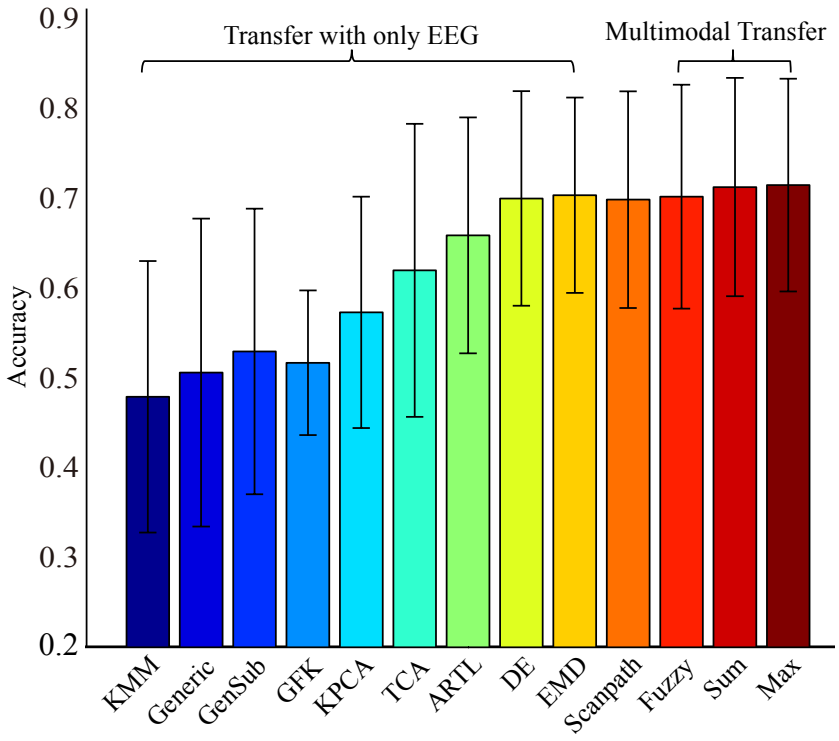


Figure 6–10 Classification accuracies of KMM, Generic, GenSub, GFK, KPCA, TCA, ARTL, DE, EMD, Scanpath, Fuzzy, Sum, and Max. The first nine methods construct subject transfer models using EEG data solely. '*Scanpath*' denotes TPT approach with scanpath-based kernels. The last three methods combine subject transfer models using EEG based and scanpath based kernels.

and target domains. According to the spatiotemporal scanpath comparison between pairs of subjects, we can find those subjects who have similar patterns with target subjects and use only these subset of data for training transfer models, referred to '*GenSub*'. How to efficiently select the right source data to enable effective knowledge transfer automatically is one of the critical techniques in transfer learning, since we usually have many choices of potential sources to use. Improper selections of source data might lead to negative transfer to some extend. Here, we select the potential source data for subject transfer based on the spatiotemporal scanpath patterns. From the experimental results, we find that *GenSub* performs slightly better than *Generic* with all source data. This result indicates that scanpath patterns contain effective information for selecting the right source data.

From Figure 6–10, we can see that TPT, TCA, and ARTL methods can significantly improve the performance of subject transfer with EEG data available only, while TPT performs the best among all the EEG-based transfer models. The mean accuracies and standard deviations of DE-

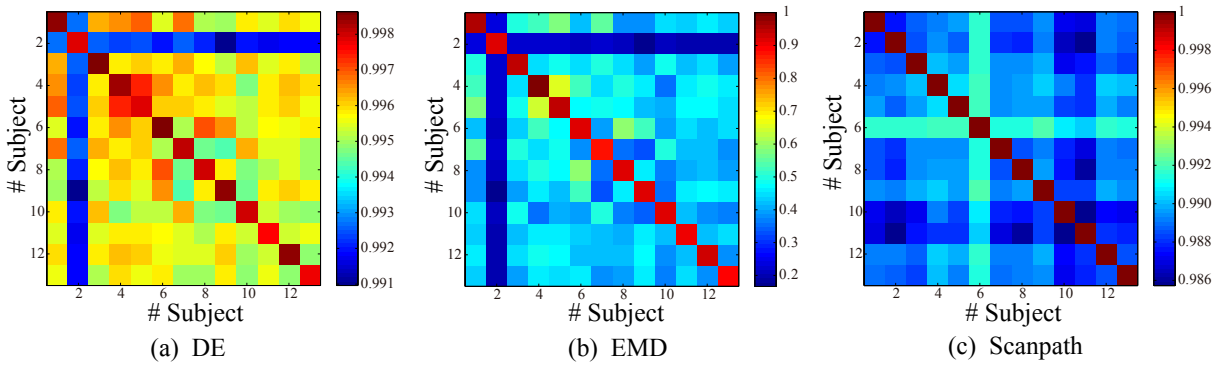


Figure 6–11 The DE-based, EMD-based, and scanpath-based kernels.

based kernel and EMD-based kernel for TPT are 69.85%/11.94% and 70.20%/10.87%, respectively. EMD-based kernel achieves slightly better accuracy than DE-based kernel. Compared to the generic classifiers (50.46%/17.14%), TPT algorithm achieves much higher performance, which shows its effectiveness for knowledge transfer between different domains.

By leveraging the heterogeneous knowledge from spatiotemporal scanpath patterns for subject transfer, the heterogeneous knowledge transfer ('*Scanpath*' in Figure 6–10) achieves comparable classification accuracy of 69.72%/12.07%. In comparison with generic classifiers, we achieve significant improvement of about 19.26% classification accuracy (one way analysis of variance, $p < 0.01$). More importantly, what is different from the above methods using homogeneous transfer learning is that the proposed approach does not need to acquire even unlabeled EEG data for target subjects in the training phase (the Heterogeneous Transfer condition shown in Figure 6–8). It is feasible in some scenarios where collecting eye tracking data is much easier, while adaptive models are still able to recycle the representational capacity of EEG recorded previously for emotion recognition. Figure 6–11 presents different kernels in the TPT framework for measuring the similarity between pairs of subjects. Higher values of the kernels indicate more similarity the patterns between subjects have. Different modalities are not always available for target subjects for calibration. For traditional transductive parameter transfer, the data modalities are the same for both source and target domains. In our case, there are not any EEG data for target subjects in calibration phase. We proposed the modified TPT framework for handling missing modality in target domain for calibration. The experimental results illustrate the effectiveness of our proposed subject transfer approach by leveraging the heterogeneous knowledge from spatiotemporal scanpath patterns from a novel aspect.

Modality fusion methods are utilized to combine different subject transfer models with EEG-based kernels and scanpath-based kernel. The last three bars in Figure 6–10 present the classification accuracies of Fuzzy Integral, Sum, and Max, respectively. Multimodal subject transfer can slightly improve the performance compared to TPT approaches, while Max strategy obtains the best mean accuracy of 71.33%. The slight improvement indicates that the discriminative information of EEG have been transferred to target subjects with the scanpath-based kernels. The modal generalization is based on EEG data, whereas eye movements help align EEG feature distributions across different subjects. Therefore, the scanpath-based subject transfer achieves comparable performance with EEG-based approach. These results demonstrate the superior performance of our approaches than those conventional methods trained solely on EEG features without taking other clues involved in affective processing into account. It is feasible in real scenarios where collecting eye tracking data is much easier, while adaptive models are still able to recycle the representational capacity of EEG recorded previously for emotion recognition taking the advantage of high performance of EEG.

6.6 Summary

In this Chapter, we focus on the challenge of subject transfer in affective brain-computer interaction. The individual differences across subjects and non-stationary characteristics limits the generalization of affective models across subjects and sessions. We proposed to personalize the EEG-based affective models with transfer learning algorithms. In Section 6.2, we introduce some transfer learning algorithms and the experimental results in Section 6.4 demonstrates the efficiency of transfer learning in tackling these problems. In previous parts of the Chapter, we focus on the homogeneous transfer learning with only EEG modality in both source and target domains. Section 6.5 describes the heterogeneous knowledge transfer from eye tracking to EEG with the proposed modified TPT framework. The scanpaths from different subjects are encoded and compared to measure the similarity of data distributions across subjects. In this way, the easily accessed eye tracking data can be used to align EEG feature distributions across different subjects in the calibration phase and the prediction models can still take the advantage of high performance of EEG for emotion recognition in the test phase. In conclusion, we present the efficiency of homogeneous transfer learning and heterogeneous transfer learning for building subject transfer affective models in affective brain-computer interactions.

Chapter 7

Multimodal Vigilance Estimation: From Simulated To Real Scenarios

In this Chapter, we present mulitmodal vigilance estimation with EEG and forehead EOG in both simulated and real driving environments. Section 7.1 introduces the research background and the motivation of vigilance estimation. The development of relevant vigilance estimation datasets have been described in Chapter 3. In Section 7.2, the detailed processing and feature extraction of EEG and forehead EOG are described. We consider the problem of vigilance estimation as continuous regression, rather than discrete classification. Since vigilance states are dynamic changing process with temporal evolution, we consider the temporal dependency between samples for robust vigilance estimation in Section 7.3. The evaluation metrics of model performance are presented in Section 7.4. SEED-VIG is the public multimodal vigilance estimation data set developed in this study, which consists of EEG and EOG data recorded by commercial wet electrodes. The experimental results, including single EEG, EOG, multimodal fusion, and their complementary characteristics, are demonstrated in Section 7.5. In order to improve the wearability and feasibility of vigilance estimation, we develop a wearable device with flexible dry electrodes and amplifier circuits for forehead EOG recording. We evaluate the performance in both laboratory simulated driving and real-world driving experiments. The experimental results on the wearable devices are described in Section 7.7. In Section 7.8, the limitations of current work and some challenges as future work are discussed. Section 7.9 summarizes the study of the Chapter. Some of the work in this Chapter have been published at Journal of Neural Engineering[186].

7.1 Introduction

Besides emotional states, vigilance states are another important components of human mental states. We concentrate our focus on current particular tasks with selective attention, even with various stimuli from the world outside simultaneously coming in. The recognition of human mental states including emotion and vigilance can help improve user-aware human-computer

interactions with adaptability, rather than in a rule-based way. Zander and Kothe proposed to combine conventional BCI systems with cognitive monitoring as passive brain-computer interfaces[25]. In our framework of affective brain-computer interactions, emotion and vigilance are two vital components to promote active interactions between users and machines. Vigilance refers to the ability to endogenously maintain attention.

Various tasks require sustained high vigilance, especially for driving scenarios. Driving fatigue is reported as one of the major factors in fatal road accidents. Therefore, active driving safety technologies are of great interests in both research and industry communities. Considerable progresses about vigilance estimation have been achieved over the past decades, as we review in Chapter 2. Various approaches with external behaviors such as head nodding, yawning and eye closure, and internal physiological signals have been proposed and evaluated. However, there are still many challenges for robust vigilance estimation in real-world applications due to its property of intrinsic mental states with temporal evolution rather than static states. The usage of intrusive stimuli and behavior probes is not feasible in real scenarios. In order to improve the feasibility, interdisciplinary studies about sensors, circuit design, and computational models should be conducted.

Previous findings support the relations between EEG and EOG and vigilance fluctuations. The combination of brain activities from EEG and behaviors from EOG is shown as an efficient approach for multimodal vigilance estimation. Most proposed systems are evaluated in laboratory simulated environments in the literature. There is a reality gap between simulations and real environments. In this study, we aim to bridge the gap from simulations to real environments by developing wearable vigilance estimation systems with EEG and forehead EOG. We firstly evaluate the performance of our computational models including preprocessing, feature extraction, regression models with data recorded by commercial devices in simulated environments. A wearable prototype device with both flexible dry electrodes and an acquisition board is further designed and evaluated in both simulated and real-world driving environments under different weather conditions. In this way, the applications of vigilance estimation systems from laboratory simulations to real-world scenarios are promoted.

Figure 7–1 shows the flowchart and structure of our proposed wearable vigilance estimation system, which includes several parts: flexible dry electrodes, EOG acquisition board, feature extraction for forehead EOG, regression models, and vigilance annotations. The experimental setups for both laboratory simulations and real-world environments are also illustrated in Figure 7–1. The subjects are asked to drive a car in the laboratory simulations and real-world environ-

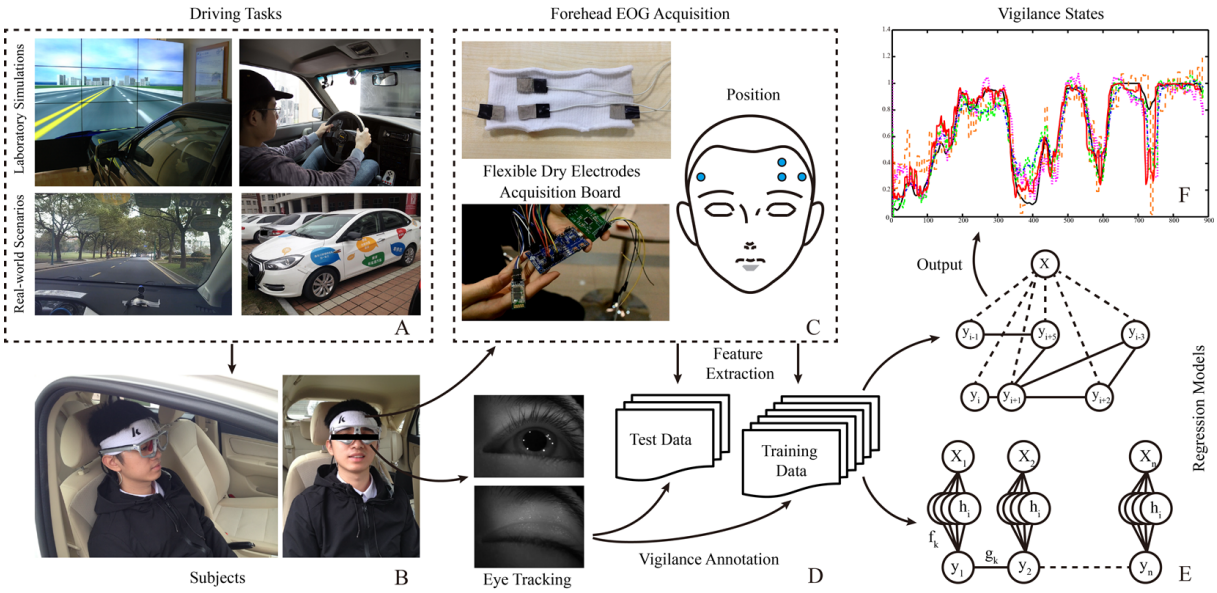


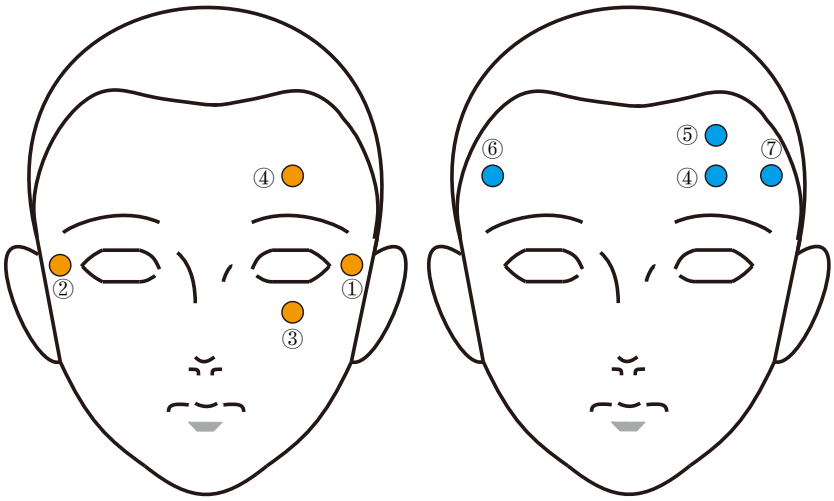
Figure 7–1 The flowchart and structure of our proposed wearable vigilance estimation system.

ments (A, B). The simulated driving experiments are performed in a real vehicle without the unnecessary engine. The driving scenes are synchronously updated according to the subjects' operations. The real-world driving experiments are performed under various weather conditions. The forehead EOGs and eye movements are simultaneously recorded using our designed wearable EOG-recording prototype and SMI eye-tracking glasses, respectively (C, D). Regression models using CCRF and CCNF are trained for continuous vigilance estimation (E, F). The detailed processing and evaluations are described in the following sections.

7.2 Feature Extraction

7.2.1 Preprocessing for Forehead EOG

Conventional EOG setups mount several electrodes around the eyes, which distract users with discomfort. In contrast, we propose a novel electrode placement for forehead EOG recordings. The design is shown more feasible for implementing wearable devices. The comparison of conventional EOG setup and our proposed forehead setup is shown in Figure 7–2. The reference and ground electrodes are placed at the left and right mastoids, respectively. Both setups consist of four electrodes for signal recording. In the EOG signal processing, the vertical EOG (VEO) and horizontal EOG (HEO) are separated by subtracting electrodes four and three and electrodes



(a) Traditional EOG (b) Forehead EOG

Figure 7–2 The electrode placements for the traditional and forehead EOG setups in vigilance estimation experiments. Electrode four is the shared electrode of both setups.

one and two, respectively. VEO and HEO provide various eye movement parameters, such as blink, and saccade, respectively.

Therefore, how to separate VEO and HEO signals from our forehead EOG setup is the first important step. We use two separation methods: the minus rule and independent component analysis (ICA). VEO_f signals are extracted from electrodes numbered four and five and HEO_f signals are extracted from electrodes seven and six. For the minus rule, the subtraction of channels four and five is an approximation of VEO, named VEO_f, and the subtraction of channels seven and six is an approximation of HEO, named HEO_f. Here, the subscript 'f' indicates 'forehead'.

ICA is a popular blind source separation algorithm to decompose a multivariate signal into independent non-Gaussian signals. The ICA algorithm decomposes the multichannel data into a sum of independent components. We have assumption that the VEO and HEO components are from different sources and the signal waves between them are largely different. We extract the VEO_f and HEO_f components using FASTICA[261] from channels four and five and channels six and seven, respectively. The comparison of the traditional EOG and forehead EOG using the minus operation and ICA separation strategies is depicted in Figure 7–3. As shown, the extracted VEO_f and HEO_f from the forehead electrodes have similar waves to the traditional ones, and the forehead VEO_f and HEO_f can capture critical eye movements, such as blinks and

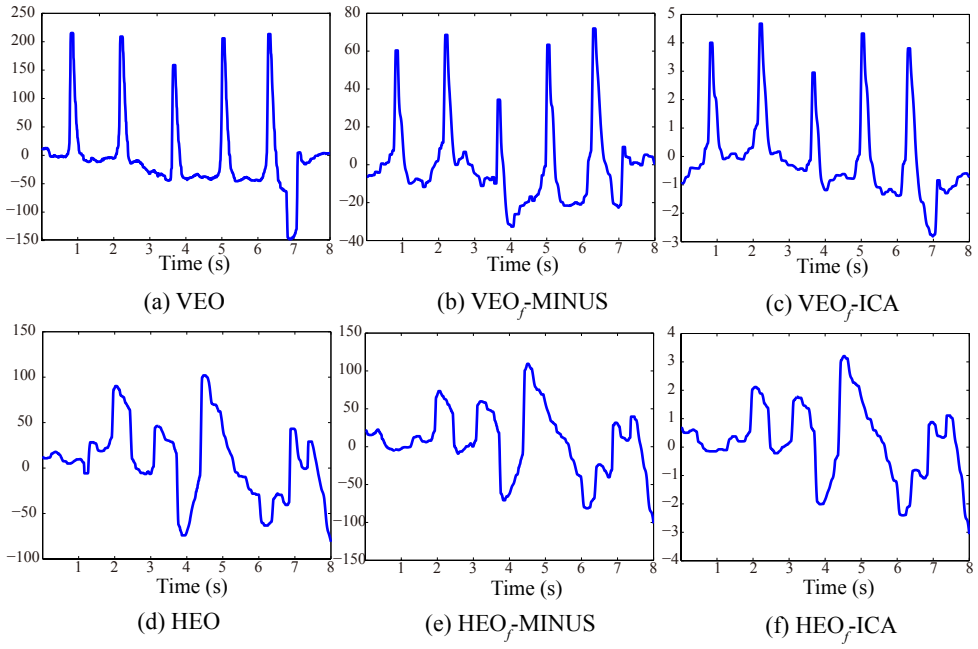


Figure 7–3 Comparison of traditional EOG and forehead EOG using minus operation and ICA separation strategies. Here, (a) and (d) are traditional VEO and HEO; (b) and (e) are extracted VEO$_f$ and HEO$_f$ from forehead EOG using the minus operation; and (c) and (f) are extracted VEO$_f$ and HEO$_f$ from forehead EOG using the ICA approach.

saccades.

