Title

Brain-computer Interfaces for Online Mental Stress Monitoring in the Real World

Author

Chang, Chi-Yuan

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Brain-computer Interfaces for Online Mental Stress Monitoring in the Real World

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Bioengineering

by

Chi-Yuan Chang

Committee in charge:

Professor Tzyy-Ping Jung, Chair
Professor Gert Cauwenberghs, Co-Chair
Professor Todd Coleman
Professor Virginia de Sa
Professor Vikash Gilja

2022

The Dissertation of Chi-Yuan Chang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my father Chin-Fa Chang, my mother Juo-Mei Cheng,
my brother Yao-Yuan Chang,
my girlfriend Longyue Wang, and all my cats.

EPIGRAPH

Even if most people don't appreciate it,
I have to do it for myself!
I have to proclaim
"Here I am!"
to the people who can appreciate each individual movement!

Keep Your Hands Off Eizouken!
Episode 7, I have to do it for myself.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

EEG	Electroencephalography or Electroencephalogram
ECG	Electrocardiography or Electrocardiogram
HR	Heart Rate
HRV	Heart Rate Variability
EDA	Electrodermal activity
GSR	Galvanic Skin Response
BCI	Brain-computer Interface(s)
ASR	Artifact Subspace Reconstruction
ICA	Independent Component Analysis
IC	Indepedent Component
VR	Virtual Reality
GIP	Gaze Intersection Point
ERP	Event Related Potential
PSD	power spectral density
1D/3D	One Dimension(al)/ Three Dimension(al)
Acc.	Accuracy
LDA	Linear Discriminant Analysis
NN	Neural Network
SMOTE	Synthetic minority over-sampling technique
LOO	Leave-one-(Session)-out

ACKNOWLEDGEMENTS

I would like to thank my department advisor Prof. Gert Cauwenberghs and my other advisor Prof. Tzyy-Ping Jung. The first time I met Prof. Gert was in his mathematics class. Through the lectures, I can feel his passion and excitement about teaching. I appreciate him strengthen my understanding of machine learning and signal processing. Tzyy-Ping (TP) is my advisor at Swartz Center for Computational Neuroscience (SCCN), and he is the one who introduced me to the field of brain-computer interfaces. He also built up my knowledge of EEG analysis and my presentation skills. I also appreciate how open-minded and encouraging TP is. I am truly lucky to join TP's lab. I greatly appreciate Prof. Gert and TP for being such supportive and caring to all the problems I have encountered in my progress toward my PhD degree. I also want to thank my PhD committee for their time and support.

I would like to thank my friends and colleagues at SCCN for their guidance and support in both the aspect of life and academics. Sheng-Hsiou Hsu and Luca Pion-Tonachini were my mentors and we worked together on many of my projects. Makoto Miyakoshi is the one I can discuss with almost all of my projects due to his solid statistical knowledge and multidisciplinary habits. Thanks to Masaki Nakanishi, Kuan-jung Chiang, Tzu-han Cheng, Clement Lee, and Hyeonseok Kim. I enjoy conversing with you and exchanging opinions about machine learning, statistics, and food. I would like to appreciate Ying Choon Wu for all the valuable discussions and instructions in Chapter 4. Thank Dung Troung, Wenchien Liu, Chieh Hsu, Siwen Wang, Darin Tsui, and Xin Wen for the support in building the experiment scheme, recording data, and analyzing data in Chapter 4 and Chapter 5. I also would like to thank Robert Buffington who has resolved my internet and computer problems for countless times, and Nicole Wells who helped me record the dataset used in Chapter 4.

Finally, I would like to thank all my family and friends for their companionship and encouragement through my PhD journey.

Chapter 2 used the dataset recorded in National Chiao Tung University (National Yang-Ming Chiao-Tung University), Hsinchu, Taiwan by Dr. Li-Wei Ko. The dissertation author was

the first investigator of this project. The authors gratefully acknowledge the support from Dr. Li-Wei Ko.

Chapter 3, in full, has been published as Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini, and Tzyy-Ping Jung, "Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings", in IEEE Transactions on Biomedical Engineering, 2019. The dissertation author was the first investigator of this paper. The authors gratefully acknowledge the valuable suggestions and comments from Makoto Miyakoshi and the author of ASR, Christian Kothe.

Chapter 4, in full, is being prepared for submission for publication as Chi-Yuan Chang, Weichen Lu, Cory Stevenson, Russell Cohen Hoffing, Steven Thurman, Tzyy-Ping Jung and Ying Choon Wu, "Free Moving Gaze-related Electroencephalograms Electroencephalography in Mobile Virtual Environments". The dissertation author was the first investigator of this paper. This research was supported by grant #1734883 from the National Science Foundation and #W911NF2120126 and #W911NF2020088 from the Army Research Laboratory. The authors gratefully acknowledge the technical support on data recordings from Nicole Wells and Xin Wen.

Chapter 5 is an ongoing project at SCCN. The dissertation author was the first investigator of this project. The authors gratefully acknowledge the support from Microsoft Research and the help from Chieh Hsu, Darin Tsui, and Siwan Wang.

VITA

2014	Bachelor of Science, National Taiwan University
2015–2016	Teaching Assistant, Department of Mechanical Engineering National Taiwan University
2022	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini and Tzyy-Ping Jung, "Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 1242-1245, doi: 10.1109/EMBC.2018.

Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini and Tzyy-Ping Jung, "Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings," in IEEE Transactions on Biomedical Engineering, vol. 67, no. 4, pp. 1114-1121, April 2020, doi: 10.1109/TBME.2019.

Chi-Yuan Chang, Weichen Lu, Cory Stevenson, Russell Cohen Hoffing, Steven Thurman, Tzyy-Ping Jung and Ying Choon Wu, "Free Moving Gaze-related Electroencephalograms Electroencephalography in Mobile Virtual Environments"

FIELDS OF STUDY

Major Field: Bioengineering

Studies in Computational Neuroscience
Professors Tzyy-Ping Jung and Gert Cauwenberghs

Studies in Machine Learning
Professors Tzyy-Ping Jung, Gert Cauwenberghs, Virginia de Sa, and Vikash Gilja

Studies in Brain-computer Interfaces
Professors Tzyy-Ping Jung, Virginia de Sa, and Todd Coleman