7.2.2 Feature Extraction for Forehead EOG

After preprocessing the raw signals and extracting the corresponding VEO and HEO signals, we detect the candidate segments of blinks and saccades using wavelet transform method. Blinks and saccades have characteristic of pulse and rectangular signal waveforms and wavelet transform method is sensitive to singularities. We compute the continuous wavelet coefficients at a scale of 8 using a Mexican hat wavelet defined by

$$\psi(t) = \frac{2}{\sqrt{3\sigma}\pi^{\frac{1}{4}}}(1 - \frac{t^2}{\sigma^2})e^{\frac{-t^2}{2\sigma^2}}, \tag{7–1}$$

where σ is the standard deviation. The peak detection algorithm on the wavelet coefficients is used to detect blinks and saccades from the forehead VEO and HEO, respectively. The detected blinks and saccades are shown in Figures 7–4 and 7–5, respectively. We apply two thresholds θ_h and θ_l on the transformed wavelet signals and detect peaks to locate blink and saccade segments.

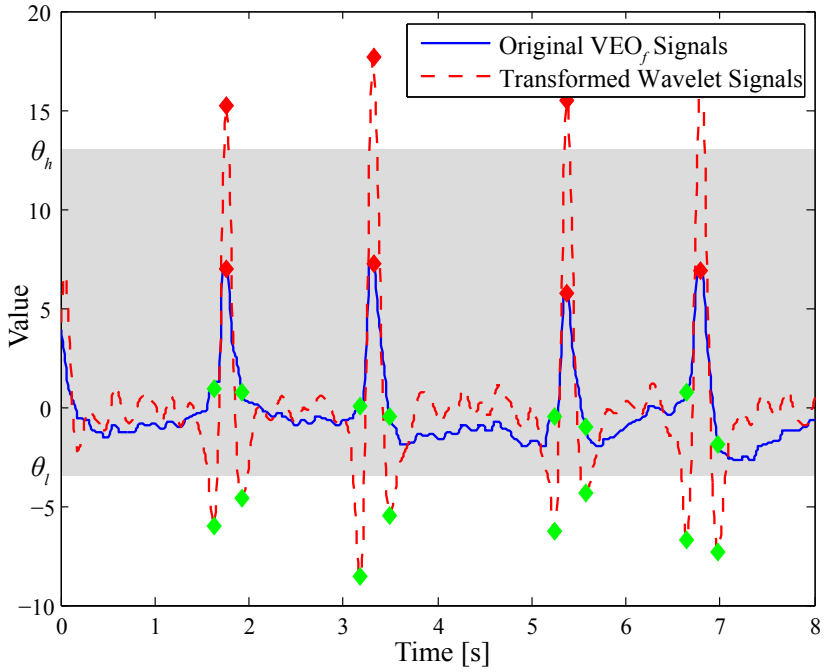


Figure 7–4 The blink detected by using continuous wavelet transform.

Red markers indicate the peaks of each blink and saccade, and green markers indicate the start and end points of each blink and saccade.

We apply thresholds on the continuous wavelet coefficients of the forehead VEO and HEO to detect positive and negative peaks and encode them into sequences. For this purpose, a positive peak is encoded as '1', and a negative peak is encoded as '0'. In continuous wavelet coefficients, positive and negative peaks denote that there are high amplitude changes in the raw signals. If the amplitude increases, there are two successive negative and positive peaks '01' and if the value decreases, two successive positive and negative peaks '10' are observed. A saccade (amplitude increases or decreases in HEO) is characterized by a sequence of two successive positive and negative peaks in the coefficients. A blink (amplitude increases and then decreases in VEO) contains three successive large peaks, namely, negative, positive, and negative, and the time between two positive peaks should be smaller than the minimum time. Therefore, segments with '01' or '10' were recognized as saccade candidates, and segments with '010' were recognized as blink candidates.

Moreover, there are some other constraints, such as slope, correlation, and maximal segment length, for guaranteeing a precise detection of blinks and saccades. After detecting blinks and saccades, we compute the statistical parameters, such as the mean, maximum, variance, and

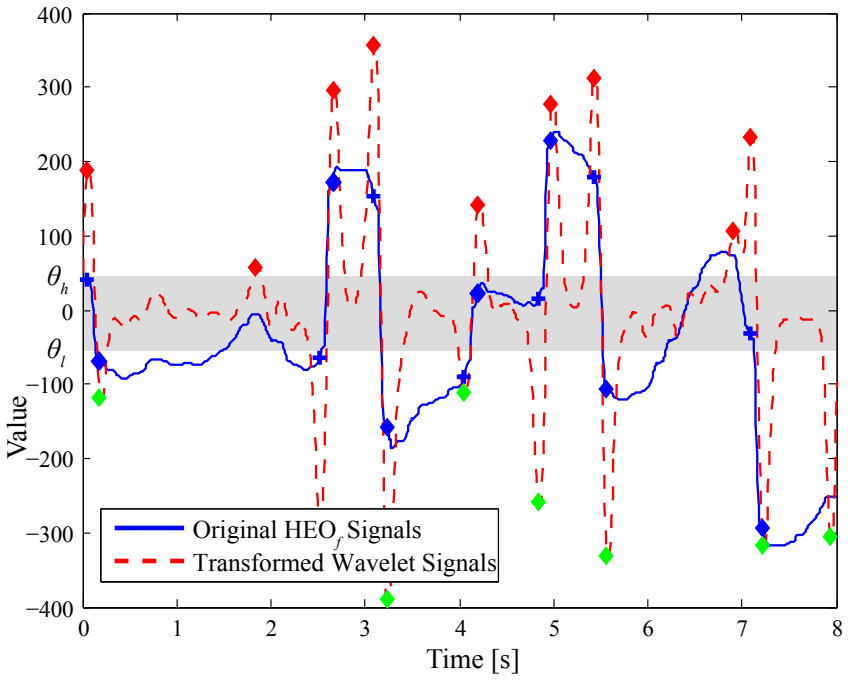


Figure 7–5 The saccade detected by using continuous wavelet transform.

derivative of different eye movements with an 8 s non-overlapping window as the forehead EOG features. We extract a total of 36 forehead EOG features from the detected blinks, saccades, and fixations, as shown in Table 7–1.

7.2.3 Forehead EEG Signal Extraction

Another advantage of the proposed forehead setup is the possibility of acquiring both EEG and EOG simultaneously without additional hardware cost. EOG is considered as a severe contamination for EEG recordings, especially in the frontal areas. Various approaches have been proposed to remove these artifacts from the recordings[261–263]. However, both EEG and EOG contain discriminative information for vigilance estimation in our task. Therefore, we try to separate EEG and EOG signals from the shared forehead electrodes.

EEG and EOG have different characteristic waveforms. EEG and EOG components are extracted from the four forehead electrodes (Nos. 4-7) using the FASTICA algorithm. Many classical artifact removal algorithms use the ICA algorithm for identifying different source components. We decompose the raw signals into several components and the forehead EEG signals are reconstructed with a weight matrix by discarding the EOG components. The raw data recorded at the four forehead channels (Nos. 4-7) are concatenated as the input matrix X for ICA as

Table 7–1 The details of the extracted 36 EOG features.

Group	Extracted Features
Blink	maximum/mean/sum of blink rate maximum/minimum/mean of blink amplitude, mean/maximum of blink rate variance and amplitude variance power/mean power of blink amplitude blink numbers
Saccade	maximum/minimum/mean of saccade rate and saccade amplitude, maximum/mean of saccade rate variance and amplitude variance, power/mean power of saccade amplitude, saccade numbers
Fixation	mean/maximum of blink duration variance and saccade duration variance maximum/minimum/mean of blink duration and saccade duration.

follows:

$$X = [Ch_4; Ch_5; -Ch_6; Ch_7], \tag{7–2}$$

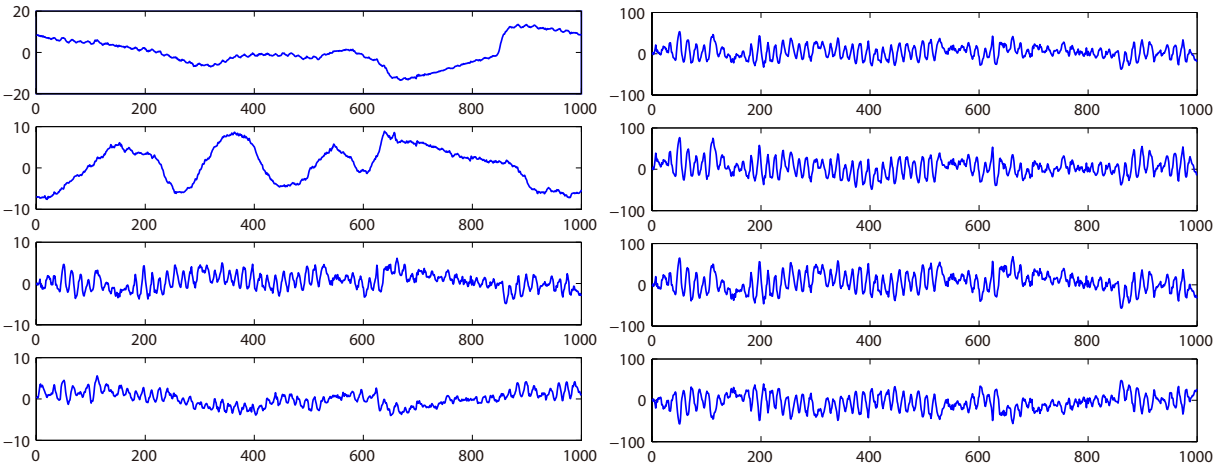
where the rows of the input matrix X are signals Ch_4, Ch_5, $-Ch_6$, and Ch_7 from channels Nos. 4-7. After ICA decomposition, the un-mixing matrix W can be obtained, which decomposes the multi-channel data into a sum of independent components as follows:

$$U = W * X, \tag{7–3}$$

where the rows of U are time courses of activations of the ICA components. The columns of the inverse matrix W^{-1} indicate the projection strengths of the corresponding components. Therefore, the clean forehead EEG signals can be derived as

$$\widetilde{X} = W^{-1} * \widetilde{U}, \tag{7–4}$$

where $\widetilde{U}$ is the matrix of activation waveforms U with rows representing EOG components set to zero.



(a) Decomposed ICA Components (b) Reconstructed Forehead EEG

Figure 7–6 (a) The decomposed independent components from the four forehead channels (Nos. 4-7) using ICA. IC 1 and IC 2 are EOG components for eye activities. (b) The reconstructed forehead EEG by filtering out the EOG components. It can be observed that strong alpha activities are verified under eye closure conditions.

The decomposed independent components and reconstructed forehead EEG of one segment under eye closure conditions are shown in Figure 7–6. From the experimental observations, the four decomposed components approximately correspond to VEO, HEO, and two-channel EEG signals, respectively. Under eye closure conditions, the alpha rhythm appears more dominant in EEG signals in previous studies[175]. From Figure 7–6 (a), we can observe that the first two rows are the corresponding eye movement components, and the last two rows contain EEG components with high alpha power values. The reconstructed signals contain characteristics of EEG waves, which are accompanied by high alpha bursts. The results presented in Figure 7–6 demonstrate the efficiency of our approach in extracting EEG signals from forehead electrodes.

7.2.4 Feature Extraction from EEG

Besides the extracted forehead EEG, EEG from critical brain regions for vigilance estimation such as temporal and posterior sites[164] are also recorded and compared. The preprocessing and feature extraction of EEG signals for vigilance estimation are similar to those in emotion recognition. The differential entropy (DE) features are extracted with with an 8-s non-overlapping window from the five frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-31 Hz), and gamma (31-50 Hz). We also compare the performance of the conventional

five frequency bands with higher frequency resolution. The DE features are extracted from the frequency band (1-50 Hz) with a 2 Hz frequency resolution.

7.3 Incorporating Temporal Dependency into Vigilance Estimation

Vigilance states are the human intrinsic mental states that involve temporal evolutions. Therefore, the samples of previous and current states have meaningful relations, and these interactions should not be discarded. Conventional regression models such as support vector regression (SVR) do not explicitly model the temporal dependency. In this paper, we apply two temporal dependency models, namely, continuous conditional neural field (CCNF) and continuous conditional random field (CCRF), to capture the vigilance dynamics and compare their performance with that of the conventional SVR model.

CCNF and CCRF are extensions of conditional random field (CRF)[264] for continuous variable modeling. These models can incorporate temporal or spatial information and have shown promising performance in various applications[265]. CCNF combines the nonlinearity of conditional neural fields[266] and the continuous output of CCRF. Compared with CCRF, CCNF can have multi-dimensional features as inputs.

The probability distribution of CCNF for a particular sequence is defined as follows:

$$P(\mathbf{y}|\mathbf{x}) = \frac{exp(\Psi)}{\int_{-\infty}^{\infty} exp(\Psi) d\mathbf{y}}, \tag{7–5}$$

where $\int_{-\infty}^{\infty} exp(\Psi) d\mathbf{y}$ is the normalization function, $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$ is a set of input observations, $\mathbf{y} = \{y_1, y_2, \cdots, y_n\}$ is a set of output variables, and n is the length of the sequence.

There are two types of features defined in these models: vertex features f_k and edge features g_k. The potential function Ψ is defined as follows:

$$\Psi = \sum_{i} \sum_{k=1}^{K_1} \alpha_k f_k(y_i, \mathbf{x}_i, \boldsymbol{\theta}_k) + \sum_{i,j} \sum_{k=1}^{K_2} \beta_k g_k(y_i, y_j), \tag{7–6}$$

where $\alpha_k > 0$, $\beta_k > 0$, the vertex features f_k denote the mapping from $\mathbf{x}_i$ to y_i with a one-layer neural network, and $\boldsymbol{\theta}_k$ is the weight vector for the neuron k.

The vertex features of CCNF are defined as

$$f_k(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}_k) = -(y_i - h(\boldsymbol{\theta}_k, \boldsymbol{x}_i))^2, and \tag{7–7}$$

$$h(\boldsymbol{\theta}, \boldsymbol{x}_i) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \boldsymbol{x}_i}}, \tag{7–8}$$

where the optimal number of vertex features K_1 is tuned through cross-validation. In our experiments, we evaluated $K_1 = \{10, 20, 30\}$.

The edge features g_k denote the similarities between observations y_i and y_j, which are defined as

$$g_k(y_i, y_j) = -\frac{1}{2} S_{i,j}^{(k)} (y_i - y_j)^2, \tag{7–9}$$

where the similarity measure $S^{(k)}$ controls the existence of the connections between two vertices.

In the experiments, K_2 was set to 1, and $S^{(k)}$ was set to 1 when two nodes i and j are neighbors; otherwise, $S^{(k)}$ was 0. The formulas for CCRF are the same as those for CCNF, except for the definition of vertex features. The vertex features of CCRF are defined as

$$f_k(y_i, \boldsymbol{x}_{i,k}) = -(y_i - \boldsymbol{x}_{i,k})^2. \tag{7–10}$$

The training of parameters in CCRF and CCNF is based on the conditional log-likelihood $P(\mathbf{y}|\mathbf{x})$ as a multivariate Gaussian. For more details regarding the learning and inference of CCRF and CCNF, please refer to[265]. The outputs of support vector regression are used to train CCRF, and the original multi-dimensional features are used to train CCNF.

7.4 Evaluation Metrics

For the continuous regression problems, we used the root mean square error (RMSE) and correlation coefficient (COR) as the evaluation metrics. RMSE is the squared error between the prediction and the ground truth, and it is defined as follows:

$$RMSE(Y, \hat{Y}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}, \tag{7–11}$$

where $Y = (y_1, y_2, ..., y_N)^T$ is the ground truth and $\hat{Y} = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_N)^T$ is the prediction. COR provides an evaluation of the linear relationship between the prediction and the ground truth, which reflects the consistency of their trends. Pearson's correlation coefficient is defined as follows:

$$COR(Y, \hat{Y}) = \frac{\sum_{i=1}^{N} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2 \sum_{i=1}^{N} (\hat{y}_i - \bar{\hat{y}})^2}}, \tag{7–12}$$

where $\bar{y}$ and $\bar{\hat{y}}$ are the means of Y and $\hat{Y}$, respectively.

For evaluation, we split the entire data from one experiment into five successive sessions and performed 5-fold cross validation. We concatenated the predictions and ground truth of the

five sessions and calculated the RMSE and COR as the evaluation metrics. In general, the more accurate the model is, the higher the COR is and the lower the RMSE is.

For parameter tuning of the regression models, we utilized support vector regression (SVR) with radial basis function (RBF) kernel as a basic regression model. The optimal values of the parameters c and g were tuned using grid search. The CCRF and CCNF regularization hyper-parameters for α_k and β_k were chosen based on a grid search in $10^{[0,1,2]}$ and $10^{[-3,-2,-1,0]}$ using the training set, respectively. The performances of CCRF and CCNF with the varying sequence length n are evaluated in the study.

7.5 Experimental Results on SEED-VIG

We evaluate the performance of our approaches on the public multimodal vigilance estimation dataset called SEED-VIG, which is developed in this study. The detailed introduction of SEED-VIG has been described in Chapter 3. The experiments were performed in the simulated driving environments, where subjects sit inside a real vehicle interacting with the virtual reality through steering wheel and gas pedal. SEED-VID consists of both EEG and forehead EOG from 23 subjects. Besides the forehead EEG, the temporal [1] and posturer [2] EEG were also recorded for comparisons. The whole experiment last for about 2 hours. There were a total of 885 samples for each experiment. The corresponding vigilance annotation was obtained from eye tracking glasses using the PERCLOS index.

7.5.1 Forehead EOG-based Vigilance Estimation

We compare the similarity of forehead EOG signals and conventional EOG signals and the performance of different separation methods. The forehead VEO$_f$ and HEO$_f$ are extracted using the minus and ICA separation approaches. We use the correlation coefficient as the similarity measurement between the forehead and traditional setups. The mean correlation coefficients of VEO$_f$-MINUS, VEO$_f$-ICA, HEO$_f$-MINUS, and HEO$_f$-ICA are 0.63, 0.80, 0.81, and 0.75, respectively. These comparative results demonstrate that the extracted forehead VEO$_f$ and HEO$_f$ contain most of the principal information of traditional EOG. VEO extracted with ICA and HEO extracted with minus approach are more proper separation methods.

Table 7–2 presents the mean RMSE, the mean COR, and their standard deviations with different separation methods. In the table, 'ICA-MINUS' denotes ICA-based VEO$_f$ and minus-

[1] FT7, FT8, T7, T8, TP7, and TP8

[2] CP1, CPZ, CP2, P1, PZ, P2, PO3, POZ, PO4, O1, OZ, and O2

ICA-MINUS		EOG-ICA		EOG-MINUS	
COR	RMSE	COR	RMSE	COR	RMSE
0.7773	**0.1188**	0.4774	0.1582	0.7193	0.1288
0.1745	**0.0391**	0.5381	0.0844	0.3492	0.0588

Table 7–2 The mean RMSE, the mean COR, and their standard deviations with different separation methods. Here, the numbers in the first and second rows are the averages and standard deviations, respectively.

based HEO_f separations, which has the highest correlation coefficients with traditional VEO and HEO. Among the separation methods, ICA-MINUS achieves the best performance for vigilance estimation in terms of both COR and RMSE. It is consistent with previous results that VEO_f-ICA and HEO_f-MINUS are more similar to the original VEO and HEO. VEO signals mainly contain blink features, which can be easily detected by ICA due to its sensitivity to singularity. In contrast, the minus method reduces the amplitude of VEO signals since the polarity of the pair electrodes is the same. Thereofore, it is more suitable to extract forehead VEO using the ICA algorithm. HEO signals mainly contain saccade features with characteristic of square waveforms, which are more difficult to be detected by ICA. But the polarity of the pair electrodes used in horizontal placements is different and the amplitude can be enhanced. Therefore, it is more suitable to extract forehead HEO using the minus approach.

7.5.2 EEG-based Vigilance Estimation

We compare the performance of EEG from different brain regions for vigilance estimation. The forehead, temporal, and posterior EEG have four, six, and twelve channels. The DE features are extracted from the five frequency bands and the total frequency band with 2 Hz frequency resolution. Table 7–3 presents the mean COR, mean RMSE and their standard deviations of different EEG features from different brain areas. The posterior site performs the best, followed by the temporal and then forehead sites. Previous findings support that the posterior brain region is one of the critical areas for vigilance estimation[164]. High frequency resolution with 2 Hz achieves better performance than the five frequency bands.

Besides the reported detection accuracy, the neural patterns under different vigilance states are discussed. Whether these neural patterns exist and are common across subjects to some extent? The investigation of relations between specific brain activities and mental states provides evidence and support for understanding the underlying brain mechanisms and promoting brain decoding. Previous studies have focused on the relevant topics. For example, Huang and col-

(a) COR

Posterior		Temporal		Forehead	
2 Hz	5 Bands	2 Hz	5 Bands	2 Hz	5 Bands
0.7001	0.6807	0.6678	0.6410	0.6502	0.5749
0.2250	0.2129	0.2349	0.2246	**0.2116**	0.2463

(b) RMSE

Posterior		Temporal		Forehead	
2 Hz	5 Bands	2 Hz	5 Bands	2 Hz	5 Bands
0.1327	0.1429	0.1385	0.1603	0.1463	0.1640
0.0303	0.0393	0.0343	0.0722	0.0383	0.0483

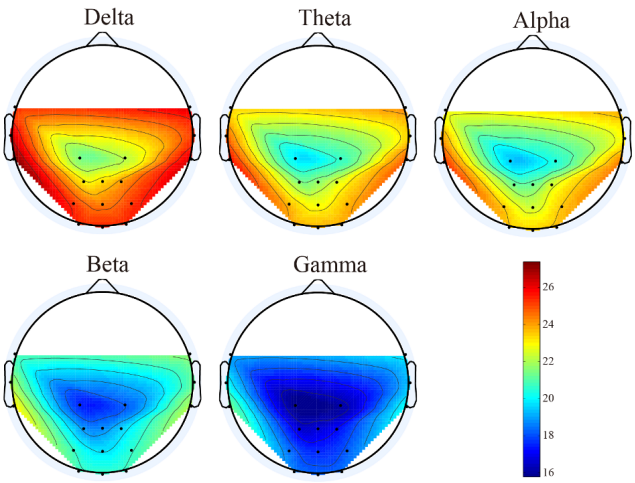
Table 7–3 The average and standard deviations of COR and RMSE for different EEG features. Here, the numbers in the first and second rows are the averages and standard deviations, respectively.

leagues found significant alpha response in the occipital site increase with reaction time increasing[267]. Ray and colleagues proposed that alpha responses can reflect attentional demands[86]. If the attentional demands decrease, the parietal alpha activities increase.

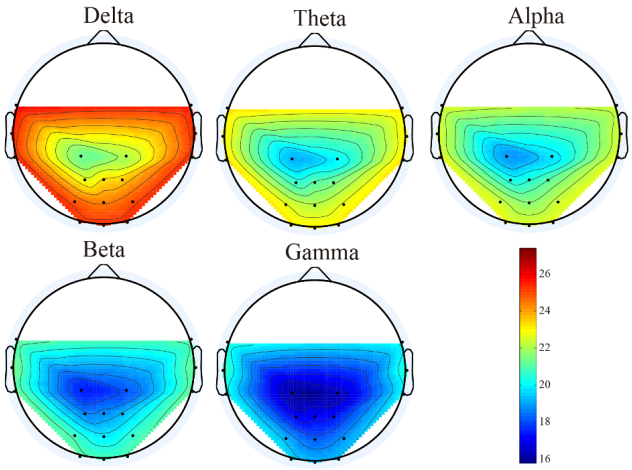
By applying two thresholds (0.35 and 0.7) on the continuous PERCLOS index, we split the EEG data into three categories: awake, tired, and drowsy. The EEG features of different categories are averaged over different experiments. Figure 7–7 presents the mean neural patterns of awake and drowsy states as well as the difference between them. As shown in Figure 7–7, we observe that increasing theta and alpha frequency activities and decreasing gamma frequency activities exist in temporal and parietal areas in drowsy states in contrast to awake states ($p < 0.01$, one-way analysis of variance, ANOVA). These results are consistent with previous findings in the literature[33–35, 86, 168, 170, 267] and support the previous findings that the increasing trend for the ratio of slow and fast waves of EEG activities reflects decreasing attentional demands[268].

7.5.3 Modality Fusion with Temporal Dependency

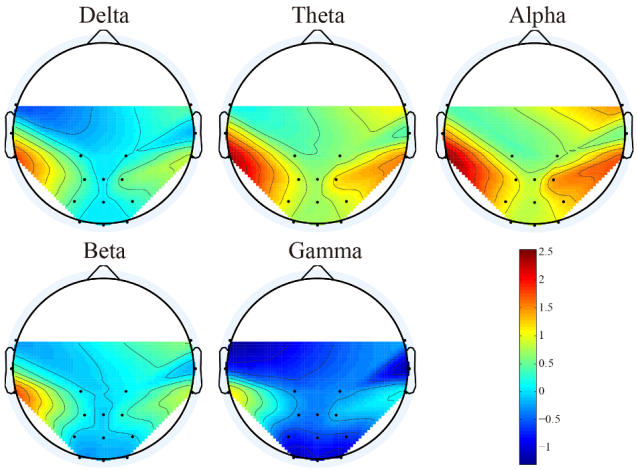
We use the feature-level fusion as the modality fusion method, where the features of EEG and EOG are directly concatenated into a larger feature vector as inputs of regression models. The forehead EOG features are extracted in the ICA-MINUS way. We combine the EOG features and the EEG features from different brain regions (forehead, temporal, and posterior) individually for multimodal vigilance estimation. The performance of using single modality solely and different



(a) Drowsy

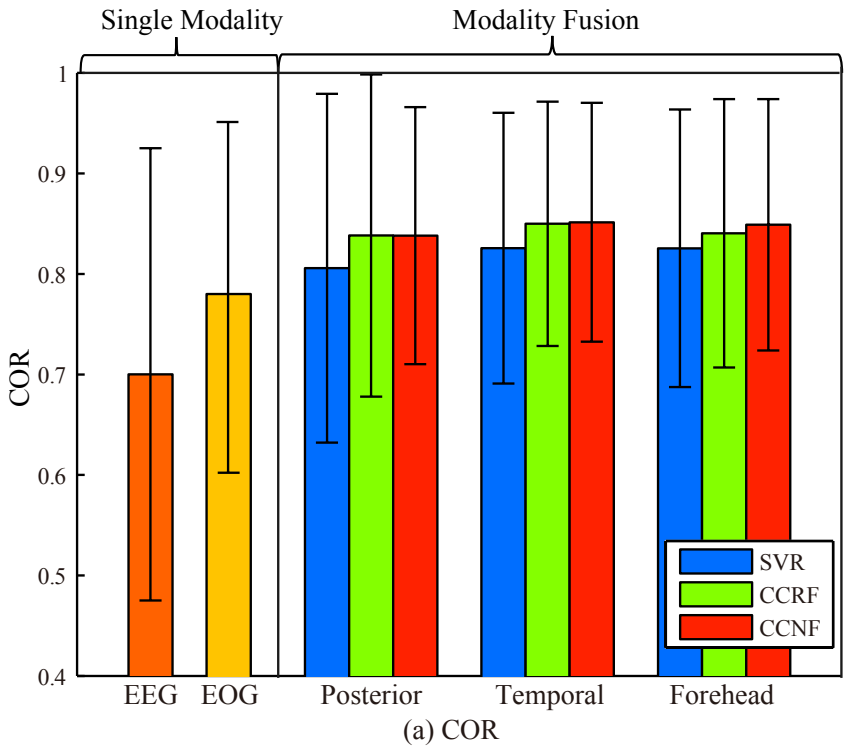


(b) Awake

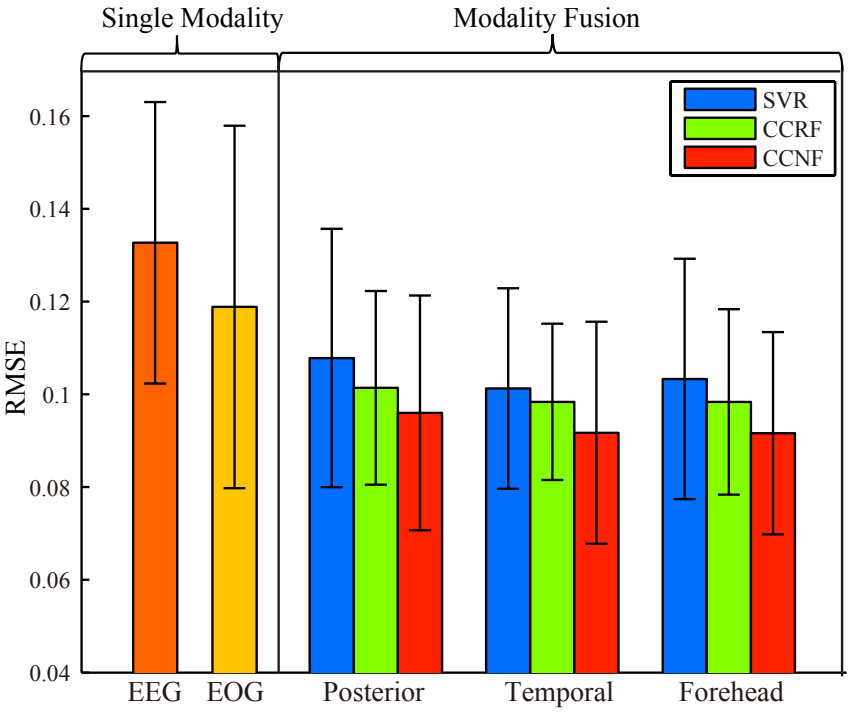


(c) Difference between (a) and (b)

Figure 7–7 The mean neural patterns of awake and drowsy states as well as the difference between these two states. From the average neural patterns, we observe that drowsy states have higher theta and alpha frequency activities and lower gamma frequency activities in temporal and parietal areas ($p < 0.01$, ANOVA).



(a) COR



(b) RMSE

Figure 7–8 The mean COR and mean RMSE of each single modality and different modality fusion strategies.

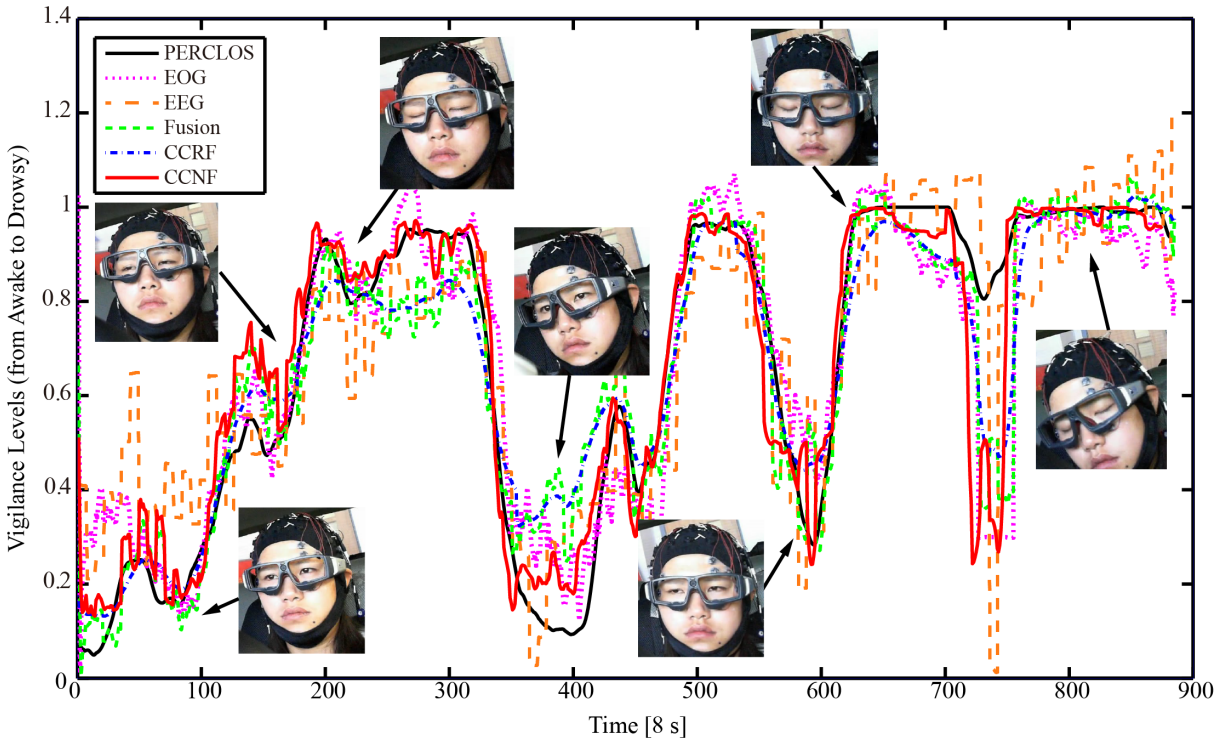


Figure 7–9 The continuous vigilance estimation of different methods in one experiment. As shown, the predictions from our proposed approaches are almost consistent with the true subjects' behaviours and cognitive states.

modality fusion strategies are presented in Figure 7–8. For the performance of using single modality, forehead EOG performs better than posterior EEG ($p = 0.2003$ for COR and $p = 0.1866$ for RMSE, ANOVA). This is because of the high relevance between forehead EOG and vigilance annotations from eye tracking data. Both approaches are based on eye movement details. Modality fusion can significantly enhance the performance in comparison with single modality with only EEG or EOG with a higher COR and a lower RMSE. The p values of COR for forehead EOG and posterior EEG are 0.2978 and 0.0264, respectively. The p values of RMSE for forehead EOG and posterior EEG are 0.0654, and 0.0002, respectively.

For comparisons of different brain regions, an interesting observation is that the combination of forehead EOG and forehead EEG achieves better performance than that of forehead EOG and posterior EEG. But the posterior EEG outperforms the other brain regions for vigilance estimation with single EEG modality. These results indicate that forehead EEG and forehead EOG have more coherent information. The fusion with the temporal EEG performs slightly

better than the forehead EEG, while requires six more electrodes for recordings. The forehead setup uses only four shared electrodes and both EEG and EOG features can be extracted in our proposed approach. The performance are relatively comparable to the best while the setup cost can be reduce a lot. Therefore, our proposed forehead approach is superior for real-world applications, which increases information flow without additional setup cost.

The CCRF and CCNF models incorporate the temporal dependency into modal training. As shown in Figure 7–8, these temporal dependency models can enhance the regression performance. For the forehead setup, the mean COR/RMSE of SVR, CCRF, and CCNF are 0.83/0.10, 0.84/0.10, and 0.85/0.09, respectively. The CCNF achieves the best performance with higher accuracies and lower standard deviations.

To verify whether the predictions from our proposed approaches are consistent with the true subjects' behaviours and cognitive states, the continuous vigilance estimation of different approaches and the corresponding ground truth are illustrated in Figure 7–9. The snapshots in Figure 7–9 show the facial frames under different vigilance levels. We can observe that our proposed multimodal approach with temporal dependency can moderately predict the continuous vigilance levels and its trends. The predictions from our proposed approaches are almost consistent with the true subjects' behaviours and cognitive states.

7.5.4 Complementary Characteristics

The fusion of EOG and EEG can significantly enhance the performance of vigilance estimation in comparison with single modality. In this section, we focus on investigating the complementary characteristics of EEG and EOG with the confusion matrices of each single modality. The confusion matrix computes the sample percentage of one class that is assigned to the other class, which indicates the strength and weakness of each modality for specific tasks. The EEG features are separated into three categories (awake, tired and drowsy states) with two thresholds (0.35 and 0.7) on the continuous PERCLOS index as described above. Multiclass SVMs are trained for predicting these three vigilance states using single EEG or EOG modality.

Figure 7–10 presents the mean confusion graph of forehead EOG and posterior EEG over all the experiments. These results demonstrate that posterior EEG and forehead EOG have important complementary characteristics. Forehead EOG has the advantage of classifying awake and drowsy states (77%/76%) compared to the posterior EEG (65%/72%), whereas posterior EEG outperforms forehead EOG in recognizing tired states (88% vs. 84%). The forehead EOG modality achieves overall better performance than the posterior EEG. Since the vigilance anno-

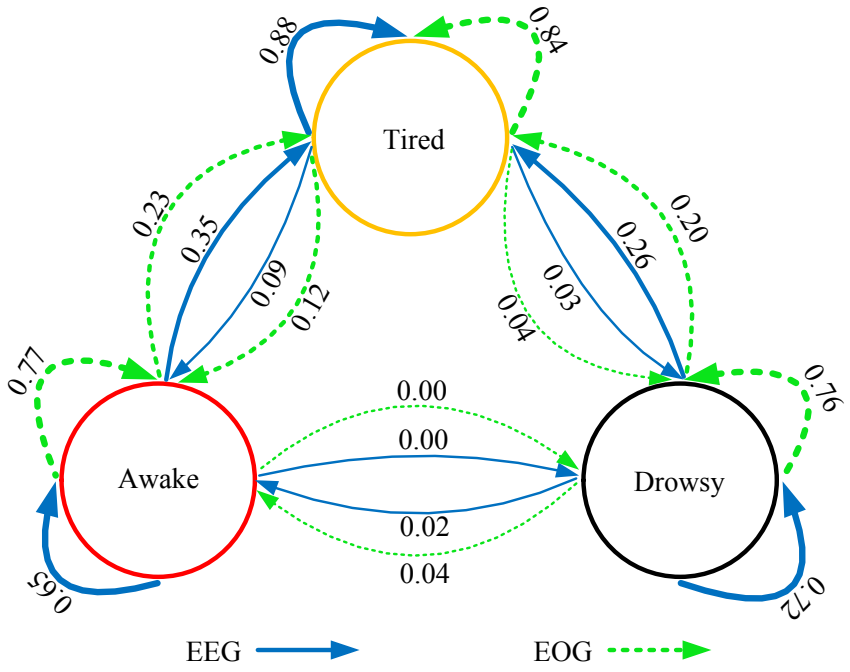


Figure 7–10 Confusion graph of forehead EOG and posterior EEG, which shows their complementary characteristics for vigilance estimation. Here, the numbers denote the percentage values of samples in the class (arrow tail) classified as the class (arrow head). Bolder lines indicate higher values.

tations are obtained from eye tracking glasses, EOG with details of eye movements has relevant information. Moreover, awake states and tired states are often misclassified with each other, and similar results are observed for drowsy and tired states. In contrast, the gap between low vigilance (drowsy) and high vigilance (awake) is relatively larger. awake states are seldom misclassified as drowsy states and vice versa for both modalities. These observations are consistent with our common sense. The differences of EEG and EOG features under awake and drowsy states are more significant. Therefore, these results demonstrates EOG and EEG contains different discriminative information for vigilance estimation. Combining the complementary information of EEG and EOG, modality fusion can improve the prediction performance for vigilance estimation.

7.6 Experimental Setups with Wearable Device

In the above experiments, we conduct the evaluations of our proposed approaches with data recorded by commercial devices in the laboratory simulations. To improve feasibility and efficiency of vigilance estimation in real driving environments, we implement a wearable EOG

device with flexible dry electrodes and an acquisition board. The detailed implementation of the designed device has been described in Chapter 3. In this section, we perform the evaluations of the whole vigilance estimation system in both simulated and real driving environments. Fourteen and ten subjects participated in the simulated and real-world driving environment experiments, respectively. Accurate eye movement parameters from eye-tracking glasses were extracted to calculate the PERCLOS index for vigilance annotation. The laboratory simulations were the same as that in SEED-VIG for fair comparisons. The whole data recording process in the laboratory simulations last for about 2 hours.