ABSTRACT OF THE DISSERTATION

Brain-computer Interfaces for Online Mental Stress Monitoring in the Real World

by

Chi-Yuan Chang

Doctor of Philosophy in Bioengineering

University of California San Diego, 2022

Professor Tzyy-Ping Jung, Chair
Professor Gert Cauwenberghs, Co-Chair

Mental stress appears in our daily life and affects our well-being and working performance. Previous studies have shown that mental stress, especially under chronic situations, can be related to obesity, heart disease, depression, or even suicide. Thus, mental stress has been recognized as the top proxy killer. However, mental stress is not something we should necessarily avoid. When under an adequate stress level, the human performance of cognitive tasks or sports is promoted. Therefore, it is essential to monitor stress levels in daily life. This dissertation aims to build a brain-computer interface for online mental stress monitoring in the real world. We first investigated the Electroencephalography features reported in previous studies and found

many artifacts in data recorded in real-world scenarios. Hence, we moved on to investigate artifact removal methods and evaluate the performance of Artifact Subspace Reconstruction (ASR) using Independent Component Analysis (ICA). Next, we further evaluated ASR's brain signal reconstruction ability and explored human behaviors in a visual-oddball task conducted in a virtual environment. We implemented ASR into our stress detection algorithms with all the results we obtained and proposed an IC projection method to remove eye activities in real time without performing ICA. Finally, to reduce the manufacturing cost and setup time of deployment, we investigated the effect of recording channel reduction on our stress detection algorithms. We found our stress detection algorithm with Linear Discriminant Analysis (LDA) can reach a 77% balanced accuracy in an online scenario with only 11 recording channels placed in the frontal region.

Chapter 1

Introduction

1.1 Background

Stress is a reaction that happened when the demands from environments exceed the adaptive capacity of an organism, resulting in physiological and biological changes [1]. These demands are named stressors and can be further categorized into physical/ environmental stressors or mental/ task-related stressors as shown in Table 1.1 [2]. In our daily life, we faced various challenges and experienced different levels of mental stress. However, people usually underestimate the effect of mental stress on our well-being, both physically and mentally. When experiencing low-level stress, people might have upset stomach [3], dizzy [4], increased heart rate [5], or muscle tension [6]. As the level of stress increases, the symptoms can be more severe. Nowadays, more and more researchers have shown that mental stress is related to immune disorders [7], obesity [8], and, heart diseases [9]. Moreover, people undergoing chronic stress conditions are prone to depression [10], anxiety [11], and stroke [12]. In the worst cases, mental stress can even lead to substance abuse and suicide [13]. Therefore, mental stress has also been recognized as the top proxy killer disease [14].

While recognizing the negative influence of mental stress, it is not something necessarily to be avoided [15]. Back in 1908, Yerkes and Dodson have already proposed a model to explain the relationship between arousal and task performance [16, 17]. As we can see in Fig. 1.1, the performance increases as the arousal increases in simple task conditions, e.g. focused

Table 1.1. Example of stressors.

Physical / Environmental	Mental / Task related
Painful stimuli Strenuous physical activity Extreme temperature condition Low oxygen levels	Cognitive capacities demanded task Rapid change task instructions Conflicting task instructions Inconsistent reward reinforcement schedule

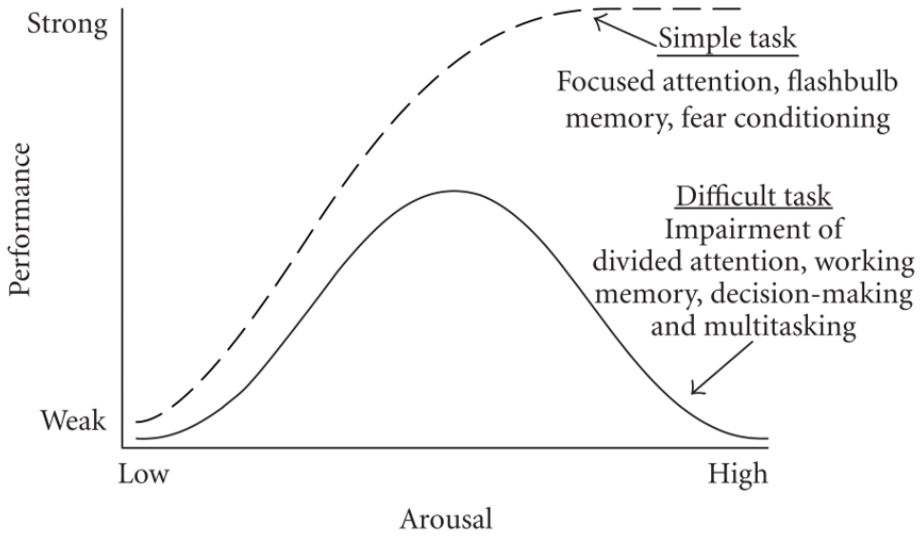

Figure 1.1. The relationship between arousal and task performance.

attention, flashbulb memory, and fear conditioning. On the other hand, the relationship between arousal and task performance is an inverted-U shape in difficult task conditions, e.g. divided attention, working memory, and decision-making. Stress can have a similar model since arousal increases when stress increases. There are studies showing that proper amounts of mental stress can increase performance in competitive sports [18] and some attention-related tasks [17]. In real-world scenarios, most of the tasks we encounter are in the category of difficult tasks. As a result, monitoring mental stress to remain at optimal task performance in daily life becomes an important problem.

Questionnaires and self-reports have been used to measure one's stress level in the field of psychology for a long time [19, 20, 21, 22]. These measurements usually serve as the ground

Table 1.2. The frequency range of each brainwave.

Name	Frequency range
δ	below 4Hz
θ	4Hz - 8Hz
α	8Hz - 13Hz
β	13Hz - 32Hz
γ	32Hz - 100Hz

truth of the stress level in stress-related studies [23, 24, 25]. However, it is impossible to monitor mental stress in real time with questionnaires and self-reports. Moreover, these measurements are subjective and might suffer from anchoring effects [26].

In contrast, physiological signals provide an objective measurement for stress detection. The Automatic Nervous System (ANS) controls the involuntary movements of the body such as skin conductivity, heart rate, and pupil dilation [27, 28, 29]. Since stress causes dynamic changes in ANS, there are several measurements to detect stress level changes [2, 28]. The most commonly used physiological signals in stress detection are Heart Rate (HR), Heart Rate Variability (HRV), Galvanic Skin Response (GSR), Blood Pressure, and Respiration Rate [30]. Previous studies have shown that, under stressful conditions, HR, GSR, Blood Pressure, and Respiration Rate increase, and the R-peak-to-R-peak interval in HRV shorten [2, 28]. Nevertheless, though HR has been wildly used in previous studies, HR might be irrelevant to mental stressors once there are body movements during measurements as the HR increases during exercises.

Electroencephalography (EEG), an non-invasive brain signals measurement, is another commonly used measurement for stress detection [28]. Since EEG has a high temporal resolution and low cost compared to other brain signals measurements, it has been widely used in the fields of neuroscience [31, 32], clinical assessment [33], and brain-computer interfaces [34]. An ordinary way to investigate EEG signals is to separate the signals into different brainwaves using different frequency ranges as shown in Table 1.2. Most common reported EEG features used in stress detection include θ power increases [35, 36], α power decreases [35, 36, 37], β power increases [36, 37], and α asymmetry [38, 39].