To evaluate the efficiency of our system, systematic experiments are performed in real scenarios under various illumination and weather conditions. The real-world driving experiments were performed inside Shanghai Jiao Tong University, Minhang Campus, in Shanghai, China. The driving speed limit was kept at approximately 30 km/h, and the duration of the driving was approximately 1.5 hours. To ensure safety when the subjects were drowsy, the subjects with the wearable device were asked to sit in the front passenger seat beside the drivers during the experiments. The forehead setup with four electrodes was utilized for EOG recordings.

There are about 885 and 675 samples for each simulated and real driving experiments, respectively. For evaluation, we split the entire data from one experiment into five successive sessions and performed 5-fold cross validation. We concatenated the predictions and ground truth of the five sessions and calculated the root mean square error (RMSE) and correlation coefficient (COR) as the evaluation metrics. The parameter tuning for support vector regression (SVR) with radial basis function (RBF) kernel, CCRF, and CCNF is the same as that in SEED-VIG.

7.7 Experimental Results on Wearable Device

7.7.1 Laboratory Driving Simulations

In this section, we present the experimental results of the simulated driving experiments in the laboratory. First, we evaluate the performances of CCRF and CCNF with varying sequence length n. The sequence length determines the strength of modeling temporal dependency. A longer sequence length makes inference based on larger time windows with large-scale information. However, it might reduce the representational capacity for the vigilance fluctuations. There is a trade-off between exploration and exploitation.

Figure 7–11 shows the performances of CCRF and CCNF with varying sequence length

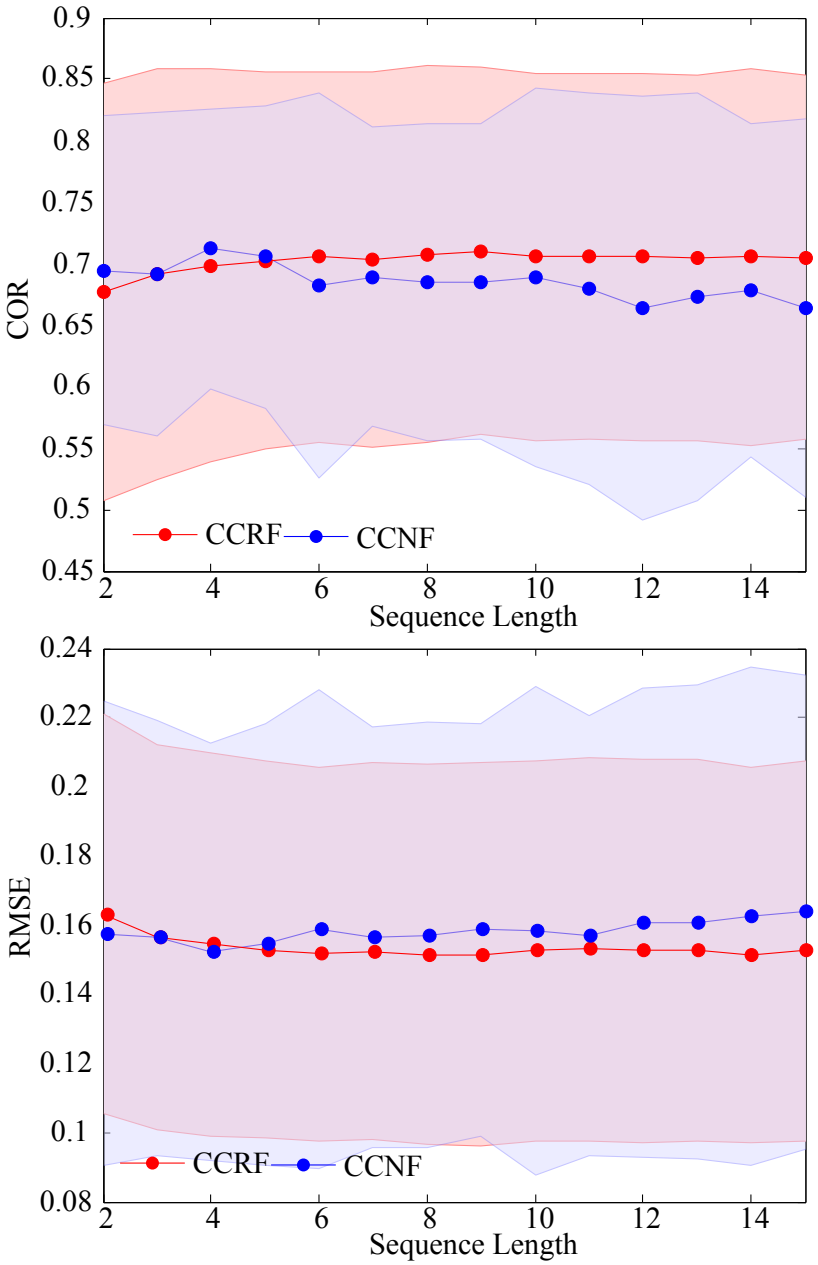


Figure 7–11 The performances of CCRF and CCNF with varying sequence length in the simulated driving experiments.

n. From the experimental results, the performance of CCRF is relatively stable, while that of CCNF is slightly degraded with a larger sequence length. The best sequence lengths of CCRF and CCNF are nine and four, where CCRF and CCNF achieve the best mean performances of 0.7104±0.1491 and 0.7118±0.1134 for COR, respectively, and their mean RMSEs

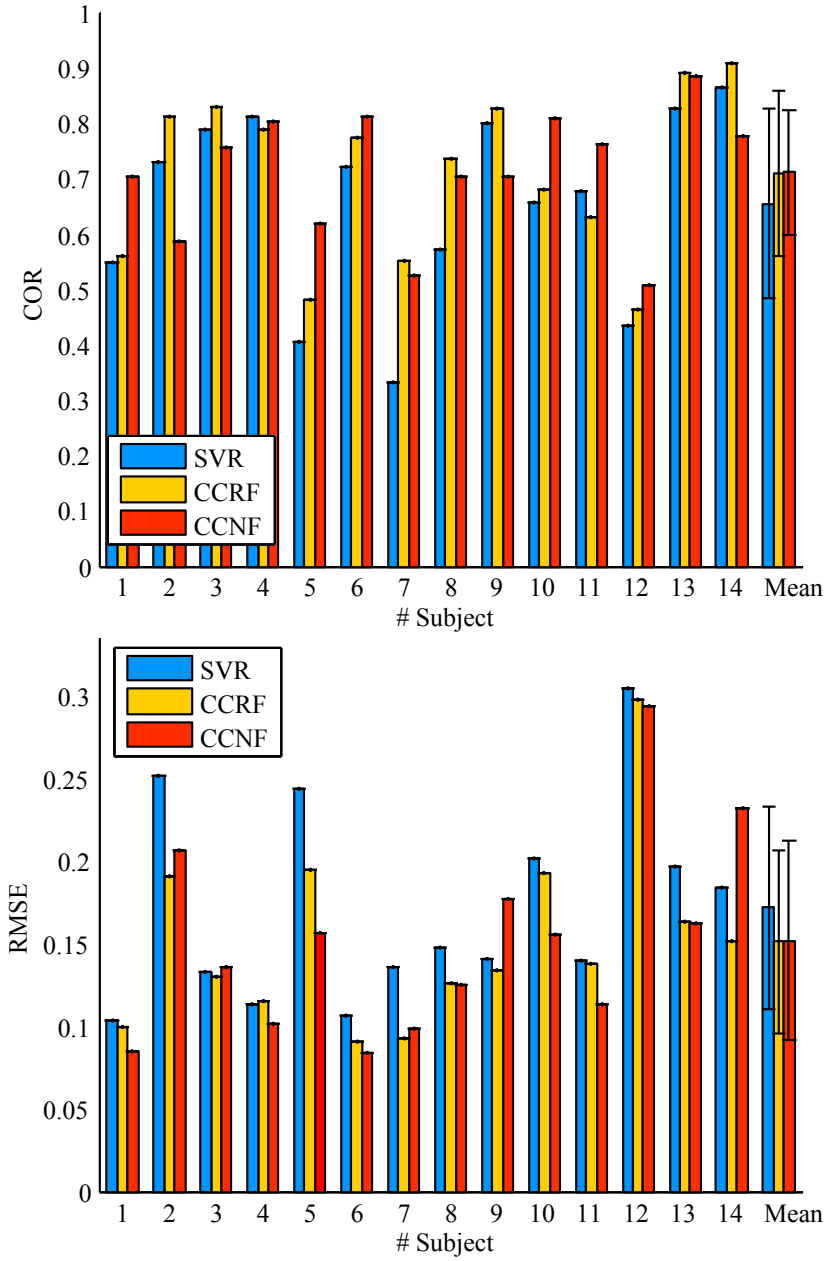


Figure 7–12 The detailed performances of different models: SVR, CCRF, and CCNF for individual subjects in the simulated driving experiments.

are 0.1515±0.0552 and 0.1523±0.0604, respectively. We observe that although the predictions of CCNF obtain higher correlation with the ground truth, they have larger mean square errors than those of CCRF. Moreover, we find that if vigilance estimation with high temporal resolution in some situations is vital, CCRF and CCNF with a much smaller sequence length (e.g., 2)

can achieve slightly lower performances of 0.6770/0.6943 and 0.1632/0.1576 for COR/RMSE, respectively.

Figure 7–12 shows the detailed performances of different models for individual subjects. The mean COR and RMSE of the baseline method SVR are 0.6554±0.1712 and 0.1721±0.0613, respectively. We evaluated the statistical significance using one-way analysis of variance (ANOVA). The temporal dependency models (CCRF and CCNF) perform better than the conventional method of SVR with p values of 0.3730/0.3609 and 0.3132/0.3970 for COR/RMSE. The effect-size correlations of CCRF and CCNF are 0.1688/-0.1738 and 0.1906/-0.1605 for COR/RMSE, respectively, in simulated driving experiments. These results indicate that incorporating temporal dependency information into vigilance estimation is efficient. These findings are consistent with previous studies of modeling temporal and spatial dependencies when analyzing time series data in the literature[265, 266].

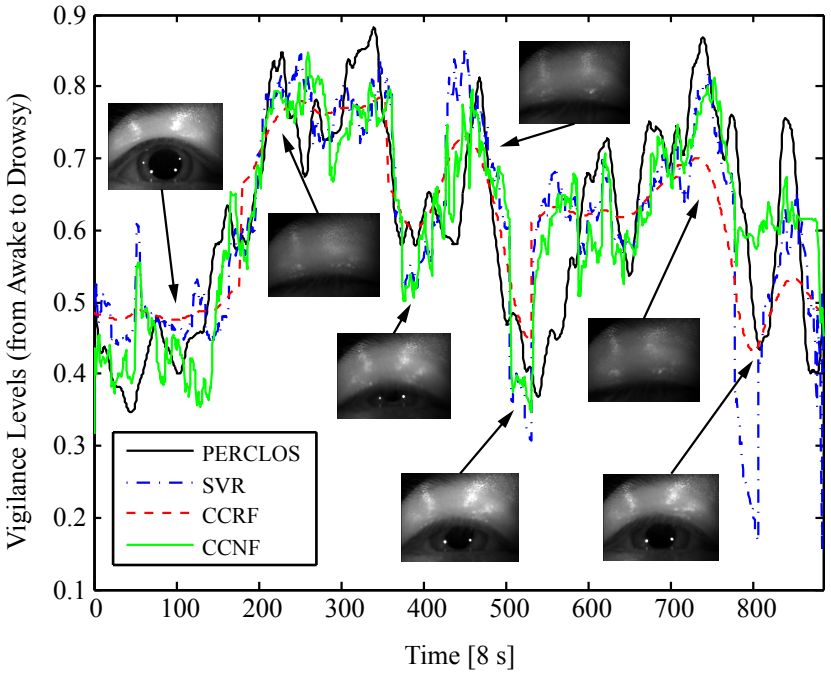


Figure 7–13 The continuous vigilance estimation of different methods in one experiment (Subject 6) in the simulated driving experiments. As shown, the predictions from our proposed system are almost consistent with the true subjects' vigilance states.

In the previous Section 7.5, we construct vigilance estimation models with forehead EOG recorded with the traditional wet electrodes using a commercial Neuroscan system, which represents the gold standard. The mean COR and RMSE of our previous wet-electrode-based

approach are 0.7773±0.1745 and 0.1188±0.0391, respectively. In comparison with the above study, we design a wearable dry EOG prototype that is feasible for real-world scenarios with a moderate decrease in prediction performance to some extent.

To verify whether the predictions from our proposed approaches are consistent with the subjects' true behaviors and cognitive states, the continuous vigilance estimation of one experiment (Subject 6) is shown in Figure 7–13. The snapshots in Figure 7–13 show the eye states corresponding to different vigilance levels, which are captured using eye-tracking glasses. We can observe that our proposed system that combines the wearable dry EOG prototype and the temporal dependency models can moderately predict the continuous vigilance levels and their trends.

7.7.2 Real-World Driving Experiments

In this section, we extend the evaluations of our designed vigilance estimation system to the real-world driving tasks. Figure 7–14 shows the performances of CCRF and CCNF with varying sequence length n. As shown, in real scenarios, CCNF performs better than CCRF with a higher COR and lower RMSE. The sequence length n does not considerably influence their performances. CCRF and CCNF achieve their peak performances with sequence lengths of 7 and 9, respectively. We compare the performances of different models individually in Figure 7–15. The mean CORs of SVR, CCRF, and CCNF are 0.5784±0.1882, 0.6133±0.1779, and 0.6620±0.2688, respectively (CCRF: $p = 0.6752$, CCNF: $p = 0.4307$, ANOVA), while their mean RMSEs are 0.1890±0.0953, 0.1630±0.0594, and 0.1307±0.0486, respectively (CCRF: $p = 0.4747$, CCNF: $p = 0.1022$, ANOVA). The effect-size correlations of CCRF and CCNF compared with SVR are 0.0948/-0.1615 and 0.1772/-0.3595 for COR/RMSE, respectively. The conclusion that temporal dependency models perform better than the baseline method SVR is validated in the real-world tasks. Figure 7–16 presents the predictions of different models and the ground truth of vigilance fluctuations.

Compared with the performance of the laboratory simulated driving, the performance of our system in real-world driving environments decreases approximately 0.0498 for COR and 0.0216 for RMSE. Considering the challenges of complex outdoor environments, including extreme weather such as rainy and cloudy conditions, our wearable system can still perform well with slightly decreased performance. These results demonstrate that our proposed vigilance estimation system with wearable dry forehead EOG and temporal dependency models is efficient in both simulated and real-world driving environments.

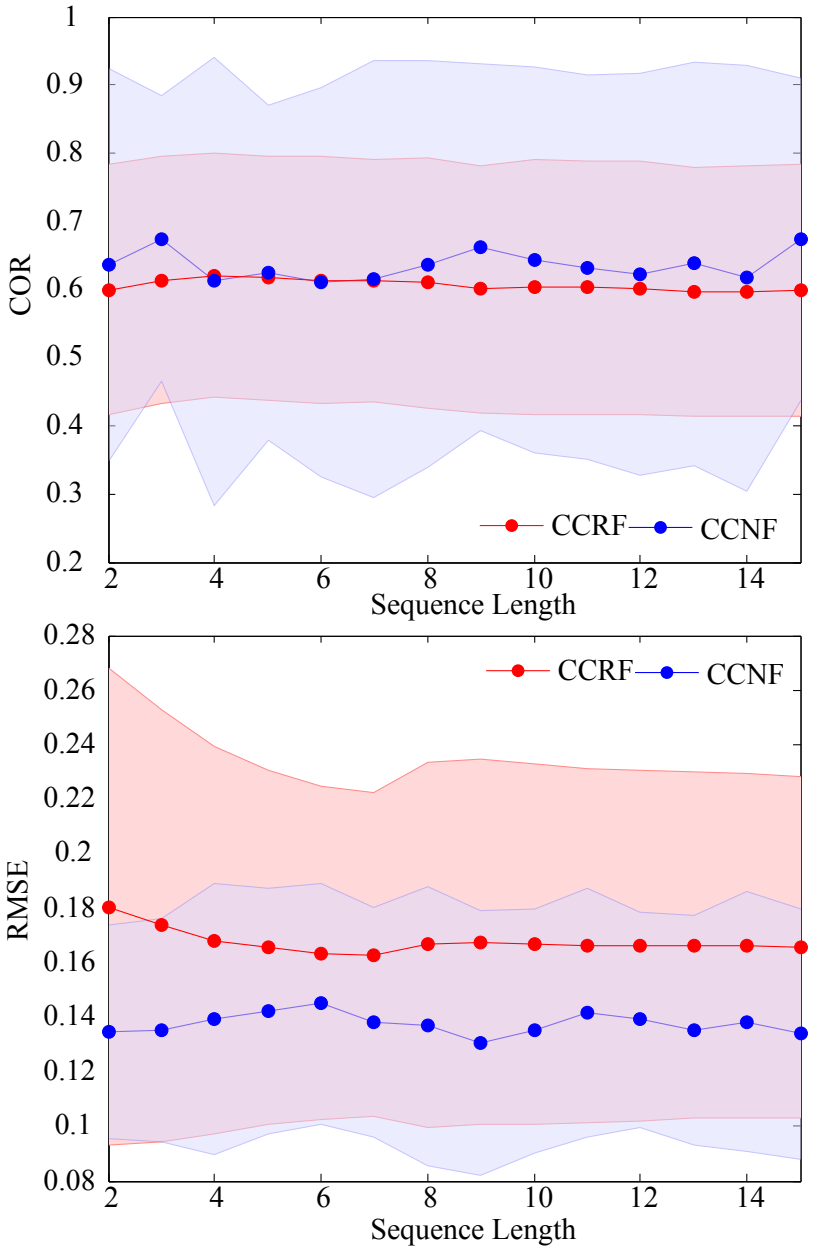


Figure 7–14 The performances of CCRF and CCNF with varying sequence length in the real-world driving experiments.

7.8 Discussion

Although considerable progress in vigilance estimation has been achieved over the past decades, few studies have been performed in real-world scenarios[175, 182, 183], and most studies are performed in controlled laboratory environments. A small number of studies in real-world environ-

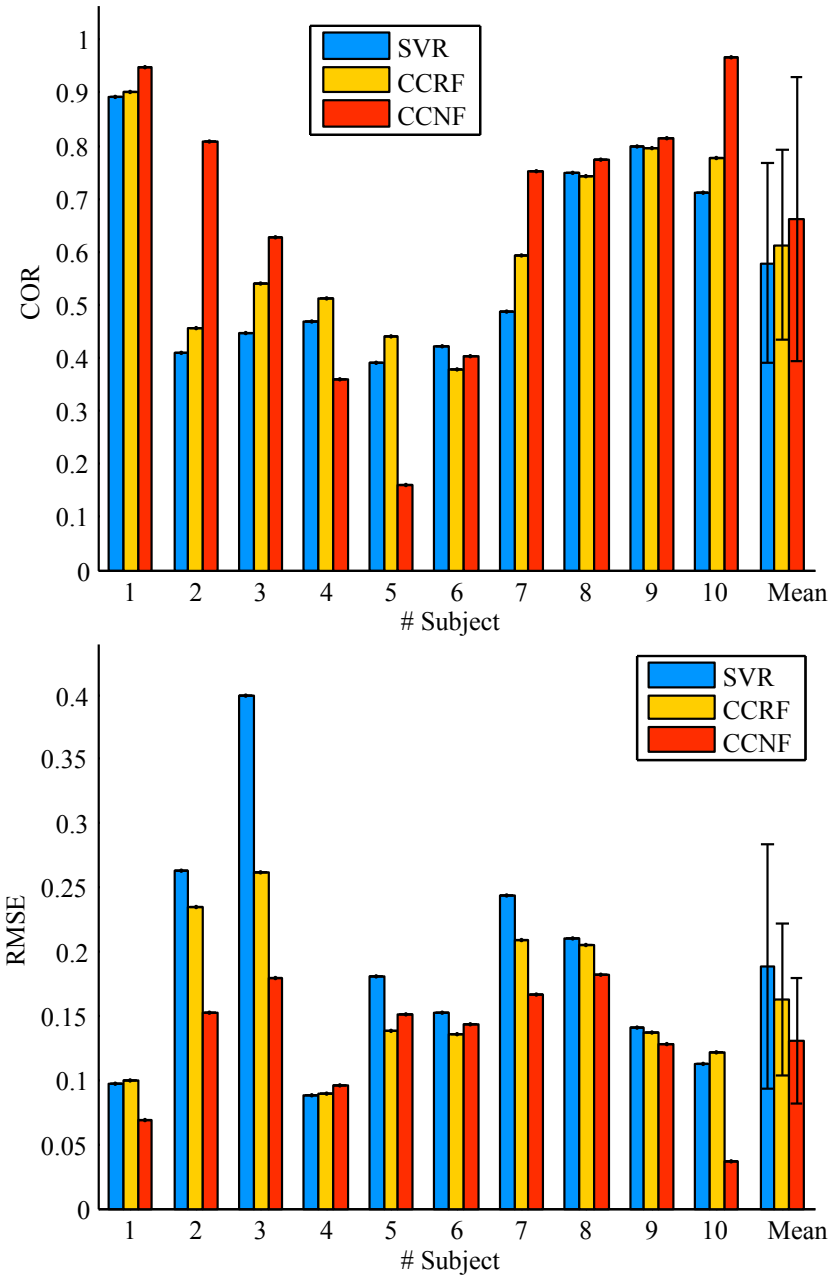


Figure 7–15 The detailed performances of different models, SVR, CCRF, and CCNF, for individual subjects in the real-world driving experiments.

ments are mostly based on facial videos due to easy setups at the cost of degraded performance. There is a large gap between simulations and real conditions. In this study, we perform both simulated and real-world experiments for our designed system to fill this gap. The experimental results demonstrate the efficiency of our system in real-world applications.

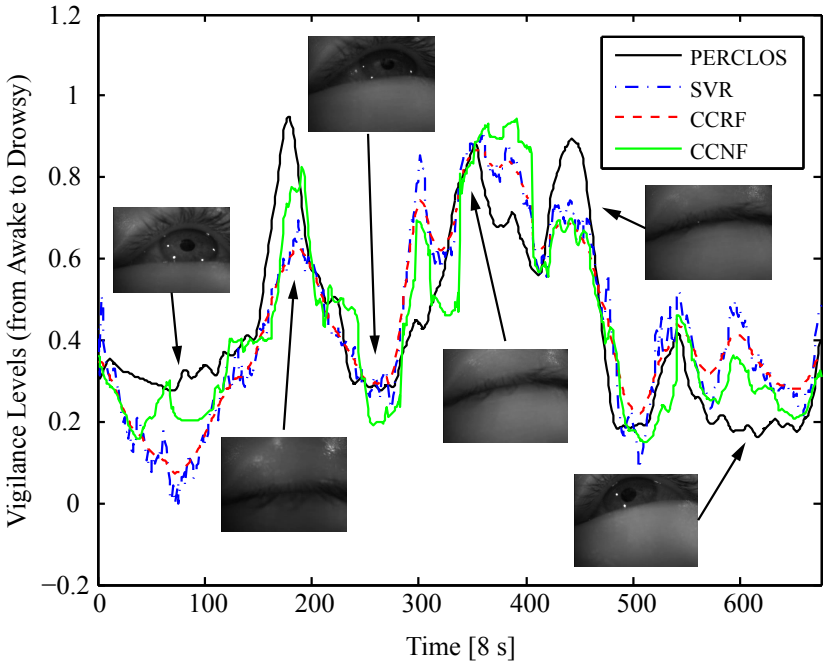


Figure 7–16 The continuous vigilance estimation of different methods in one experiment (Subject 9) in the real-world driving experiments.

Considering safety issues, we actually analyze the data of subjects sitting in the front passenger seat rather than drivers in real driving environments. Since the operations are different for drivers and passengers, the performance of the system for real drivers requires further investigation. Note that although we have evaluated the performance of our system in real-world driving environments, the real-time online performance has not been tested, which should be evaluated in the near future.

One of the limits for constructing robust vigilance estimation models is the high cost of collecting a large amount of physiological data. By leveraging the feasibility and wearability of our vigilance estimation prototype, we are able to collect a large amount of related data, and these large datasets can actively help generalize our computational models and novel prototype designs as feedback. Through these progressive procedures, we can pave the way for robust vigilance estimation in complex real environments. Therefore, it is very important to implement such wearable systems and test them on board vehicles.

To reduce the calibration time and improve the generalization performance of vigilance estimation models, an intuitive approach is to adaptively recycle the previously recorded data. However, there are individual differences in neurophysiological signals across subjects and sessions.

The performance of vigilance estimation models may be dramatically degraded. To address this problem, one efficient approach is to perform subject transfer and session transfer using transfer learning techniques, which aim to reduce the differences of feature distributions between source domains and target domains[129].

Another similar problem is about the reality gap between simulated and real-world environments. There are many different characteristics of simulated and real scenarios. Real driving environments contain many more problems (e.g., device setups and artifacts) compared with laboratory simulations. In real-world experiments, there is usually much more noise caused by vehicle engine and vibration when recording EOG signals. However, it is very time consuming and expensive to collect a large number of high quality data in real-world environments. Therefore, an intuitive and straightforward way to dealing with this problem is to train models with data from simulations and make inference on real-world data in a cross-environment way. However, the performance of the overall systems might be dramatically degraded due to the reality gap. Due to the reality gap and the high cost of data collections in real scenarios, how to transfer knowledge from simulations to reality and generalize computational models to real-world applications are still open questions. To address this problem, transfer learning should be further investigated for knowledge transfer from simulations to the real world and generalizing the models to real applications.

Compared with the EEG-based vigilance estimation approaches, EOG-based methods are much easier to implement with feasibility and wearability in real sceneries with the advantages of high amplitudes of EOGs and good placements of electrode mounting. The limitation of EOG-based vigilance estimation is that it requires a larger time window for extracting sufficient information since most eye movements such as blink and saccade are based on second scales. For example, the time window for EOG feature extraction is 8 s. In contrast, EEG signals have high temporal resolution, and they are able to capture the vigilance dynamics. In the study of Davidson *et al.*[34], they developed an EEG-based lapse detection system on the temporal scale of 4 s.

There are also some limitations for the work. The subject groups are mainly young students and the performance evaluations with more age groups should be further conducted. Although we have considered the weather and illumination factors and evaluated the performance under these conditions, the detailed influence of these factors in our proposed system needs systematical evaluations and further investigations with more experiments in the future. More temporal dependency models like Recurrent Neural Networks (RNN) and Long Short Time Memory

(LSTM) can be introduced to capture vigilance dynamics and further improve the prediction performance. Recent studies with deep neural networks achieve better performance on SEED-VIG[269]. Moreover, the current work mainly focuses on the detection of vigilance states without any feedbacks. How to introduce efficient feedbacks after detecting vigilance decrement to ensure driving safety is another research topic. And the assessment about the efficiency of feedbacks should be also studied. The closed-loop brain-computer interactions consist of both the detection and the feedback phases.

7.9 Summary

In this Chapter, multimodal vigilance estimation with EEG and EOG is introduced. The proposed approach with fusion of EEG and EOG and temporal dependency models is evaluated on the public vigilance estimation dataset called SEED-VIG in the simulated driving environment. The novel forehead setup can extract both EEG and EOG features simultaneously with only four shared electrodes. From the experimental results in Section 7.5, the performance is comparable to the optimal posterior setup with easy setup. The complementary characteristics of EEG and EOG are revealed. And the neural patterns of different vigilance states are investigated.

For dealing with the problem of vigilance estimation in real driving environments, we have proposed both wearable device implementations for easy signal acquisition and efficient algorithms for precise data modeling in this study. The designed wearable device has integrated only four flexible dry electrodes and a low-cost acquisition board for real-world forehead EOG recordings in comparison with the existing devices. We have performed the systematical evaluations of our designed system in real-world driving environment under various weather conditions and validated its efficiency for continuous vigilance estimation in Section 7.7. This study has provided an efficient approach to real-time vigilance estimation in real-world applications and the sceneries can be extended from driving tasks to various tasks that require sustained attention.

Chapter 8

Conclusions and Future Work

Affective brain-computer interactions try to fill the sematic gap between highly emotional humans and emotionally challenged machines. For the goal of developing machine intelligence with emotion intelligence, affective computing is still a very young research field in general. Recent years have witnessed growing interests of affective computing in research and industry communities. In this thesis, we have explored several research problems in the field of affective brain-computer interactions. Whether we can decode human mental states with neurophysiological signals and behaviors? Specifically, emotion and vigilance are studied using EEG, EOG, and behaviors with machine learning algorithms. Three public benchmark datasets for emotion recognition and vigilance estimation have been developed to promote scientific cooperations. The proposed computational models and prototype demonstration are systematically evaluated.

8.1 Summary of Contributions

The main aim of the work is to find the mapping between observed brain signals and convert mental states of emotion and vigilance, and develop the computational models for automatic mental state detection with machine learning approaches. Both theoretical and practical implementations for multimodal emotion recognition and vigilance estimation are explored.

In Chapter 3, we developed several benchmark datasets for emotion recognition and vigilance estimation. The details of emotion elicitation, materials, subject information, experimental protocols and data recordings were introduced. We released two public emotion recognition datasets called SJTU Emotion EEG Dataset (SEED) for three emotions (happy, sad, and neutral) and SEED-IV for four emotions (happy, sad, fear, and neutral). The performance of various proposed computational models were mainly evaluated on these two datasets. The detailed preprocessing and feature extraction of EEG and eye movements were presented. For vigilance estimation, the simulated driving experiments were performed in the laboratory and the multimodal dataset with EEG and EOG for vigilance estimation (SEED-VIG) was released. In order to improve the feasibility and wearability, we implemented a wearable prototype with flexible dry electrodes and an acquisition board and performed evaluations in both simulated and real

driving environments. All the datasets were freely available to the academic community via the project website of SEED.

In Chapter 4, the experimental results and discussions of EEG-based emotion recognition were presented. Deep neural networks were introduced to improve the performance in comparison with shallow models. The critical frequency bands and brain regions that are associated with emotion recognition were also explored. Based these findings, electrode numbers were reduced and the optimal electrode placements were investigated. The performance of various feature extraction, feature smoothing, feature reduction, and classification methods for EEG-based emotion recognition were systematically evaluated on the public datasets: DEAP and SEED. The neural patterns of three emotions (happy, sad, and neutral) were identified. The beta and gamma responses at the temporal areas increased for happy emotion. The neural patterns of neutral emotion and sad emotion were similar. Neutral emotion had higher alpha responses at parietal and occipital sites, while sad emotion had higher delta responses at parietal and occipital sites and higher gamma responses at prefrontal sites. These neural patterns were shown commonality across individuals and sessions. The stability of the computational models over time was also investigated with training data and test data from different sessions in different days.

In Chapter 5, we proposed to a multimodal emotion recognition framework called *EmotionMeter* with fusion of EEG and eye movements. EEG and eye movements represent the internal cognitive states and external subconscious behaviors, respectively. We adopted various modality fusion strategies, including feature-level fusion, decision-level fusion, and bimodal deep autoencoder. The experimental results demonstrated significant improvements with about 10% accuracy using fuzzy integral algorithm on SEED and with about 15% accuracy using multimodal deep neural networks on SEED-IV in comparison with single modality. The best accuracies of 87.59% and 85.11% were achieved on SEED and SEED-IV, respectively. The complementary characteristics between EEG and eye movements for emotion recognition were also discussed. EEG modality had the advantage of recognizing happy emotion, while eye movements performed much better for classifying neutral and sad emotions.

In Chapter 6, transfer learning algorithms were introduced to deal with the challenges of individual differences across subjects and non-stationary characteristics of EEG. Several popular transfer learning methods were adopted to recycle the previously collected labeled data from a group of subjects and make inference on new data from target subjects. The experimental results demonstrated the efficiency of transfer learning in tackling the subject-to-subject

variability. The transductive parameter transfer algorithm achieved the best performance with 76.31% accuracy and a significant improvement of about 20% accuracy in comparison with the baseline without domain adaptation. The heterogeneous knowledge transfer from eye tracking to EEG was also explored. Homogeneous transfer learning still needs time-consuming and high cost for collecting unlabeled EEG data. For heterogeneous knowledge transfer, the scanpath sequences were encoded and compared as domain discrepancy to measure the similarity of EEG feature distributions from different subjects in the calibration phase. Collecting eye tracking data is much easier for calibration and the adaptive models are still able to leverage the representational property of EEG.

In Chapter 7, mutimodal vigilance estimation with EEG and forehead EOG in both laboratory stimulations and real driving environments was presented. In our study, the vigilance estimation problem was modeled as a regression rather than classification problem. The details of data processing and feature extraction of EEG and forehead EOG were described. In order to capture vigilance dynamics with temporal evolutions, the temporal dependency models were introduced. We first performed the evaluations of the computational models on the public vigilance estimation dataset called SEED-VIG, where the EEG and EOG data were recorded by commercial wet electrodes. The experimental results indicated the superiority of modality fusion over single modality and the complementary characteristics of EEG and forehead EOG were demonstrated. The forehead setup achieved the mean correlation coefficient of 0.85. In order to improve the wearability and feasibility of vigilance estimation, we developed a wearable device with flexible dry electrodes and amplifier circuits for forehead EOG recordings. The experimental results on both laboratory driving simulations and real-world driving environments demonstrated the efficiency of the proposed vigilance estimation systems.

8.2 Future Work

Although much progresses have been achieved in affective brain-computer interactions over the past decades, many open questions remain unanswered. Some possible directions for future research include:

- Nowadays, there are many mental disorders that are associated with emotion processing, e.g., anxiety and depression. However, little is known about the underlying neural mechanism of emotion processing in the brain. Emotion research across different subjects should cooperate through analysis of neural and behavior data and find the treatments for mood disorders. Maybe, we will have an efficient approach to activating or suppressing

particular neural circuits for mood disorder treatments.

- The closed-loop brain-computer interactions consist of both the detection and the stimulations. In this work, we mainly focus on the forward detection phase. Efficient neurofeedbacks or neural circuit manipulations should be introduced for emotion regulation. And the assessment about the efficiency of feedbacks can be also studied.
- The applications of affective brain-computer interactions to real scenarios should be further investigated. The application scenarios can be extended to various areas like games, MOOC, and robots. Current studies are mainly performed and evaluated in laboratory controlled settings. The generalizations from stimulations to real scenarios need further investigations. There are many different characteristics of simulated and real scenarios, including adverse environments, individual variations, effects of chronic and acute fatigue, stress, anxiety and so on. How to fill the reality gap between the simulations and real scenarios is an interesting research topic. The performance with recent new devices in real scenarios should be evaluated.
- The subject groups are mainly young students and the performance evaluations with more age groups should be further conducted. Moreover, the gender and cultural differences in emotion recognition are also interesting. Facial expressions show commonality across various cultures. Whether there are any stable neural patterns of individual emotions across cultures, e.g., Chinese, Japanese, and English. The systemical experiments and comparisons can be conducted.
- This work utilizes only a subset of possible modalities for emotion recognition and vigilance estimation. More sensor modalities can be introduced and the interactions between each other can be investigated. How to efficiently combine various modalities with different sampling rates in different levels still remains an open question.
- The emotion labels are annotated into several discrete categories in this work, which does not consider the temporal dynamic of emotion over time. How to accurately recognize emotions in continuous scale and perform emotion annotations in higher temporal resolutions is one of the key problems that need to be addressed.
- The current experimental protocols are isolated from social context. The subjects are exposed passively to stimuli and do not actively interact with environments. In the future, the experiments can be improved with introducing social interactions, for example, more participants together in some social context. Recent developed virtual reality technologies can also help improve the emotion experiments in a natural way.

References

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2012: 1097–1105.

[2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436–444.

[3] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484–489.

[4] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of Go without human knowledge[J]. Nature, 2017, 550(7676): 354–359.

[5] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision (IJCV), 2015, 115(3): 211–252. DOI: 10.1007/s11263-015-0816-y.

[6] HE K, ZHANG X, REN S, et al. Delving Deep Into Rectifiers: Surpassing Human-Level Performance On Imagenet Classification[C]// IEEE International Conference on Computer Vision. [S.l.]: [s.n.], 2015: 1026–1034.

[7] DOLAN R J. Emotion, Cognition, and Behavior[J]. Science, 2002, 298(5596): 1191.

[8] SALOVEY P, MAYER J D. Emotional Intelligence[J]. Imagination Cognition & Personality, 1989, 9(6): 217–236.

[9] GROSS J J, BARRETT L F. Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View[J]. Emotion Review Journal of the International Society for Research on Emotion, 2011, 3(1): 8.

[10] PICARD R W. Affective Computing[M]. Cambridge, MA, USA: MIT Press, 1997. ISBN: 0-262-16170-2.

[11] PICARD R W, VYZAS E, HEALEY J. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 23. 10. [S.l.]: [s.n.], 2001: 1175–1191.

[12] PICARD R W. Affective Computing: From Laughter to IEEE[J]. IEEE Transactions on Affective Computing, 2010, 1(1): 11–17.

[13] D'MELLO S K, KORY J. A Review and Meta-Analysis of Multimodal Affect Detection Systems[J]. Acm Computing Surveys, 2015, 47(3): 1–36.

[14] MAUSS I B, ROBINSON M D. Measures of emotion: A review[J]. Cognition and Emotion, 2009, 23(2): 209–237.

[15] CALVO R A, D'MELLO S. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications[J]. IEEE Transactions on Affective Computing, 2010, 1(1): 18–37.

[16] SITARAM R, ZHANG H, GUAN C, et al. Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain–computer interface[J]. NeuroImage, 2007, 34(4): 1416–1427.

[17] LANCE B J, KERICK S E, RIES A J, et al. Brain–computer interface technologies in the coming decades[J]. Proceedings of the IEEE, 2012, 100(Special Centennial Issue): 1585–1599.

[18] BRUNNER C, BLANKERTZ B, CINCOTTI F, et al. BNCI horizon 2020–towards a roadmap for brain/neural computer interaction[C]// International Conference on Universal Access in Human-Computer Interaction. Springer. [S.l.]: [s.n.], 2014: 475–486.

[19] NICOLELIS M A, LEBEDEV M A. Principles of neural ensemble physiology underlying the operation of brain–machine interfaces[J]. Nature Reviews Neuroscience, 2009, 10(7): 530–540.

[20] HOCHBERG L R, BACHER D, JAROSIEWICZ B, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm[J]. Nature, 2012, 485(7398): 372–375.

[21] CAPOGROSSO M, MILEKOVIC T, BORTON D, et al. A brain-spine interface alleviating gait deficits after spinal cord injury in primates.[J]. Nature, 2016, 539(7628): 284–288.

[22] YUAN H, LIU T, SZARKOWSKI R, et al. Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: an EEG and fMRI study of motor imagery and movements[J]. Neuroimage, 2010, 49(3): 2596–2606.

[23] LOTTE F, CONGEDO M, LÉCUYER A, et al. A review of classification algorithms for EEG-based brain–computer interfaces[J]. Journal of Neural Engineering, 2007, 4(2): R1.

[24] FERNÁNDEZ-RODRÍGUEZ Á, VELASCO-ÁLVAREZ F, RON-ANGEVIN R. Review of real brain-controlled wheelchairs[J]. Journal of Neural Engineering, 2016, 13(6): 061001.

[25] ZANDER T O, KOTHE C. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general[J]. Journal of Neural Engineering, 2011, 8(2): 025005.

[26] NIJBOER F, MORIN F O, CARMIEN S P, et al. Affective brain-computer interfaces: Psychophysiological markers of emotion in healthy persons and in persons with amyotrophic lateral sclerosis[C]// International Conference on Affective Computing and Intelligent Interaction and Workshops. IEEE. [S.l.]: [s.n.], 2009: 1–11.

[27] MÜHL C, ALLISON B, NIJHOLT A, et al. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges[J]. Brain-Computer Interfaces, 2014, 1(2): 66–84.

[28] AL-KAYSI A M, AL-ANI A, LOO C K, et al. Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification[J]. Journal of Affective Disorders, 2017, 208: 597–603.