Table 1.3. Previous studies of stress detection.

Study	Stimuli	Biosignals used	Accuracy
Xia et al.[41]	Mental Arithmetic Task	EEG, ECG	79.54%
Minguillon et al.[42]	Montreal Imaging Stress Task	EEG, ECG, EMG, EDA	94.60%
Asif et al.[43]	Music Tracks	EEG	98.76%
Saeed et al.[39]	No stimuli	EEG	85.20%

To focus on deploying mental stress detection in the real world, we selected EEG as the measurement for stress detection in this dissertation. Because the human brain determines whether a situation is stressful, mental stress influences the human brain the most [40]. We hypothesized EEG can reveal information in addition to arousal and thus have a more robust measurement in real-world scenarios.

1.2 Problem statement

There are many studies that demonstrate the capability of detecting stress in laboratory environments. Nevertheless, stress detection in realistic environments is still in early exploration. Table 1.3 shows the previous studies in stress detection. As we can see in the table, most of the studies induce mental stress by extra stimuli except for Saeed et al. [39].

In this dissertation, we emphasize deploying BCI for mental stress monitoring in the real world. As a result, we focused on three main problems. First, the BCI must be robust, i.e. removing non-brain signals while preserving brain signals. Second, the BCI needs to be online capable. Third, the BCI needs to be easy to set up. While we do not build a commercialized BCI for mental stress monitoring, we do provide a potential solution by improving the robustness of BCI and reducing the difficulty of deployment.

Chapter 2

Mental Stress Detection

2.1 Introduction

To approach online mental stress monitoring, instead of continuous stress measurement, we simplified the problem into a binary mental stress detection, i.e. stress level increases or remains. In this chapter, we built two stress detection algorithms without any preprocessing of the EEG signals. We first introduced the dataset used in building our stress detection algorithms. Next, we explained the processes of data labeling, feature extraction, and feature preprocessing. Finally, we introduced the models used in our algorithms and the metrics to evaluate their performance. We found that EEG signals are highly contaminated by non-brain signals and thus damage the performance of the algorithms. The performance of our algorithms has been improved in Chapter 5 after implementing our results in Chapter 3.

2.2 Data acquisition and preprocessing

To detect mental stress in the real world, we used a dataset recorded from a longitudinal stress and fatigue experiment in a classroom [44]. In this experiment, students were asked to participate in their lectures with their EEG and ECG recorded for two semesters. At the beginning of the lectures, students were asked to complete a Depression, Anxiety, Stress Scales-21 (DASS-21) questionnaire [22] as subjective reports. Next, we recorded an eye-open session for about 5 mins. Afterward, the students attended their lectures normally for about 60 mins.

At the end of the lectures, students were asked to perform another eye-open resting for 5 mins. We also recorded another DASS-21 report after the eye open resting in the second semester. Eighteen subjects participated in the first semester and eight subjects participated in the second semester. Each subject recorded 4 to 13 sessions on different days. The EEG and ECG data were recorded using a 32-channel NeuroScan System at 1000 Hz sampling rate. In this study, we analyzed the first semester only since there were insufficient self-reports in the second semester. To classify mental stress levels using non-task related data, we first extract the 5 mins eye open resting at the beginning of the experiment. Then, we performed bandpass filtering from 1Hz to 50Hz to remove high-frequency noise.

2.3 Data labeling, feature extraction, and feature preprocessing

2.3.1 Data labeling

In this study, we use the stress index in the DASS-21 questionnaire to quantify subjects' stress levels. Since the stress level varied from 0 to 38 across all the sessions and the number of sessions is not enough for regression analysis, we separated the sessions into an increased stress level group and a normal stress level group according to DASS-21. If the sessions had a stress index higher than 14, we assigned them to be in the increased stress level group and the rest of the sessions to be in the normal stress level group. After the sessions without DASS-21 reports, we ended up with 71 normal stress level groups and 21 increased stress level groups. Fig. 2.1 shows the number of sessions recorded from each subject in the first semester.

2.3.2 Feature extraction

In this study, we calculated each channel's power spectral density (PSD) from the eye-open resting sessions and created 50 frequency bins with a bin width equal to 1Hz. To reduce the number of features, we selected PSD from 3Hz to 7Hz frequency bins since these frequency bins overlapped with θ band which shows high correlations with cognitive state changes in previous

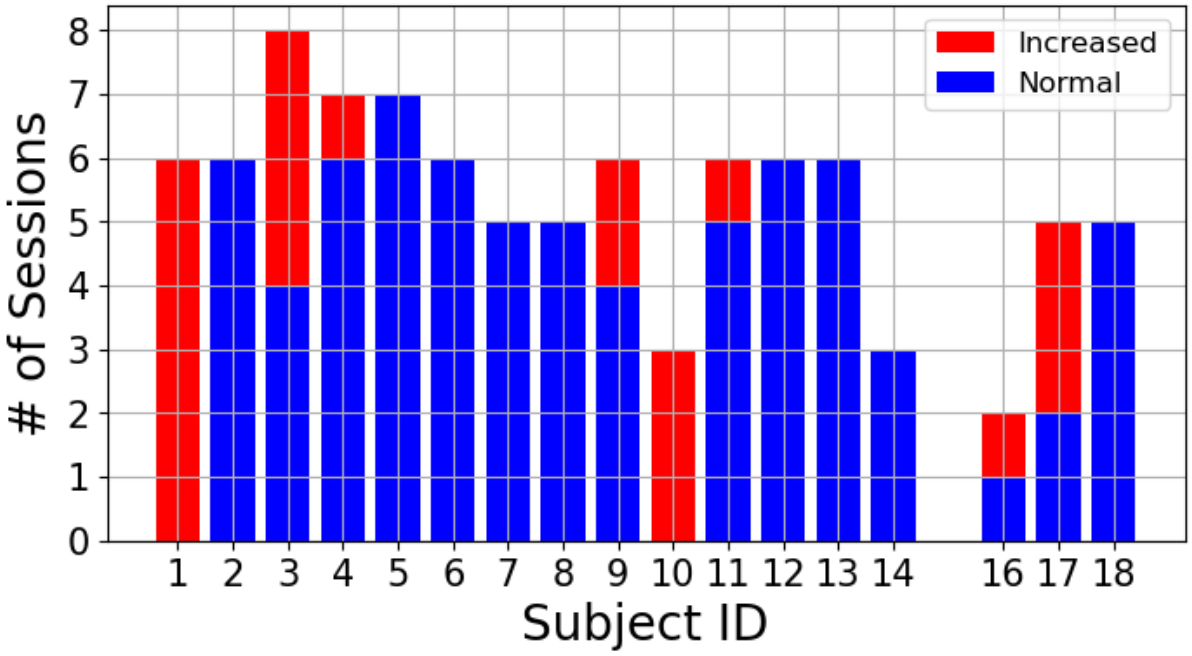

Figure 2.1. Number of sessions recorded from each subject. The increased stress level sessions are indicated in red. The normal stress level sessions are indicated in blue.

studies [35, 36]. Though some studies report the power of α band decreases under stressful conditions, we observed an increase instead. Hence, we did not cover α frequency range for feature extraction. A detailed analysis is provided in Section 5.5.1. We did not include β and γ bands because the frequency range is highly overlapped with muscle activities [45] and can be unreliable in realistic scenarios. The number of features we have extracted is 150 which is 5 frequency bins times 30 channels.

2.3.3 Synthetic minority over-sampling technique

Training classifiers on the imbalance classes can potentially bias the classifiers' results toward the majority group. To overcome the imbalance classes problem, we evaluated the performance of the synthetic minority over-sampling technique (SMOTE) [46]. First, SMOTE randomly selects one data point out of the minority group. Next, SMOTE calculates k-nearest neighbors in the minority group of the selected data point. Finally, SMOTE randomly selects one of the neighbors and generates a synthetic minority data point between the selected data

point and the neighbor. By repeating this process, the number of data points in the majority and minority groups is equal.

2.4 Model selection and evaluation

2.4.1 Linear Discriminant Analysis

The first classification model we have investigated is Linear Discriminant Analysis (LDA) [47]. LDA aims to find a linear combination of features to separate different classes into different distributions. In this study, we implement LDA in `scikit-learn` [48], a python library, with equal probability prior. Since LDA is sensitive to outlier [49], we further selected 3Hz to 7Hz from Fz, FCz, and Cz out of 150 features to prevent potential outlier effects. These channels are on the midline of the head topology (Fig. 2.4) and previous studies have shown these channels are highly related to cognitive state changes [50].

2.4.2 Neural Network

Besides the linear classifier, we also investigated the performance of the non-linear classifier. In this study, we built a 4-layers fully connected neural network (NN) using `pyTorch` [51] and trained this NN using the features from all the recording channels. To compare NN with LDA model, we also built another 4-layers fully connected NN with different numbers of nodes in each layers and trained this NN using the same features as the LDA model (Fz, FCz, and Cz). The architectures of the NN are shown in Fig. 2.2. We used ReLU as an activation function between layers and a Sigmoid function at the output layers. The number of training epochs is 1000 and the loss function for backpropagation is binary cross-entropy loss.

2.4.3 Model evaluation

To evaluate the performance of our model, we performed leave-one-session out (LOO) cross-validation after model training. LOO cross-validation first picks out a test session from the dataset and then trains the model on the rest of the sessions. Next, the trained model is applied to

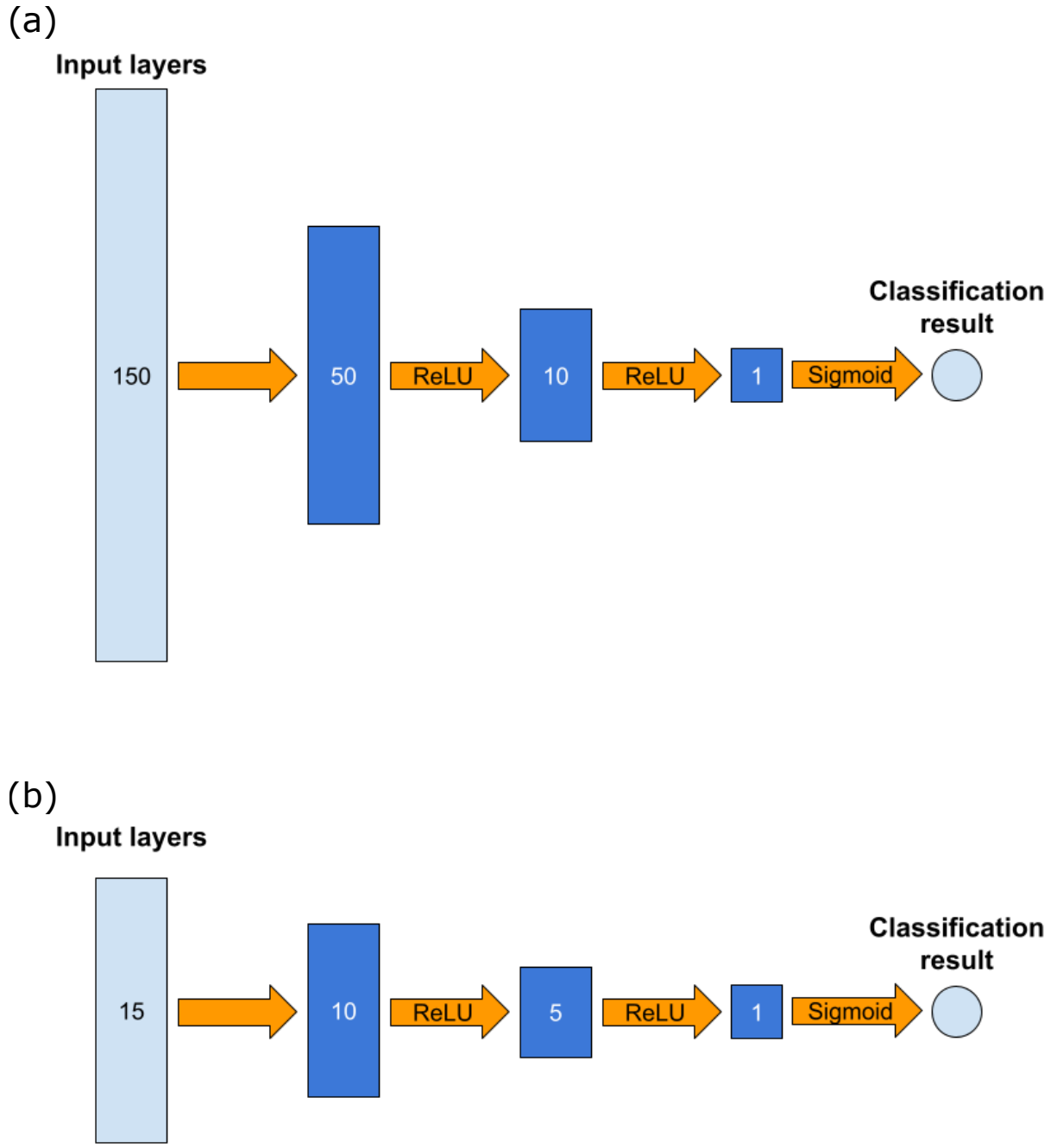

Figure 2.2. The architectures of the Neural Network used in this study. The number of nodes is indicated by the number in each layer. All the layers are fully connected to the next layer. (a) The NN model uses the features from all the recording channels. (b) The NN model uses the features from Fz, FCz, and Cz.

the test session. LOO cross-validation performs the same procedure for all of the sessions and finally reports the prediction results of all the sessions.