[29] BOCHAROV A V, KNYAZEV G G, SAVOSTYANOV A N. Depression and implicit emotion processing: An EEG study[J]. Neurophysiologie Clinique/Clinical Neurophysiology, 2017.

[30] TANAKA T, YAMAMOTO T, HARUNO M. Brain response patterns to economic inequity predict present and future depression indices[J]. Nature Human Behaviour, 2017: 748 C756.

[31] HIGGINS J S, MICHAEL J, AUSTIN R, et al. Asleep at the Wheel The Road to Addressing Drowsy Driving[J]. Sleep, 2017, 40(2): zsx001.

[32] GRIER R A, WARM J S, DEMBER W N, et al. The vigilance decrement reflects limitations in effortful attention, not mindlessness[J]. Human Factors, 2003, 45(3): 349–359.

[33] PEIRIS M T, DAVIDSON P R, BONES P J, et al. Detection of lapses in responsiveness from the EEG[J]. Journal of Neural Engineering, 2011, 8(1): 016003.

[34] DAVIDSON P R, JONES R D, PEIRIS M T. EEG-based lapse detection with high temporal resolution[J]. IEEE Transactions on Biomedical Engineering, 2007, 54(5): 832–839.

[35] MARTEL A, DÄHNE S, BLANKERTZ B. EEG predictors of covert vigilant attention[J]. Journal of Neural Engineering, 2014, 11(3): 035009.

[36] ZHENG W.-L, LU B.-L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks[J]. IEEE Transactions on Autonomous Mental Development, 2015, 7(3): 162–175.

[37] JAMES W. What is an emotion?[J]. Mind, 1884, 9(34): 188–205.

[38] SCHERER K R. What are emotions? And how can they be measured?[J]. Social Science Information, 2005, 44(4): 695–729.

[39] KIM J, ANDRÉ E. Emotion recognition based on physiological changes in music listening[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(12): 2067–2083.

[40] ADOLPHS R. Neural systems for recognizing emotion[J]. Current Opinion in Neurobiology, 2002, 12(2): 169–177.

[41] KLEINGINNA P R, KLEINGINNA A M. A categorized list of emotion definitions, with suggestions for a consensual definition[J]. Motivation and Emotion, 1981, 5(4): 345–379.

[42] GROSS J J. Handbook of emotion regulation[M]. [S.l.]: Guilford Publications, 2013.

[43] BARRETT L F. Are emotions natural kinds?[J]. Perspectives on Psychological Science, 2006, 1(1): 28–58.

[44] BAVEYE Y, CHAMARET C, DELLANDRÉA E, et al. Affective Video Content Analysis: A Multidisciplinary Insight[J]. IEEE Transactions on Affective Computing, 2017.

[45] EKMAN P, FRIESEN W V. Constants across cultures in the face and emotion[J]. Journal of Personality and Social Psychology, 1971, 17(2): 124.

[46] EKMAN P. An argument for basic emotions[J]. Cognition & Emotion, 1992, 6(3-4): 169–200.

[47] PLUTCHIK R. A general psychoevolutionary theory of emotion[J]. Theories of Emotion, 1980, 1(3-31): 4.

[48] MOHAMMAD S M, TURNEY P D. Crowdsourcing a word–emotion association lexicon[J]. Computational Intelligence, 2013, 29(3): 436–465.

[49] OLSON D H, SPRENKLE D H, RUSSELL C S. Circumplex model of marital and family systems: I. Cohesion and adaptability dimensions, family types, and clinical applications[J]. Family Process, 1979, 18(1): 3–28.

[50] RUSSELL J A, MEHRABIAN A. Evidence for a three-factor theory of emotions[J]. Journal of Research in Personality, 1977, 11(3): 273–294.

[51] BAVEYE Y, DELLANDREA E, CHAMARET C, et al. Liris-accede: A video database for affective content analysis[J]. IEEE Transactions on Affective Computing, 2015, 6(1): 43–55.

[52] LIU Y, SOURINA O. EEG databases for emotion recognition[C]// International Conference on Cyberworlds. IEEE. [S.l.]: [s.n.], 2013: 302–309.

[53] FONTAINE J R, SCHERER K R, ROESCH E B, et al. The world of emotions is not two-dimensional[J]. Psychological Science, 2007, 18(12): 1050–1057.

[54] COWEN A S, KELTNER D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients[J]. Proceedings of the National Academy of Sciences, 2017, 114(38): E7900–E7909.

[55] KASSAM K S, MARKEY A R, CHERKASSKY V L, et al. Identifying emotions on the basis of neural activation[J]. PloS One, 2013, 8(6): e66032.

[56] DAVIDSON R J, FOX N A. Asymmetrical Brain Activity Discriminates between Positive and Negative[J]. Science, 1981, 211: 840.

[57] DAVIDSON R J. Anterior cerebral asymmetry and the nature of emotion[J]. Brain and Cognition, 1992, 20(1): 125–151.

[58] MAUSS I B, ROBINSON M D. Measures of emotion: A review[J]. Cognition and Emotion, 2009, 23(2): 209–237.

[59] DALGLEISH T. The emotional brain[J]. Nature Reviews Neuroscience, 2004, 5(7): 583–589.

[60] DARWIN C. The expression of the emotions in man and animals[M]. Vol. 526. [S.l.]: University of Chicago press, 1965.

[61] CANNON W B. The James-Lange theory of emotions: A critical examination and an alternative theory[J]. The American Journal of Psychology, 1927, 39(1/4): 106–124.

[62] CANNON W B. Again the James-Lange and the thalamic theories of emotion.[J]. Psychological Review, 1931, 38(4): 281.

[63] PAPEZ J W. A proposed mechanism of emotion[J]. Archives of Neurology & Psychiatry, 1937, 38(4): 725–743.

[64] MACLEAN P D. Psychosomatic Disease and the" Visceral Brain": Recent Developments Bearing on the Papez Theory of Emotion.[J]. Psychosomatic Medicine, 1949, 11(6): 338–353.

[65] LEDOUX J. The emotional brain: The mysterious underpinnings of emotional life[M]. [S.l.]: Simon, Schuster, 1998.

[66] ETKIN A, BÜCHEL C, GROSS J J. The neural bases of emotion regulation[J]. Nature Reviews Neuroscience, 2015, 16(11): 693–700.

[67] PESSOA L. A Network Model of the Emotional Brain[J]. Trends in Cognitive Sciences, 2017.

[68] MALHI G S, BYROW Y, FRITZ K, et al. Mood disorders: neurocognitive models[J]. Bipolar Disorders, 2015, 17(S2): 3–20.

[69] LEDOUX J E. Emotion circuits in the brain[J]. Annual Review of Neuroscience, 2000, 23(1): 155–184.

[70] JANAK P H, TYE K M. From circuits to behaviour in the amygdala[J]. Nature, 2015, 517(7534): 284.

[71] TYE K M, PRAKASH R, KIM S.-Y, et al. Amygdala circuitry mediating reversible and bidirectional control of anxiety[J]. Nature, 2011, 471(7338): 358.

[72] PHAN K L, WAGER T, TAYLOR S F, et al. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI[J]. Neuroimage, 2002, 16(2): 331–348.

[73] NUMMENMAA L, GLEREAN E, VIINIKAINEN M, et al. Emotions promote social interaction by synchronizing brain activity across individuals[J]. Proceedings of the National Academy of Sciences, 2012, 109(24): 9599–9604.

[74] KRAGEL P A, KNODT A R, HARIRI A R, et al. Decoding spontaneous emotional states in the human brain[J]. PLoS Biology, 2016, 14(9): e2000106.

[75] VUILLEUMIER P, POURTOIS G. Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging[J]. Neuropsychologia, 2007, 45(1): 174–194.

[76] SAARIMÄKI H, GOTSOPOULOS A, JÄÄSKELÄINEN I P, et al. Discrete neural signatures of basic emotions[J]. Cerebral Cortex, 2015, 26(6): 2563–2573.

[77] COHEN M X. Where does EEG come from and what does it mean?[J]. Trends in Neurosciences, 2017, 40(4): 208–218.

[78] BUZSÁKI G, ANASTASSIOU C A, KOCH C. The origin of extracellular fields and currents EEG, ECoG, LFP and spikes[J]. Nature Reviews Neuroscience, 2012, 13(6): 407.

[79] HUANG Y.-J, WU C.-Y, WONG A M.-K, et al. Novel active comb-shaped dry electrode for EEG measurement in hairy site[J]. IEEE Transactions on Biomedical Engineering, 2015, 62(1): 256–263.

[80] GROZEA C, VOINESCU C D, FAZLI S. Bristle-sensors low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications[J]. Journal of Neural Engineering, 2011, 8(2): 025008.

[81] HAIRSTON W D, WHITAKER K W, RIES A J, et al. Usability of four commercially-oriented EEG systems[J]. Journal of Neural Engineering, 2014, 11(4): 046018.

[82] MICHEL C M, MURRAY M M, LANTZ G, et al. EEG source imaging[J]. Clinical Neurophysiology, 2004, 115(10): 2195–2222.

[83] ACAR Z A, MAKEIG S. Effects of forward model errors on EEG source localization[J]. Brain Topography, 2013, 26(3): 378–396.

[84] BISWAL S, KULAS J, SUN H, et al. SLEEPNET: Automated Sleep Staging System via Deep Learning[J]. ArXiv preprint arXiv:1707.08262, 2017.

[85] HUANG C.-S, LIN C.-L, KO L.-W, et al. Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels[J]. Frontiers in Neuroscience, 2014, 8: 263.

[86] RAY W J, COLE H W. EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes[J]. Science, 1985, 228(4700): 750–752.

[87] GROSS J J, LEVENSON R W. Emotion elicitation using films[J]. Cognition and Emotion, 1995, 9(1): 87–108.

[88] SCHAEFER A, NILS F, SANCHEZ X, et al. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers[J]. Cognition and Emotion, 2010, 24(7): 1153–1172.

[89] LANG P J, BRADLEY M M, CUTHBERT B N, et al. International affective picture system (IAPS): Instruction manual and affective ratings[J]. The Center for Research in Psychophysiology, University of Florida, 1999.

[90] BRADLEY M, LANG P J. The International affective digitized sounds (IADS)[: stimuli, instruction manual and affective ratings[M]. [S.l.]: NIMH Center for the Study of Emotion, Attention, 1999.

[91] KRUMHANSL C L. An exploratory study of musical emotions and psychophysiology.[J]. Canadian Journal of Experimental Psychology, 1997, 51(4): 336.

[92] PHILIPPOT P. Inducing and assessing differentiated emotion-feeling states in the laboratory[J]. Cognition and Emotion, 1993, 7(2): 171–193.

[93] TOMARKEN A J, DAVIDSON R J, HENRIQUES J B. Resting frontal brain asymmetry predicts affective responses to films.[J]. Journal of Personality and Social Psychology, 1990, 59(4): 791.

[94] IZARD C E. The Differential Emotions Scale: DES IV-A;[a Method of Measuring the Meaning of Subjective Experience of Discrete Emotions][M]. [S.l.]: University of Delaware, 1993.

[95] WATSON D, CLARK L A, TELLEGEN A. Development and validation of brief measures of positive and negative affect: the PANAS scales.[J]. Journal of Personality and Social Psychology, 1988, 54(6): 1063.

[96] DOUGLAS-COWIE E, COWIE R, SNEDDON I, et al. The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data[C]// International Conference on Affective Computing and Intelligent Interaction. Springer. [S.l.]: [s.n.], 2007: 488–500.

[97] KOELSTRA S, MUHL C, SOLEYMANI M, et al. DEAP: A database for emotion analysis; using physiological signals[J]. IEEE Transactions on Affective Computing, 2012, 3(1): 18–31.

[98] SOLEYMANI M, LICHTENAUER J, PUN T, et al. A multimodal database for affect recognition and implicit tagging[J]. IEEE Transactions on Affective Computing, 2012, 3(1): 42–55.

[99] CARVALHO S, LEITE J, GALDO-ÁLVAREZ S, et al. The emotional movie database (EMDB): A self-report and psychophysiological study[J]. Applied Psychophysiology and Biofeedback, 2012, 37(4): 279–294.

[100] DEMARTY C.-H, PENET C, GRAVIER G, et al. A benchmarking campaign for the multimodal detection of violent scenes in movies[C]// European Conference on Computer Vision. Springer. [S.l.]: [s.n.], 2012: 416–425.

[101] BRADLEY M M, LANG P J. Measuring emotion: the self-assessment manikin and the semantic differential[J]. Journal of Behavior Therapy and Experimental Psychiatry, 1994, 25(1): 49–59.

[102] GROSS J J. The emerging field of emotion regulation: an integrative review.[J]. Review of General Psychology, 1998, 2(3): 271.

[103] ALARCAO S M, FONSECA M J. Emotions recognition using EEG signals: a survey[J]. IEEE Transactions on Affective Computing, 2017.

[104] KIM M.-K, KIM M, OH E, et al. A review on the computational methods for emotional state estimation from the human EEG[J]. Computational and Mathematical Methods in Medicine, 2013, 2013.

[105] JENKE R, PEER A, BUSS M. Feature extraction and selection for emotion recognition from EEG[J]. IEEE Transactions on Affective Computing, 2014, 5(3): 327–339.

[106] WANG X.-W, NIE D, LU B.-L. Emotional state classification from EEG data using machine learning approach[J]. Neurocomputing, 2014, 129: 94–106.

[107] BOS D O. EEG-based emotion recognition[J]. The Influence of Visual and Auditory Stimuli, 2006: 1–17.

[108] HERAZ A, FRASSON C. Predicting the three major dimensions of the learner s emotions from brainwaves[J]. International Journal of Computer Science, 2007, 2(3): 187–193.

[109] MURUGAPPAN M, RAMACHANDRAN N, SAZALI Y. Classification of human emotion from EEG using discrete wavelet transform.[J]. Journal of Biomedical Science and Engineering, 2010, 2(4).

[110] LIN Y.-P, WANG C.-H, JUNG T.-P, et al. EEG-based emotion recognition in music listening[J]. IEEE Transactions on Biomedical Engineering, 2010, 57(7): 1798–1806.

[111] BROWN L, GRUNDLEHNER B, PENDERS J. Towards wireless emotional valence detection from EEG[C]// Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. [S.l.]: [s.n.], 2011: 2188–2191.

[112] PETRANTONAKIS P C, HADJILEONTIADIS L J. A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition[J]. IEEE Transactions on Information Technology in Biomedicine, 2011, 15(5): 737–746.

[113] HADJIDIMITRIOU S K, HADJILEONTIADIS L J. Toward an EEG-based recognition of music liking using time-frequency analysis[J]. IEEE Transactions on Biomedical Engineering, 2012, 59(12): 3498–3510.

[114] WANG X.-W, NIE D, LU B.-L. Emotional state classification from EEG data using machine learning approach[J]. Neurocomputing, 2014, 129: 94–106.

[115] DALY I, MALIK A, HWANG F, et al. Neural correlates of emotional responses to music: an EEG study[J]. Neuroscience Letters, 2014, 573: 52–57.

[116] LIU Y.-J, YU M, ZHAO G, et al. Real-time movie-induced discrete emotion recognition from EEG signals[J]. IEEE Transactions on Affective Computing, 2017.

[117] LEE Y.-Y, HSIEH S. Classifying different emotional states by means of EEG-based functional connectivity patterns[J]. PloS One, 2014, 9(4): e95415.

[118] VALENZI S, ISLAM T, JURICA P, et al. Individual classification of emotions using EEG[J]. Journal of Biomedical Science and Engineering, 2014, 7(08): 604.

[119] LI M, LU B.-L. Emotion classification based on gamma-band EEG[C]// Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. [S.l.]: [s.n.], 2009: 1223–1226.

[120] MÜLLER M M, KEIL A, GRUBER T, et al. Processing of affective pictures modulates right-hemispheric gamma band EEG activity[J]. Clinical Neurophysiology, 1999, 110(11): 1913–1920.

[121] NIE D, WANG X.-W, SHI L.-C, et al. EEG-based emotion recognition during watching movies[C]// 5th International IEEE/EMBS Conference on Neural Engineering. IEEE. [S.l.]: [s.n.], 2011: 667–670.

[122] BALCONI M, BRAMBILLA E, FALBO L. Appetitive vs. defensive responses to emotional cues. Autonomic measures and brain oscillation modulation[J]. Brain Research, 2009, 1296: 72–84.

[123] SATO W, KOCHIYAMA T, YOSHIKAWA S, et al. Emotional expression boosts early visual processing of the face: ERP recording and its decomposition by independent component analysis[J]. Neuroreport, 2001, 12(4): 709–714.

[124] BALCONI M, LUCCHIARI C. Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A gamma band analysis[J]. International Journal of Psychophysiology, 2008, 67(1): 41–46.

[125] MCEVOY L, SMITH M, GEVINS A. Test–retest reliability of cognitive EEG[J]. Clinical Neurophysiology, 2000, 111(3): 457–463.

[126] ALLEN J J, URRY H L, HITT S K, et al. The stability of resting frontal electroencephalographic asymmetry in depression[J]. Psychophysiology, 2004, 41(2): 269–280.

[127] LAN Z, SOURINA O, WANG L, et al. Stability of features in real-time EEG-based emotion recognition algorithm[C]// International Conference on Cyberworlds. IEEE. [S.l.]: [s.n.], 2014: 137–144.

[128] LIN Y.-P, HSU S.-H, JUNG T.-P. Exploring day-to-day variability in the relations between emotion and EEG signals[C]// International Conference on Augmented Cognition. Springer. [S.l.]: [s.n.], 2015: 461–469.

[129] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345–1359.

[130] JAYARAM V, ALAMGIR M, ALTUN Y, et al. Transfer learning in brain-computer interfaces[J]. IEEE Computational Intelligence Magazine, 2016, 11(1): 20–31.

[131] SAMEK W, MEINECKE F C, MÜLLER K.-R. Transferring subspaces between subjects in brain–computer interfacing[J]. IEEE Transactions on Biomedical Engineering, 2013, 60(8): 2289–2298.

[132] CALVO R A, D'MELLO S. Affect detection: An interdisciplinary review of models, methods, and their applications[J]. IEEE Transactions on Affective Computing, 2010, 1(1): 18–37.

[133] RAVÌ D, WONG C, DELIGIANNI F, et al. Deep learning for health informatics[J]. IEEE Journal of Biomedical and Health Informatics, 2017, 21(1): 4–21.

[134] D'MELLO S K, KORY J. A review and meta-analysis of multimodal affect detection systems[J]. ACM Computing Surveys, 2015, 47(3): 43.

[135] KOELSTRA R A L. Affective and Implicit Tagging using Facial Expressions and Electroencephalography.[D]. Queen Mary University of London, 2012.

[136] NICOLAOU M A, GUNES H, PANTIC M. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space[J]. IEEE Transactions on Affective Computing, 2011, 2(2): 92–105.

[137] LIN Y.-P, YANG Y.-H, JUNG T.-P. Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening[J]. Frontiers in Neuroscience, 2014, 8: 94.

[138] WAGNER J, ANDRE E, LINGENFELSER F, et al. Exploring fusion methods for multimodal emotion recognition with missing data[J]. IEEE Transactions on Affective Computing, 2011, 2(4): 206–218.

[139] TSALAMLAL M Y, AMORIM M.-A, MARTIN J.-C, et al. Combining facial expression and touch for perceiving emotional valence[J]. IEEE Transactions on Affective Computing, 2016.

[140] SCHIRMER A, ADOLPHS R. Emotion perception from face, voice, and touch: comparisons and convergence[J]. Trends in Cognitive Sciences, 2017, 21(3): 216–228.

[141] NGIAM J, KHOSLA A, KIM M, et al. Multimodal deep learning[C]// International Conference on Machine Learning. [S.l.]: [s.n.], 2011: 689–696.

[142] TZIRAKIS P, TRIGEORGIS G, NICOLAOU M A, et al. End-to-end multimodal emotion recognition using deep neural networks[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(8): 1301–1309.

[143] VALTCHANOV D, HANCOCK M. EnviroPulse: providing feedback about the expected affective valence of the environment[C]// ACM Conference on Human Factors in Computing Systems. ACM. [S.l.]: [s.n.], 2015: 2073–2082.