Since our dataset is imbalanced, containing more normal stress level groups than increased stress level groups, the accuracy calculated over all the sessions is 77.17%. Therefore, instead of using accuracy as an indicator of performance, we used balanced accuracy shown in eq. 2.1 which gives us a random probability of 0.5. We also used F1-score shown in eq. 2.2 to indicate the harmonic balance between recall and precision.

$$Balanced\ Accuracy\ =\ \frac{1}{N_{class}}\sum_{i=1}^{N_{class}}\frac{Number\ of\ correct\ prediction\ in\ class\ i}{Number\ of\ samples\ in\ class\ i} \tag{2.1}$$

$$F_1\ =\ 2\,\frac{precision\cdot recall}{precision+recall} \tag{2.2}$$

2.5 Results and discussion

Fig. 2.3 shows the PSDs of Fz, FCz, and Cz which are the selected channels used in LDA stress detection. We found there is no significant difference between the Increased stress level group and the Normal stress level group in the feature frequency for all the feature channels. Moreover, the scale of the PSD goes up to 20 dB which is higher than empirical observations.

Table. 2.1 shows the performance of our stress detection algorithms using bandpass-filtered data. The LOO Acc. increases with SMOTE for LDA stress detection while decreases in NN stress detection. On the other hand, LOO Balanced Acc. and F1 increase after SMOTE in both stress detection algorithms. These results show that SMOTE increases the ability to detect minority groups which is the increased stress level group in our study. However, the results are not satisfying enough since our LOO Acc. is lower than the 77.17% chance level accuracy which predicts all sessions into the majority class, and our LOO Balanced Acc. is slightly above 50%.

Fig. 2.4 shows an example of EEG signals with contamination from artifacts, i.e. non-

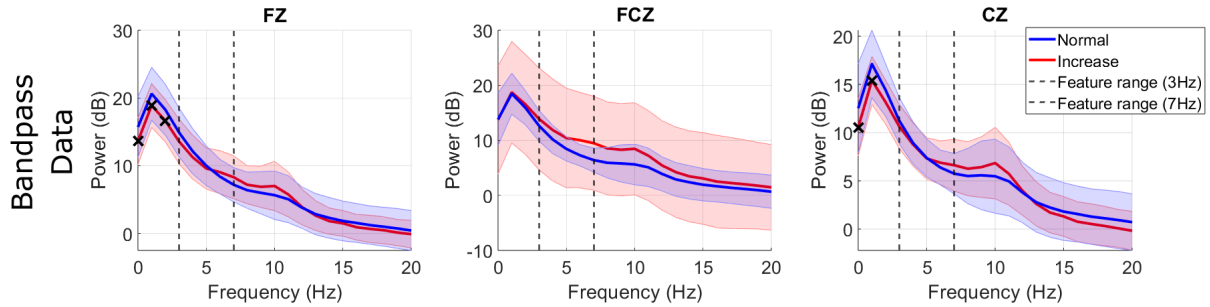

Figure 2.3. PSD features of selected channels after bandpass filtering. The results of the normal and increased stress level groups are shown in blue and red respectively. The solid lines indicate the mean and the shaded area indicates the standard deviation across sessions. The cross dots show the significant difference between the two groups ($p < 0.05$). The dashed line indicates the frequency range used in the analysis.

Table 2.1. Stress detection using bandpass filtered data. BP refers to bandpass filtering.

	LOO Acc.	LOO Balanced Acc.	F1
LDA (w/o SMOTE)	68.48%	62.81%	0.43
LDA (w/ SMOTE)	72.83%	67.30%	0.49
NN (w/o SMOTE)	75.00%	56.98%	0.30
NN (w/ SMOTE)	73.91%	62.98%	0.43

brain signals. We can see there are several artifacts caused by eye blinks and there is an artifact caused by jaw clenching at around 74.5 sec. Previous studies have shown muscle-related artifacts contaminates EEG PSD from 8Hz and upwards, especially having high powers in the range of 20Hz to 60Hz [52], and eye-related artifacts contaminate EEG PSD from 0Hz to 12Hz [53]. These artifact-contaminated frequency range overlapped with the frequency range of the features used in this study. Based on the results, we hypothesize the classification performance is deprecated because EEG signals are contaminated by artifacts.

2.6 Conclusion

This study aims to detect mental stress in the real world. We built two stress detection algorithms using the PSD features calculated from eye-open resting EEG signals. However, the EEG signals are strongly contaminated by the artifacts. Therefore, we removed artifacts before

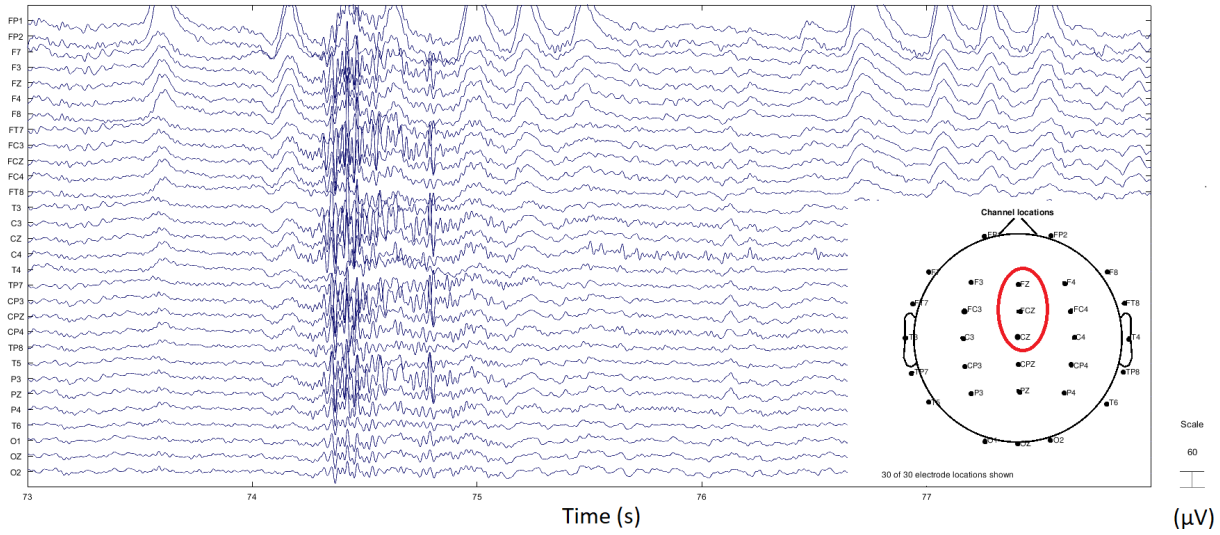

Figure 2.4. Example of artifact contaminated EEG recordings after bandpass filtering. The 30 channel locations are shown in the lower right corner with Fz, FCz, and Cz marked in the red circle.

building our stress detection algorithms.

2.7 Acknowledgement

The dataset used in Chapter 2 is recorded in National Chiao Tung University (National Yang Ming Chiao Tung University), Hsinchu, Taiwan by Dr. Li-Wei Ko. The dissertation author was the first investigator of this project. The authors gratefully acknowledge the support from Dr. Li-Wei Ko.

Chapter 3

Artifact removal

3.1 Introduction

One of the greatest challenges that hinder the decoding and application of EEG data is that EEG recordings are almost always contaminated by artifacts such as electrode impedance changes caused by headset motion as well as eye blink, eye-movement, neck muscle, and scalp muscle activities. Traditionally, these artifacts were removed manually by visual inspection [54], which could be time-consuming, laborious, subjective, and incompatible with online and real-time applications [55].

To automate the artifact removal process, earlier methods have used channel-based statistical thresholding approaches to remove abnormal activities [56, 57] or adaptive filters with additional reference channels to regress out targeted artifacts [58, 59]. Unfortunately, these methods either cannot reconstruct clean data from spatially outspread artifacts or require auxiliary channels for specific artifacts.

Another popular approach is to separate artifacts from brain signals using blind source separation (BSS), especially independent component analysis (ICA) [54, 55]. Since BSS cannot identify artifact components automatically, a classifier is needed to identify and reject the artifact-related components. Most of the classifiers are pre-trained and are not adaptive to different datasets [60, 61, 62, 63] or they require pre-recorded target-artifact sections [64] or auxiliary channels [65]. Moreover, the ICA-based methods are usually less effective in removing transient,

non-biological artifacts such as abrupt impedance changes from headset motions.

Recently, joint blind source separation (JBSS) [66, 67] and embedded empirical mode decomposition (EEMD) [68, 69] have shown great performance in removing muscle artifacts and high-amplitude artifacts. However, JBSS-based methods and EEMD-based methods are unavailable for real-time applications due to their high computational cost. Singular spectrum analysis (SSA) is another effective method for separating artifacts in single-channel recordings [70]. However, SSA-based methods require users to manually select a threshold for decomposition and may not be online-capable when applied to multi-channel recordings due to the computation of singular value decomposition for each channel.

To address the challenges the above methods encountered, Kothe and Jung [71] proposed artifact subspace reconstruction (ASR), which is an automatic, online-capable, component-based artifact removal method for removing transient or large-amplitude artifacts in multi-channel EEG recordings. Table 1 summarizes the state-of-the-art automatic, online-capable artifact removal methods for multi-channel EEG recordings [72]. ASR is similar to principal-component-analysis-based (PCA-based) methods in which large-variance components are rejected and channel data are reconstructed from remaining components. The main difference is that ASR automatically identifies and utilizes clean portions of data as a reference to determine thresholds for rejecting components.

In the EEG community, there has been increasing use of ASR as a powerful, automatic data-cleaning method [32, 73, 74, 75, 76]. In addition, recent studies on EEG artifact removal have used ASR as a benchmark for comparison, such as Kilicarslan et al. [59], Gabard-Durnam et al. [77], and Ojeda et al. [78]. However, they used a suboptimal parameter (current default value) of ASR, which motivate the current study.

This study systematically evaluates the effectiveness of ASR using twenty actual EEG recordings where various artifacts were present. We first characterize the effects of ASR on overall signal reduction with different cutoff parameters. Next, extending from the previous study [79], this study applies ICA and further employs a recently-developed, automatic IC

classifier (*ICLabel*) [80] to separately identify brain activities and various types of artifact signals. This allows quantitative and objective assessment of ASR's effectiveness in removing different artifacts and preserving brain activity. Finally, the results of dipole fitting to decomposed ICs after ASR cleaning are reported to provide evidence that ASR could improve the quality of ICA decomposition. The online and real-time capability and shortcomings of ASR are also discussed. Collectively, this study provides a guideline for choosing an optimal ASR parameter and demonstrates that ASR can be an effective automatic, online-capable artifact removal method.

Table 3.1. State-of-the-art automatic, online-capable artifact removal methods for multi-channel EEG. **Some features cannot be used in online classification. **Functional Link Neural Network and Adaptive Neural Fuzzy Inference System.*

Author	Year	Method category	Type of artifacts	Reference	Toolbox
Guarnieri et al. [65]	2018	ICA-REG	EOG	EOG	N/A
Radüntz et al. [60]	2017	ICA	EOG, Heart	None	N/A
de Cheveigné [56]	2016	Statistic	Channel Noise	None	STAR
Zhang et al. [64]	2015	wavelet ICA	EOG, EMG	None	N/A
Frølich* et al. [61]	2015	ICA	EOG, EMG, Heart	None	IC_MARC
Hu et al. [81]	2015	FLNN-ANFIS**	EOG, EMG	EOG, EMG	N/A
Kothe and Jung [71]	**2014**	**PCA-statistic**	**Large-amplitude artifacts**	**None**	**ASR**
Winkler et al. [62]	2014	ICA	EOG,EMG	None	MARA
Bigdely et al. [63]	2013	ICA	EOG	None	EyeCatch
Noureddin et al. [58]	2012	Adaptive Filter	EOG	Eye Tracker	N/A
Gao et al. [82]	2010	CCA	EMG	None	N/A

3.2 Dataset and data preprocessing

3.2.1 Experiment and data collection

To evaluate ASR, we used twenty EEG recordings from 10 subjects performing a sustained attention task in a driving simulator [83]. The simulated driving experiment had subjects driving in a real car mounted on a motion platform with a 360-degree virtual reality environment and real steering-wheels control, hoping to realistically simulate the actual driving scenario. In the 90-minute experiments, subjects reacted to randomly occurring lane-departure events by steering the car back to the center of their lane. Therefore, there were intermittent artifacts in the EEG data from electrical interference, EEG headset motions, and activities from neck and scalp muscles, eye blinks, and eye movements. For each subject, 32-channel EEG data were recorded using a NeuroScan System at 500 Hz sampling rate. Wet electrodes (Ag/AgCl) were placed on the scalp following the international 10-20 system [84].