[144] MICHAL P, PAWEL D, WENHAN S, et al. Towards context aware emotional intelligence in machines: computing contextual appropriateness of affective states[C]// International Joint Conference on Artificial Intelligence. AAAI. [S.l.]: [s.n.], 2009: 1469–1474.

[145] BULLING A, WARD J A, GELLERSEN H, et al. Eye movement analysis for activity recognition using electrooculography[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(4): 741–753.

[146] DUCHOWSKI A T. Eye tracking methodology[J]. Theory and Practice, 2007, 328.

[147] HENDERSON J M, SHINKAREVA S V, WANG J, et al. Predicting cognitive state from eye movements[J]. PloS One, 2013, 8(5): e64937.

[148] BUSCHER G, DENGEL A, BIEDERT R, et al. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond[J]. ACM Transactions on Interactive Intelligent Systems, 2012, 1(2): 9.

[149] PARTALA T, JOKINIEMI M, SURAKKA V. Pupillary responses to emotionally provocative stimuli[C]// The Symposium on Eye Tracking Research & Applications. ACM. [S.l.]: [s.n.], 2000: 123–129.

[150] BRADLEY M M, MICCOLI L, ESCRIG M A, et al. The pupil as a measure of emotional arousal and autonomic activation[J]. Psychophysiology, 2008, 45(4): 602–607.

[151] ZEKVELD A A, HESLENFELD D J, JOHNSRUDE I S, et al. The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load[J]. Neuroimage, 2014, 101: 76–86.

[152] LANGER N, HO E J, ALEXANDER L M, et al. A resource for assessing information processing in the developing brain using EEG and eye tracking[J]. Scientific Data, 2017, 4: 170040.

[153] WANG C, ONG J L, PATANAIK A, et al. Spontaneous eyelid closures link vigilance fluctuation with fMRI dynamic connectivity states[J]. Proceedings of the National Academy of Sciences, 2016, 113(34): 9653–9658.

[154] LAENG B, SIROIS S, GREDEBÄCK G. Pupillometry: a window to the preconscious?[J]. Perspectives on Psychological Science, 2012, 7(1): 18–27.

[155] MURPHY P R, O'CONNELL R G, O'SULLIVAN M, et al. Pupil diameter covaries with BOLD activity in human locus coeruleus[J]. Human Brain Mapping, 2014, 35(8): 4140–4154.

[156] KRAFKA K, KHOSLA A, KELLNHOFER P, et al. Eye Tracking for Everyone[C]// IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: [s.n.], 2016: 2176–2184.

[157] JI Q, ZHU Z, LAN P. Real-time nonintrusive monitoring and prediction of driver fatigue[J]. IEEE Transactions on Vehicular Technology, 2004, 53(4): 1052–1068.

[158] OKEN B, SALINSKY M, ELSAS S. Vigilance, alertness, or sustained attention: physiological basis and measurement[J]. Clinical Neurophysiology, 2006, 117(9): 1885–1901.

[159] CHUA E C.-P, TAN W.-Q, YEO S.-C, et al. Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation[J]. Sleep, 2012, 35(3): 325–334.

[160] DONG Y, HU Z, UCHIMURA K, et al. Driver inattention monitoring system for intelligent vehicles: A review[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(2): 596–614.

[161] DINGES D F, GRACE R. PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance[J]. US Department of Transportation, Federal Highway Administration, Publication Number FHWA-MCRT-98-006, 1998.

[162] LIN C.-T, WU R.-C, LIANG S.-F, et al. EEG-based drowsiness estimation for safety driving using independent component analysis[J]. IEEE Transactions on Circuits and Systems, 2005, 52(12): 2726–2738.

[163] KHUSHABA R N, KODAGODA S, LAL S, et al. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm[J]. IEEE Transactions on Biomedical Engineering, 2011, 58(1): 121–131.

[164] SHI L.-C, LU B.-L. EEG-based vigilance estimation using extreme learning machines[J]. Neurocomputing, 2013, 102: 135–143.

[165] QIAN D, WANG B, QING X, et al. Drowsiness detection by Bayesian-copula discriminant classifier based on EEG signals during daytime short nap[J]. IEEE Transactions on Biomedical Engineering, 2017, 64(4): 743–754.

[166] OLBRICH S, MULERT C, KARCH S, et al. EEG-vigilance and BOLD effect during simultaneous EEG/fMRI measurement[J]. Neuroimage, 2009, 45(2): 319–332.

[167] DE GENNARO L, VECCHIO F, FERRARA M, et al. Antero-posterior functional coupling at sleep onset: changes as a function of increased sleep pressure[J]. Brain Research Bulletin, 2005, 65(2): 133–140.

[168] LIN C.-T, HUANG K.-C, CHUANG C.-H, et al. Can arousing feedback rectify lapses in driving? Prediction from EEG power spectra[J]. Journal of Neural Engineering, 2013, 10(5): 056024.

[169] ROSENBERG M D, FINN E S, SCHEINOST D, et al. A neuromarker of sustained attention from whole-brain functional connectivity[J]. Nature Neuroscience, 2016, 19(1): 165.

[170] O'CONNELL R G, DOCKREE P M, ROBERTSON I H, et al. Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20 s before they occur[J]. Journal of Neuroscience, 2009, 29(26): 8604–8611.

[171] LIN C.-T, CHEN Y.-C, HUANG T.-Y, et al. Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver's drowsiness detection and warning[J]. IEEE Transactions on Biomedical Engineering, 2008, 55(5): 1582–1591.

[172] LIN C.-T, CHUANG C.-H, HUANG C.-S, et al. Wireless and wearable EEG system for evaluating driver vigilance[J]. IEEE Transactions on Biomedical Circuits and Systems, 2014, 8(2): 165–176.

[173] HA S, KIM C, CHI Y M, et al. Integrated circuits and electrode interfaces for noninvasive physiological monitoring[J]. IEEE Transactions on Biomedical Engineering, 2014, 61(5): 1522–1537.

[174] MULLEN T R, KOTHE C A, CHI Y M, et al. Real-time neuroimaging and cognitive monitoring using wearable dry EEG[J]. IEEE Transactions on Biomedical Engineering, 2015, 62(11): 2553–2567.

[175] PAPADELIS C, CHEN Z, KOURTIDOU-PAPADELI C, et al. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents[J]. Clinical Neurophysiology, 2007, 118(9): 1906–1922.

[176] CAI H.-Y, MA J.-X, SHI L.-C, et al. A novel method for EOG features extraction from the forehead[C]// Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. [S.l.]: [s.n.], 2011: 3075–3078.

[177] MA J.-X, SHI L.-C, LU B.-L. An EOG-based vigilance estimation method applied for driver fatigue detection[J]. Neuroscience and Biomedical Engineering, 2014, 2(1): 41–51.

[178] DAMOUSIS I G, TZOVARAS D. Fuzzy fusion of eyelid activity indicators for hypovigilance-related accident prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2008, 9(3): 491–500.

[179] ZHANG L, WADE J, BIAN D, et al. Cognitive load measurement in a virtual reality-based driving system for autism intervention[J]. IEEE Transactions on Affective Computing, 2017, 8(2): 176–189.

[180] AHN S, NGUYEN T, JANG H, et al. Exploring neuro-physiological correlates of drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data[J]. Frontiers in Human Neuroscience, 2016, 10: 219.

[181] NGUYEN T, AHN S, JANG H, et al. Utilization of a combined EEG/NIRS system to predict driver drowsiness[J]. Scientific Reports, 2017, 7: 43933.

[182] HEALEY J A, PICARD R W. Detecting stress during real-world driving tasks using physiological sensors[J]. IEEE Transactions on Intelligent Transportation Systems, 2005, 6(2): 156–166.

[183] WANG H, ZHANG C, SHI T, et al. Real-time EEG-based detection of fatigue driving danger for accident prediction[J]. International Journal of Neural Systems, 2015, 25(02): 1550002.

[184] ZHENG W.-L, ZHU J.-Y, LU B.-L. Identifying stable patterns over time for emotion recognition from EEG[J]. IEEE Transactions on Affective Computing, 2017.

[185] ZHENG W.-L, LIU W, LU Y, et al. EmotionMeter: A Multimodal Framework for Recognizing Human Emotions[J]. IEEE Transactions on Cybernetics, 2018.

[186] ZHENG W.-L, LU B.-L. A multimodal approach to estimating vigilance using EEG and forehead EOG[J]. Journal of Neural Engineering, 2017, 14(2): 026017.

[187] EYSENCK S B, EYSENCK H J, BARRETT P. A revised version of the psychoticism scale[J]. Personality and Individual Differences, 1985, 6(1): 21–29.

[188] SHI L.-C, JIAO Y.-Y, LU B.-L. Differential entropy feature for EEG-based vigilance estimation[C]// 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. [S.l.]: [s.n.], 2013: 6627–6630.

[189] DUAN R.-N, ZHU J.-Y, LU B.-L. Differential entropy feature for EEG-based emotion classification[C]// 6th International IEEE/EMBS Conference on Neural Engineering. IEEE. [S.l.]: [s.n.], 2013: 81–84.

[190] GIBBS J W. Elementary principles in statistical mechanics[M]. [S.l.]: Courier Corporation, 2014.

[191] SHI L.-C, LU B.-L. Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning[C]// Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. [S.l.]: [s.n.], 2010: 6587–6590.

[192] DUAN R.-N, WANG X.-W, LU B.-L. EEG-based emotion recognition in listening music by using support vector machine and linear dynamic system[C]// International Conference on Neural Information Processing. Springer. [S.l.]: [s.n.], 2012: 468–475.

[193] CASELLA G, FIENBERG S, OLKIN I. Time Series Analysis and Its Applications - With R Examples[J]. Journal of the American Statistical Association, 2006, 97(458): 656–657.

[194] JAIN A K, DUIN R P W, MAO J. Statistical pattern recognition: A review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(1): 4–37.

[195] PENG H, LONG F, DING C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1226–1238.

[196] PARTALA T, SURAKKA V. Pupil size variation as an indication of affective processing[J]. International Journal of Human-Computer Studies, 2003, 59(1-2): 185–198.

[197] FERRARA M, DE GENNARO L. How much sleep do we need?[J]. Sleep Medicine Reviews, 2001, 5(2): 155–179.

[198] ZHANG Y.-F, GAO X.-Y, ZHU J.-Y, et al. A novel approach to driving fatigue detection using forehead EOG[C]// 7th International IEEE/EMBS Conference on Neural Engineering. IEEE. [S.l.]: [s.n.], 2015: 707–710.

[199] HUO X.-Q, ZHENG W.-L, LU B.-L. Driving fatigue detection with fusion of EEG and forehead EOG[C]// International Joint Conference on Neural Networks. IEEE. [S.l.]: [s.n.], 2016: 897–904.

[200] MAKEIG S, INLOW M. Lapse in alertness: coherence of fluctuations in performance and EEG spectrum[J]. Electroencephalography and Clinical Neurophysiology, 1993, 86(1): 23–35.

[201] WANG Y.-K, JUNG T.-P, LIN C.-T. EEG-based attention tracking during distracted driving[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2015, 23(6): 1085–1094.

[202] LIN C.-T, HUANG K.-C, CHAO C.-F, et al. Tonic and phasic EEG and behavioral changes induced by arousing feedback[J]. NeuroImage, 2010, 52(2): 633–642.

[203] BERGASA L M, NUEVO J, SOTELO M A, et al. Real-time system for monitoring driver vigilance[J]. IEEE Transactions on Intelligent Transportation Systems, 2006, 7(1): 63–77.

[204] TRUTSCHEL U, SIROIS B, SOMMER D, et al. PERCLOS: An alertness measure of the past[J]. 2011.

[205] SHEN S.-C, ZHENG W.-L, LU B.-L. Online object tracking based on depth image with sparse coding[C]// International Conference on Neural Information Processing. Springer. [S.l.]: [s.n.], 2014: 234–241.

[206] ZHENG W.-L, SHEN S.-C, LU B.-L. Online depth image-based object tracking with sparse representation and object detection[J]. Neural Processing Letters, 2017, 45(3): 745–758.

[207] GAO X.-Y, ZHANG Y.-F, ZHENG W.-L, et al. Evaluating driving fatigue detection algorithms using eye tracking glasses[C]// 7th International IEEE/EMBS Conference on Neural Engineering. IEEE. [S.l.]: [s.n.], 2015: 767–770.

[208] OKEN B S, SALINSKY M C. Sleeping and driving: Not a safe dual-task[J]. Clinical Neurophysiology, 2007, 118(9): 1899.

[209] LEE J H, LEE S M, BYEON H J, et al. CNT/PDMS-based canal-typed ear electrodes for inconspicuous EEG recording[J]. Journal of Neural Engineering, 2014, 11(4): 046014.

[210] JUNG H.-C, MOON J.-H, BAEK D.-H, et al. CNT/PDMS composite flexible dry electrodesfor long-term ECG monitoring[J]. IEEE Transactions on Biomedical Engineering, 2012, 59(5): 1472–1479.

[211] LEE S M, KIM J H, PARK C, et al. Self-adhesive and capacitive carbon nanotube-based electrode to record electroencephalograph signals from the hairy scalp[J]. IEEE Transactions on Biomedical Engineering, 2016, 63(1): 138–147.

[212] NATHAN V, JAFARI R. Design principles and dynamic front end reconfiguration for low noise EEG acquisition with finger based dry electrodes[J]. IEEE transactions on Biomedical Circuits and Systems, 2015, 9(5): 631–640.

[213] PINEGGER A, WRIESSNEGGER S C, FALLER J, et al. Evaluation of different EEG acquisition systems concerning their suitability for building a brain–computer interface: case studies[J]. Frontiers in Neuroscience, 2016, 10: 441.

[214] MARTINEZ H P, BENGIO Y, YANNAKAKIS G N. Learning deep physiological models of affect[J]. IEEE Computational Intelligence Magazine, 2013, 8(2): 20–33.

[215] LÄNGKVIST M, KARLSSON L, LOUTFI A. Sleep stage classification using unsupervised feature learning[J]. Advances in Artificial Neural Systems, 2012, 2012.

[216] LI K, LI X, ZHANG Y, et al. Affective state recognition from EEG with deep belief networks[C]// IEEE International Conference on Bioinformatics and Biomedicine. IEEE. [S.l.]: [s.n.], 2013: 305–310.

[217] SCHIRRMEISTER R T, SPRINGENBERG J T, FIEDERER L D J, et al. Deep learning with convolutional neural networks for brain mapping and decoding of movement-related information from the human EEG[J]. ArXiv preprint arXiv:1703.05051, 2017.

[218] BASHIVAN P, RISH I, YEASIN M, et al. Learning representations from EEG with deep recurrent-convolutional neural networks[J]. ArXiv preprint arXiv:1511.06448, 2015.

[219] LI Y, DZIRASA K, CARIN L, et al. Targeting EEG/LFP Synchrony with Neural Nets[C]// Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2017: 4623–4633.

[220] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504–507.

[221] WANG H, RAJ B, XING E P. On the origin of deep learning[J]. ArXiv preprint arXiv:1702.07800, 2017.

[222] WULSIN D, GUPTA J, MANI R, et al. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement[J]. Journal of Neural Engineering, 2011, 8(3): 036015.

[223] CHANG C.-C, LIN C.-J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 27.

[224] ZHENG W.-L, ZHU J.-Y, PENG Y, et al. EEG-based emotion classification using deep belief networks[C]// IEEE International Conference on Multimedia and Expo. IEEE. [S.l.]: [s.n.], 2014: 1–6.

[225] HAYKIN S S, HAYKIN S S, HAYKIN S S, et al. Neural networks and learning machines[M]. Vol. 3. [S.l.]: Pearson Upper Saddle River, NJ, USA: 2009.

[226] SALINSKY M, OKEN B, MOREHEAD L. Test-retest reliability in EEG frequency analysis[J]. Electroencephalography and Clinical Neurophysiology, 1991, 79(5): 382–392.

[227] GUDMUNDSSON S, RUNARSSON T P, SIGURDSSON S, et al. Reliability of quantitative EEG features[J]. Clinical Neurophysiology, 2007, 118(10): 2162–2171.

[228] PENG Y, WANG S, LONG X, et al. Discriminative graph regularized extreme learning machine and its application to face recognition[J]. Neurocomputing, 2015, 149: 340–353.

[229] FAN R.-E, CHANG K.-W, HSIEH C.-J, et al. LIBLINEAR: A library for large linear classification[J]. Journal of Machine Learning Research, 2008, 9(Aug): 1871–1874.

[230] HUANG G.-B, ZHU Q.-Y, SIEW C.-K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1-3): 489–501.

[231] CHUNG S Y, YOON H J. Affective classification using Bayesian classifier and supervised learning[C]// 12th International Conference on Control, Automation and Systems. IEEE. [S.l.]: [s.n.], 2012: 1768–1771.

[232] LIU Y, SOURINA O. Real-time fractal-based valence level recognition from EEG[G]// Transactions on Computational Science XVIII. [S.l.]: Springer, 2013: 101–120.

[233] ZHANG X, HU B, CHEN J, et al. Ontology-based context modeling for emotion recognition in an intelligent web[J]. World Wide Web, 2013, 16(4): 497–513.

[234] YANG Y, WU Q J, ZHENG W.-L, et al. EEG-based emotion recognition using hierarchical network with subnetwork nodes[J]. IEEE Transactions on Cognitive and Developmental Systems, 2017.

[235] TRIPATHI S, ACHARYA S, SHARMA R D, et al. Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset.[C]// AAAI. [S.l.]: [s.n.], 2017: 4746–4752.

[236] KLIMESCH W, DOPPELMAYR M, RUSSEGGER H, et al. Induced alpha band power changes in the human EEG and attention[J]. Neuroscience Letters, 1998, 244(2): 73–76.

[237] LU Y, ZHENG W.-L, LI B, et al. Combining Eye Movements and EEG to Enhance Emotion Recognition[C]// International Joint Conference on Artificial Intelligence. Vol. 15. [S.l.]: [s.n.], 2015: 1170–1176.

[238] MACLEAN D, ROSEWAY A, CZERWINSKI M. MoodWings: a wearable biofeedback device for real-time stress intervention[C]// The 6th International Conference on Pervasive Technologies Related to Assistive Environments. ACM. [S.l.]: [s.n.], 2013: 66.

[239] WILLIAMS M A, ROSEWAY A, O'DOWD C, et al. SWARM: an actuated wearable for mediating affect[C]// The Ninth International Conference on Tangible, Embedded, and Embodied Interaction. ACM. [S.l.]: [s.n.], 2015: 293–300.

[240] HASSIB M, SCHNEEGASS S, EIGLSPERGER P, et al. EngageMeter: A system for implicit audience engagement sensing using electroencephalography[C]// The 2017 CHI Conference on Human Factors in Computing Systems. ACM. [S.l.]: [s.n.], 2017: 5114–5119.

[241] MUROFUSHI T, SUGENO M. An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure[J]. Fuzzy Sets and Systems, 1989, 29(2): 201–227.

[242] LIU W, ZHENG W.-L, LU B.-L. Emotion recognition using multimodal deep learning[C]// International Conference on Neural Information Processing. Springer. [S.l.]: [s.n.], 2016: 521–529.

[243] TANG H, LIU W, ZHENG W.-L, et al. Multimodal Emotion Recognition Using Deep Neural Networks[C]// International Conference on Neural Information Processing. Springer. [S.l.]: [s.n.], 2017: 811–819.

[244] ZHENG W.-L, LU B.-L. Personalizing EEG-based affective models with transfer learning[C]// International Joint Conference on Artificial Intelligence. AAAI Press. [S.l.]: [s.n.], 2016: 2732–2738.

[245] ABRAHAM A, PEDREGOSA F, EICKENBERG M, et al. Machine learning for neuroimaging with scikit-learn[J]. Frontiers in Neuroinformatics, 2014, 8: 14.

[246] MORIOKA H, KANEMURA A, HIRAYAMA J.-I, et al. Learning a common dictionary for subject-transfer decoding with resting calibration[J]. NeuroImage, 2015, 111: 167–178.