3.2.2 Data preprocessing

To remove high-frequency noise, the EEG data were cleaned using a band-pass FIR filter (0.5-100 Hz) and then were down-sampled to 250 Hz. Next, we used `clean_rawdata`, and EEGLAB plugin function [85], to remove channels with negligible activity (flat line threshold: 5), noisy signals (noisy line threshold: 4), or a poor correlation with adjacent channels (correlation threshold: 0.8).

3.3 Artifact Subspace Reconstruction (ASR)

This section describes the ASR algorithm with emphasis on the key aspects and advantages of ASR. A more detailed description of ASR is available in [71].

The underlying concept is that the data segment X_t can be decomposed into latent components S_t using the mixing matrix M_r: $X_t = M_r S_t$. Artifact rejection is performed in the principal component (PC) space $Y_t = V_t^T X_t = V_t^T M_r S_t$, and thus the clean latent components $(S_t)_{clean}$ can

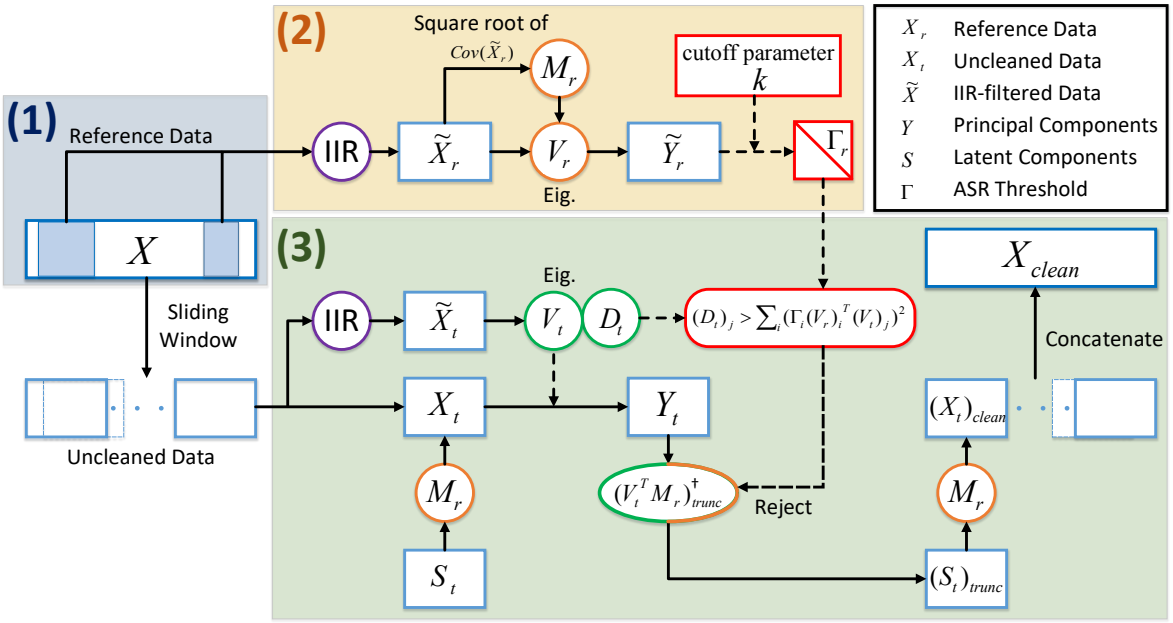

Figure 3.1. Flow chart for ASR.

be reconstructed using the pseudoinverse of the truncated $V_t^T M_r$: $(S_t)_{clean} = (V_t^T M_r)^+_{trunc} Y_t = (V_t^T M_r)^+_{trunc} V_t^T X_t$ where A^+ is the pseudoinverse of A. Projecting $(S_t)_{clean}$ back to channel-space using M_r yields the cleaned data in Eq. 3.1.

The ASR process consists of three steps: (1) extracting reference data from raw data, (2) determining thresholds for identifying artifact components, and (3) rejecting the artifact components and reconstructing the resulting data. Fig. 3.1 shows an overview of the three steps. The concepts and implementation details are described as follows.

3.3.1 Extract reference data

ASR automatically selects clean portions of EEG data based on the distribution of signal variance. Specifically, ASR calculates channel-wise root-mean-square (RMS) values on 1-second windows, z-scores the values across all windows from each individual channel identifies clean windows in which the z-scored values are within -3.5 and 5.5 [1], and concatenates the clean windows to obtain reference data X_r. A tolerance value, here we used 7.5%, is set to allow a

[1]The RMS values are fit into a truncated Gaussian distribution.

small percentage of bad channels to remain in X_r otherwise the criteria for choosing X_r is too restrictive and there will not be enough reference data to calibrate ASR. It is worth noting that the length of the reference data found will vary with the noise level of the data.

3.3.2 Determine thresholds for identifying artifact components

ASR applies a carefully-designed IIR filter to the reference data X_r to suppress specific frequency-band activities typically associated with brain oscillations, obtaining $\widetilde{X}_r$. ASR computes the mixing matrix M_r, i.e. the square root of $\mathrm{Cov}(\widetilde{X}_r)$, and the eigenvalue decomposition of M_r to obtain the eigenvectors matrix V_r and eigenvalues vector D_r. Each column in V_r is the eigenvector corresponding to the eigenvalue in D_r. Once the data are projected onto the PC space $\widetilde{Y}_r = V_r^T \cdot \widetilde{X}_r$, ASR calculates the mean μ_i and standard deviation σ_i of RMS values across all 0.5-second windows of $\widetilde{Y}_r$ for each component i, and defines rejection thresholds $\Gamma_i = \mu_i + k \cdot \sigma_i$ where k is the user-defined cutoff parameter.

3.3.3 Reject artifact components and reconstruct cleaned data

ASR applies an eigenvalue decomposition to the covariance matrix taken across channels of the IIR-filtered uncleaned EEG segments $\mathrm{Cov}(\widetilde{X}_t) = V_t D_t V_t^T$ along a sliding window with a window size of 0.5 seconds and a step size of 0.25 seconds. The IIR filter here is the same as in step (2). For each window, ASR identifies whether j^{th} PC $(V_t)_j$ with variance $(D_t)_j$ is larger than the rejection thresholds Γ_i projected from V_r onto V_t: $(D_t)_j > \sum_i (\Gamma_i (V_r)_i{}^T (V_t)_j)^2$. If the inequality holds, then the values of that component's activities are replaced with zero vectors: $(V_t^T M_r)_{trunc}$. Finally, ASR reconstructs the cleaned data segment using the equation:

$$(X_t)_{clean} = M_r (V_t^T M_r)^+_{trunc} V_t^T X_t \tag{3.1}$$

The MATLAB scripts for performing ASR are available as an open-source plug-in function `clean_rawdata` in EEGLAB [85]. While many settings can be optimized, the most

important user-defined parameter is the cutoff parameter k for determining the rejection thresholds in units of standard deviations. This study aims to characterize the effectiveness of ASR in removing artifacts and how k affects its performance.