[247] SUGIYAMA M, KRAULEDAT M, MÃŽLLER K.-R. Covariate shift adaptation by importance weighted cross validation[J]. Journal of Machine Learning Research, 2007, 8(May): 985–1005.

[248] KRAULEDAT M, TANGERMANN M, BLANKERTZ B, et al. Towards zero training for brain-computer interfacing[J]. PloS One, 2008, 3(8): e2967.

[249] PAN S J, TSANG I W, KWOK J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199–210.

[250] SANGINETO E, ZEN G, RICCI E, et al. We are not all equal: Personalizing models for facial expression analysis with transductive parameter transfer[C]// ACM International Conference on Multimedia. ACM. [S.l.]: [s.n.], 2014: 357–366.

[251] GRETTON A, BORGWARDT K M, RASCH M, et al. A kernel method for the two-sample-problem[C]// Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2007: 513–520.

[252] SCHÖLKOPF B, SMOLA A, MÜLLER K.-R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(5): 1299–1319.

[253] BLANCHARD G, LEE G, SCOTT C. Generalizing from several related classification tasks to a new unlabeled sample[C]// Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2011: 2178–2186.

[254] RICCI E, ZEN G, SEBE N, et al. A prototype learning framework using emd: Application to complex scenes analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 513–526.

[255] COLLOBERT R, SINZ F, WESTON J, et al. Large scale transductive SVMs[J]. Journal of Machine Learning Research, 2006, 7(Aug): 1687–1712.

[256] ZHOU J T, PAN S J, TSANG I W, et al. Hybrid Heterogeneous Transfer Learning through Deep Learning.[C]// AAAI. [S.l.]: [s.n.], 2014: 2213–2220.

[257] SALVADOR S, CHAN P. Toward accurate dynamic time warping in linear time and space[J]. Intelligent Data Analysis, 2007, 11(5): 561–580.

[258] GONG B, SHI Y, SHA F, et al. Geodesic flow kernel for unsupervised domain adaptation[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE. [S.l.]: [s.n.], 2012: 2066–2073.

[259] LONG M, WANG J, DING G, et al. Adaptation regularization: A general framework for transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5): 1076–1089.

[260] HUANG J, GRETTON A, BORGWARDT K M, et al. Correcting sample selection bias by unlabeled data[C]// Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2007: 601–608.

[261] DELORME A, MAKEIG S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis[J]. Journal of Neuroscience Methods, 2004, 134(1): 9–21.

[262] URIGÜEN J A, GARCIA-ZAPIRAIN B. EEG artifact removal state-of-the-art and guidelines[J]. Journal of Neural Engineering, 2015, 12(3): 031001.

[263] DALY I, SCHERER R, BILLINGER M, et al. FORCe: Fully online and automated artifact removal for brain-computer interfacing[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2015, 23(5): 725–736.

[264] LAFFERTY J, MCCALLUM A, PEREIRA F, et al. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]// International Conference on Machine Learning. Vol. 1. [S.l.]: [s.n.], 2001: 282–289.

[265] BALTRUŠAITIS T, ROBINSON P, MORENCY L.-P. Continuous conditional neural fields for structured regression[C]// European Conference on Computer Vision. Springer. [S.l.]: [s.n.], 2014: 593–608.

[266] PENG J, BO L, XU J. Conditional neural fields[C]// Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2009: 1419–1427.

[267] HUANG R.-S, JUNG T.-P, MAKEIG S. Tonic changes in EEG power spectra during simulated driving[G]// Foundations of Augmented Cognition, Neuroergonomics and Operational Neuroscience. [S.l.]: Springer, 2009: 394–403.

[268] JAP B T, LAL S, FISCHER P, et al. Using EEG spectral components to assess algorithms for detecting fatigue[J]. Expert Systems with Applications, 2009, 36(2): 2352–2359.

[269] DU L.-H, LIU W, ZHENG W.-L, et al. Detecting driving fatigue with multimodal deep learning[C]// International IEEE/EMBS Conference on Neural Engineering. IEEE. [S.l.]: [s.n.], 2017: 74–77.

Acknowledgements

The brain is the most complex organ in our body, where the functions of memory, learning, and cognition exist. As a college student, I got interested in the brain science. How the intelligence emerges from unintelligence with numerous tiny neurons connected with each other. It is lucky for me to study one subfield called affective brain-computer interactions during my Ph.D. study at Shanghai Jiao Tong University (SJTU). During this period, I had a wonderful research experience studying some computer science, some psychology, and some neuroscience. Through Ph.D. training, I became a researcher that can propose scientific problems and conduct the research independently from a college student.

Firstly, I especially must thank my Ph.D. supervisor Prof. Bao-Liang Lu for his generous help in both my research and daily life. Without the continuous support from Prof. Lu, I would never be able to finish this Ph.D. thesis. It is a great pleasure to work with Prof. Lu on such a cutting-edge research. I will never forget the time when we prepared proposals, conducted experiments, revised papers, and had discussions.

Thanks for Dr. Andrzej Cichocki at RIKEN Brain Science Institute, Japan and Prof. Jose A. Lozano and Dr. Roberto Santana at University of the Basque Country, Spain. The visits were really wonderful experiences during my stays in the laboratories. Thanks for the insightful comments and generous help in my academic life.

Science and technology development is by nature a highly collaborative process. I wish to thank most sincerely my friends and colleagues for their generous help: Dan Peng, Haohua Zhao, Wei Liu, Yun Luo, Liming Zhao, Rui Li, Rui Wang, Xiao-Wei Wang, Xiao-Lin Wang, Yong Peng, Yimin Yang, Keting Zhang, Chuangxin He, Yingying Jiao, Shu Jiang, Ruo-Nan Duan, Jia-Yi Zhu, Yifei Lu, Si-Yuan Wu, Jia-Jun Tong, He Li, Hao Tang, Xun Wu, Bo-Qun Ma, Xingzan Zhang, Xue Yan, Lihuan Du, Nan Zhang, Shan-Chun Shen, Xuemin Zhu, Jincheng Mei, Yu-Fei Zhang, Xiang-Yu Gao, and Xue-Qin Huo.

Last but not least, I appreciate the love and support of my family. They gave me a lot. Without their continuous support, I would never finish my Ph.D. study.

Publications

[1] **Wei-Long Zheng**, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki, EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. **IEEE Transactions on Cybernetics**, 2018.

[2] **Wei-Long Zheng** and Bao-Liang Lu, A Multimodal Approach to Estimating Vigilance Using EEG and Forehead EOG. **Journal of Neural Engineering**, 14(2): 026017, 2017.

[3] **Wei-Long Zheng**, Jia-Yi Zhu, and Bao-Liang Lu, Identifying Stable Patterns over Time for Emotion Recognition from EEG, **IEEE Transactions on Affective Computing**, 2017.

[4] **Wei-Long Zheng**, and Bao-Liang Lu, Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks, **IEEE Transactions on Autonomous Mental Development**, vol. 7, no. 3, pp. 162-175, 2015. (**2018 IEEE TAMD Outstanding Paper Award**)

[5] **Wei-Long Zheng**, Shan-Chun Shen and Bao-Liang Lu, Online Depth Image-Based Object Tracking with Sparse Representation and Object Detection, **Neural Processing Letters**, pp. 1-14, 2016.

[6] **Wei-Long Zheng** and Bao-Liang Lu, Personalizing EEG-based Affective Models with Transfer Learning, in Proc. of the 25th **International Joint Conference on Artificial Intelligence (IJCAI'16)**, 2016: 2732-3738.

[7] Yifei Lu*, **Wei-Long Zheng***, Binbin Li, and Bao-Liang Lu, Combining Eye Movements and EEG to Enhance Emotion Recognition, in Proc. of the **International Joint Conference on Artificial Intelligence (IJCAI'15)**, 2015:1170-1176. (**contributed equally as joint first authors*)

[8] **Wei-Long Zheng**, Yong-Qi Zhang, Jia-Yi Zhu, and Bao-Liang Lu, Transfer Components between Subjects for EEG-based Emotion Recognition, in Proc. of the sixth International Conference on **Affective Computing and Intelligent Interaction (ACII'15)**, 2015: 917-922.

[9] **Wei-Long Zheng**, Roberto Santana, and Bao-Liang Lu, Comparison of Classification Methods for EEG-based Emotion Recognition, in Proc. of the **World Congress on Medical Physics and Biomedical Engineering (WC'15)**. IFMBE, 2015: 1184-1187

[10] **Wei-Long Zheng**, Hao-Tian Guo, and Bao-Liang Lu, Revealing Critical Channels and Frequency Bands for EEG-based Emotion Recognition with Deep Belief Network, in Proc. of the 7th **International IEEE EMBS Conference on Neural Engineering (IEEE NER'15)**. IEEE, 2015: 154-157.

[11] **Wei-Long Zheng**, Jia-Yi Zhu, Yong Peng, and Bao-Liang Lu, EEG-Based Emotion Classification Using Deep Belief Networks. in Proc. of the **IEEE International Conference on Multimedia and Expo (ICME'14)**. IEEE, 2014: 1-6.

[12] **Wei-Long Zheng**, Jia-Yi Zhu, and Bao-Liang Lu, Multimodel Emotion Analysis in Response to Multimedia. in Proc. of the **IEEE International Conference on Multimedia and Expo Workshops (ICMEW'14)**. IEEE, 2014: 1-2.

[13] **Wei-Long Zheng**, Bo-Nan Dong, and Bao-Liang Lu, Multimodal Emotion Recognition using EEG and Eye Tracking Data. in Proc. of the 36th Annual **International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'14)**. IEEE, 2014: 5040-5043.

[14] Yimin Yang, Q. M. Jonathan Wu, **Wei-Long Zheng**, and Bao-Liang Lu, EEG-based Emotion Recognition Using Hierarchical Network with Subnetwork Nodes, **IEEE Transactions on Cognitive and Developmental Systems**, 2017.

[15] Yong Peng, **Wei-Long Zheng** and Bao-Liang Lu, An Unsupervised Discriminative Extreme Learning Machine And Its Applications To Data Clustering. **Neurocomputing** 174: 250-264 (2016).

[16] He Li, **Wei-Long Zheng** and Bao-Liang Lu, Multimodal Vigilance Estimation with Adversarial Domain Adaptation Networks, in Proc. of the **International Joint Conference on Neural Networks (IJCNN'18)**, 2018.

[17] Jia-Jun Tong, Yun Luo, Bo-Qun Ma, **Wei-Long Zheng**, Bao-Liang Lu, Xiao-Qi Song and Shi-Wei Ma, Sleep Quality Estimation with Adversarial Domain Adaptation: From Laboratory to Real Scenario, in Proc. of the **International Joint Conference on Neural Networks (IJCNN'18)**, 2018.

[18] Xue Yan, **Wei-Long Zheng**, Wei Liu, and Bao-Liang Lu, Identifying Gender Differences in Multimodal Emotion Recognition Using Bimodal Deep AutoEncoder, in Proc. of the **International Conference on Neural Information Processing (ICONIP'17)**, 2017: 533-542.

[19] Xue Yan, **Wei-Long Zheng**, Wei Liu, and Bao-Liang Lu, Investigating Gender Differences of Brain Areas in Emotion Recognition Using LSTM Neural Network, in Proc. of the **International Conference on Neural Information Processing (ICONIP'17)**, 2017: 820-829.

[20] Xing-Zan Zhang, **Wei-Long Zheng**, and Bao-Liang Lu, EEG-Based Sleep Quality Evaluation with Deep Transfer Learning, in Proc. of the **International Conference on Neural Information Processing (ICONIP'17)**, 2017: 543-552.

[21] Hao Tang, Wei Liu, **Wei-Long Zheng**, and Bao-Liang Lu, Multimodal Emotion Recognition Using Deep Neural Networks, in Proc. of the **International Conference on Neural Information Processing (ICONIP'17)**, 2017: 811-819.

[22] Wei-Ye Zhao, Sheng Fang, Ting Ji, Qian Ji, **Wei-Long Zheng**, Bao-Liang Lu, Emotion Annotation Using Hierarchical Aligned Cluster Analysis, in Proc. of the **International Conference on Neural Information Processing (ICONIP'17)**, 2017: 572-580.

[23] Si-Yuan Wu, Moritz Schaefer, **Wei-Long Zheng**, Bao-Liang Lu, Neural Patterns between Chinese and Germans for EEG-based Emotion Recognition, in Proc. of the 8th **International IEEE EMBS Conference on Neural Engineering (IEEE NER'17)**, 2017: 94-97.

[24] Zhen-Feng Shi, Chang Zhou, **Wei-Long Zheng** and Bao-Liang Lu, Attention Evaluation with Eye Tracking Glasses for EEG-based Emotion Recognition, in Proc. of the 8th **International IEEE EMBS Conference on Neural Engineering (IEEE NER'17)**, 2017: 86-89.

[25] Li-Huan Du, Wei Liu, **Wei-Long Zheng** and Bao-Liang Lu, Detecting Driving Fatigue with Multimodal Deep Learning, in Proc. of the 8th **International IEEE EMBS Conference on Neural Engineering (IEEE NER'17)**, 2017:74-77.

[26] Yi-Ming Jin, Yu-Dong Luo, **Wei-Long Zheng**, Bao-Liang Lu, EEG-based Emotion Recognition Using Domain Adaptation Network, in Proc. of the **International Conference on Orange Techology**, 2017: 222-225.

[27] Xue-Qin Huo, **Wei-Long Zheng**, and Bao-Liang Lu, Driving Fatigue Detection with Fusion of EEG and Forehead EOG, in Proc. of **International Joint Conference on Neural Networks (IJCNN'16)**, 2016: 897-904.

[28] Li-Li Wang, **Wei-Long Zheng**, Hai-Wei Ma and Bao-Liang Lu, Measuring Sleep Quality from EEG with Machine Learning Approaches, in Proc. of **International Joint Conference on Neural Networks (IJCNN'16)**, 2016:905-912.

[29] Wei Liu, **Wei-Long Zheng**, and Bao-Liang Lu, Emotion Recognition Using Multimodal Deep Learning. in Proc. of the 23nd **International Conference on Neural Information Processing (ICONIP'16)**, 2016: 521-529.

[30] Nan Zhang, **Wei-Long Zheng**, Wei Liu, and Bao-Liang Lu,Continuous Vigilance Estimation using LSTM Neural Networks. in Proc. of the 23nd **International Conference on Neural Information Processing (ICONIP'16)**, 2016: 530-537.

[31] Jia-Yi Zhu, **Wei-Long Zheng**, and Bao-Liang Lu. Cross-subject and Cross-gender Emotion Classification from EEG, in Proc. of the **World Congress on Medical Physics and Biomedical Engineering (WC'15)**. IFMBE, 2015: 1188-1191

[32] Xiang-Yu Gao, Yu-Fei Zhang, **Wei-Long Zheng**, and Bao-Liang Lu, Evaluating Driving Fatigue Detection Algorithms Using Eye Tracking Glasses, in Proc. of the 7th **International IEEE EMBS Conference on Neural Engineering (IEEE NER'15)**. IEEE, 2015: 767-770.

[33] Yu-Fei Zhang, Xiang-Yu Gao, Jia-Yi Zhu, **Wei-Long Zheng**, and Bao-Liang Lu, A Novel Approach to Driving Fatigue Detection Using Forehead EOG, in Proc. of the 7th **International IEEE EMBS Conference on Neural Engineering (IEEE NER'15)**. IEEE, 2015: 707-710.

[34] Yong-Qi Zhang, **Wei-Long Zheng**, and Bao-Liang Lu, Transfer Components between Subjects for EEG-based Driving Fatigue Detection, in Proc. of the 22rd **International Conference of Neural Information Processing (ICONIP'15)**, 2015: 61-68.

[35] Yong Peng, Jia-Yi Zhu, **Wei-Long Zheng**, and Bao-Liang Lu, EEG-Based Emotion Recognition with Manifold Regularized Extreme Learning Machine, in Proc. of the 36th Annual **International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'14)**, IEEE, 2014: 974-977.

[36] Jia-Yi Zhu, **Wei-Long Zheng**, Ruo-Nan Duan, Yong Peng and Bao-Liang Lu. EEG-based Emotion Recognition Using Discriminative Graph Regularized Extreme Learning Machine, in Proc. of the **International Joint Conference on Neural Networks (IJCNN'14)**. IEEE, 2014: 525-532.

[37] Xuemin Zhu, **Wei-Long Zheng**, Bao-Liang Lu, Xiaoping Chen, Shanguang Chen and Chunhui Wang, EOG-based Drowsiness Detection Using Convolutional Neural Networks, in Proc. of the **International Joint Conference on Neural Networks (IJCNN'14)**. IEEE, 2014: 128-134.

[38] Shan-Chun Shen, **Wei-Long Zheng**, and Bao-Liang Lu, Online Object Tracking based on Depth Image with Sparse Coding, in Proc. of the 21st **International Conference of Neural Information Processing (ICONIP'14)**. 2014: 234-241.

Under Review

[39] **Wei-Long Zheng**, Kunpeng Gao, Gang Li, Wei Liu, Chao Liu, Jing-Quan Liu, Guoxing Wang, and Bao-Liang Lu, Vigilance Estimation Using a Wearable EOG Device in Real Driving Environments, submitted to **IEEE Transactions on Intelligent Transportation Systems** (under second round review).

Project Paticipation

[1] National Key Research and Development Program of China "The Key Technologies, Platform Devices and Applications of Brain Machine Collaboration and Hybrid Intelligence" (Grant No. 2017YFB1002501)

[2] National Natural Science Foundation of China "A Study on the Discrimination Ability, Representation Characteristics and Stability of EEG and Eye Movement Signals for Multimodal Emotion Recognition" (Grant No. 61673266)

[3] Major Basic Research Program of Shanghai Science and Technology Committee "A Study on Brain Information Processing based Deep Neural Networks and Intelligent Systems" (Grant No. 15JC1400103)

[4] National Basic Research Program (973) of China "A Study on Visual Cognition Coding and Information Integration Mechanism of the Brain" (Grant No. 2013CB329401)

[5] National Science Foundation of China "A Study on Key Techniques for Fatigue Monitoring by Using Forehead EOG" (Grant No. 61272248)

[6] Technology Research and Development Program of China Railway Corporation "A Study on Human Factors Analysis of High Speed Railway Drivers" (Grant No. 2016Z003-B)

[7] Science and Technology Commission of Shanghai Municipality "EOG and EEG based Driving Fatigue Detection System Research" (Grant No.13511500200)

[8] Open Funding Project of National Key Laboratory of Human Factors Engineering "Astronaut Fatigue Detection with EEG and forehead EOG" (Grant No. HF2012-K-01)

[9] European Union Seventh Framework Program (Grant No.247619)

List of Patents

[1] Bao-Liang Lu, **Wei-Long Zheng**, and Yifei Lu, “An EEG Feature Recognition System and Method”, Application ID: 201710372387.7

[2] Bao-Liang Lu, **Wei-Long Zheng**, Zhen-Feng Shi, and Chang Zhou, “An Eye Movement Based EEG Experiment Evaluation System and Method”, Application ID: 201710378148.2

Resume

Basic Information

Wei-Long Zheng, Male, Born in October, 1988 in Guangdong, China, Ph.D. Candidate of Computer Science and Technology, The School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University.

Education

September 2012-June 2018, Shanghai Jiao Tong University, Ph.D.

June 2016-September 2016, Brain Science Institute, RIKEN, Visiting Researcher

September 2014-March 2015, University of the Basque Country, Visiting Researcher

September 2008-July 2012, South China University of Technology, Bachelor

Research Interest

Affective Computing, Affective Brain-Computer Interactions, Machine Learning, Pattern Recognition, and Brain-Computer Interface

Scholarships and Honors

- IEEE Transactions on Autonomous Mental Development Outstanding Paper Award, IEEE Computational Intelligence Society 2018
- Excellent Ph.D. Student Scholarship of Yang Yuanqing Education Fund, Shanghai Jiao Tong University 2016
- Ruth Mulan Chu Chao Ph.D. Student Scholarship, Shanghai Jiao Tong University 2016
- National Scholarship for Graduate Student, Ministry of Education, China 2015
- Graduate Academic Excellence Scholarship of Shanghai Jiao Tong University 2014

Contact Information

Address: Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China, 200240

Email: weilong@sjtu.edu.cn