3.4 Evaluating performance of ASR using independent component analysis

ASR is a general-purpose automatic artifact removal method that can be used as a preprocessing step. It is therefore important to maximize the artifact signals rejected while minimizing the brain activities removed. Without the ground truth in actual experimental EEG data, we propose to use this trade-off between removal of artifact and brain signals, separated by ICA and identify by an automatic IC classifier (*ICLabel*) [80], as an objective and quantitative measure of the performance of artifact removal method.

3.4.1 Independent Component Analysis (ICA)

ICA has been widely used for separating stereotyped brain processes and various types of artifacts such as muscle, eye-blink, and lateral eye-movement activities [54]. ICA assumes EEG data, x, can be modeled as a linear mixture A of statistically independent sources, s, and learns an unmixing matrix, W, such that the independent components (IC) recover the original sources, y.

$$x = As$$

$$y = Wx \approx s$$

In this study, we employ extended Infomax ICA [86], which is available in the `runica` function in EEGLAB, an open-source Matlab Toolbox.

3.4.2 Changes in spatial distribution and temporal activities of independent components

With the ICA decompositions of ASR-cleaned data, we can quantitatively assess the extent to which ASR affects the activities of the brain and artifactual ICs in two ways. First, we compute the component-wise correlation coefficients of the best-matched ICs across ICA decompositions of EEG data with and without ASR cleaning, that is, rearranging the order of ICs to maximize $\sum_i \mathrm{Corr}((A^{(k)})_i, (A^{(*)})_i)$ where A refers to the linear mixing matrix of ICA, k refers to ASR's cutoff parameter, $*$ refers to no ASR cleaning, and i is the column index. The matching process was performed using the Hungarian method [87] in the `matcorr` function in EEGLAB. This enables assessment of the stability of ICs across different ASR thresholds (k) by examining whether ICs disappear, change, or remain the same. Second, we apply the spatial filter $W^{(*)} = (A^{(*)})^{+}$, obtained from the ICA decomposition of raw data, to ASR-cleaned data $X^{(k)}$. Then we calculate the IC activities,

$$Y^{(k)} = W^{(*)} X^{(k)} \tag{3.2}$$

and compare the mean power reduction for the IC activities.

$$\text{Power reduction} = \mathrm{Mean}(\mathrm{Var}(Y^{(k)})) - \mathrm{Mean}(\mathrm{Var}(Y^{(*)})) \tag{3.3}$$

Since ASR removes artifacts based on the variance of component activities, it could remove both brain and artifact activity, especially when an aggressive threshold is applied. By examining the power reduction of the activities of brain-related and non-brain-related ICs, we can reveal the effectiveness of ASR for removing non-brain signals while retaining brain-related signals and provide guidelines for choosing the cutoff parameter k.

3.4.3 IC classification

To summarize the ICA results across subjects, we classify the ICs from each decomposition using an automated IC classifier. We utilize the `iclabel` function in EEGLAB [80] to classify ICs into 7 classes: Brain, Eye, Muscle, Heart, Line noise, Channel noise, and, a class for ICs which do not fit into the first 6 classes: Other.

3.4.4 Changes in dipole fitting result of ICs

The quality of an ICA decomposition can be measured by the number of dipolar ICs, whose spatial distribution over the scalp can be modeled by a current dipole in the brain, as suggested by Delorme et al [88]. Dipolar ICs are important because EEG is assumed to capture brain signals from near-synchronous cortical patches, generated by well-aligned pyramidal cells. With this assumption, it is reasonable to model each brain component in EEG recordings as an equivalent current dipole in the cortex [89]. Since large-amplitude artifacts usually disrupt ICA decompositions, we expect that ICA will find more dipolar ICs if the artifacts are removed from the data. In this study, we employ the `dipfit` function in EEGLAB [90] and consider ICs with residual variance, the mismatch between IC's spatial distribution over the scalp and the projection of fitting dipole, lower than 5% to be "dipolar sources" [91, 88].

3.5 Results

3.5.1 Data modification and variance reduction through ASR cleaning

Fig. 3.2 shows the percentage of data points modified by ASR (i.e. rejecting at least one component) and the average variance reduction of the data before and after ASR cleaning using different cutoff parameters k. The average portion of the reference data selected by ASR across twenty EEG recordings is 43.9% with a standard deviation of 14.6%.

In Fig. 3.2(a), when the cutoff parameter $k = 100$, less than 3% of data was modified while still reducing variance by more than 20%. When k is between 5 and 7 as previously suggested in

[73], ASR modified nearly 80% of data and reduced 80% of signal variance.

Fig. 3.2(b) shows that the percentage of data modified and variance reduced started to increase when $k \leq 30$. When k is between 5 and 7, ASR modified 50% of reference data and reduced the signal variance by 30%. One thing to note is that, by visual inspection, there are still some eye and muscle activities in the reference data and ASR starts to reduce variance when the threshold falls below $k = 1000$ in Fig. 3.2(b).

3.5.2 Stability of ICs across choices of the ASR parameter

We examined the stability of IC by calculating component-wise correlation coefficients across ICA decompositions of the EEG data with and without ASR cleaning.

Through visual inspection of the results shown in Fig. 3.3, we found that those ICs which were preserved by ASR with $k = 1$ (shown in the green box in Fig. 3.3(b)) were likely to be associated with brain activities (IC2, IC4 and IC5). These ICs were visually characterized by spatially homogeneous scalp maps [88]. Interestingly, the ICs accounting for eye-blink (IC7) and eye-movement (IC3) activities were also consistently present when different values of k were used. On the contrary, those ICs which disappeared when the value of k was smaller than 70 were likely to account for artifacts due to single-channel noise (IC25 and IC29) or localized muscle activities (IC27 and IC28), visually characterized by scalp maps with sparse and localized activity.

To quantify the above results across subjects, Fig. 3.4(a) depicts the percentage of preserved ICs at each ASR threshold from all twenty EEG recordings, categorized into 5 groups using *ICLabel* classifications (see Section 3.4.3) and IC dipolarity (see Section 3.4.4). Fig. 3.4(a) shows that, when $k = 5$, ASR altered 50% of ICs. However, the ratio of Dipolar brain sources in preserved ICs increased, compared to the ratio in ICs without ASR cleaning, from 20% to 30%.

Fig. 3.4(b) shows the percentage of preserved ICs within each group at each ASR threshold. When $k = 100$, almost 20% of ICs in the Eye and Muscle classes were removed. When $k \geq 20$, 90% of Dipolar brain ICs were preserved while less than 70% of ICs in the other four

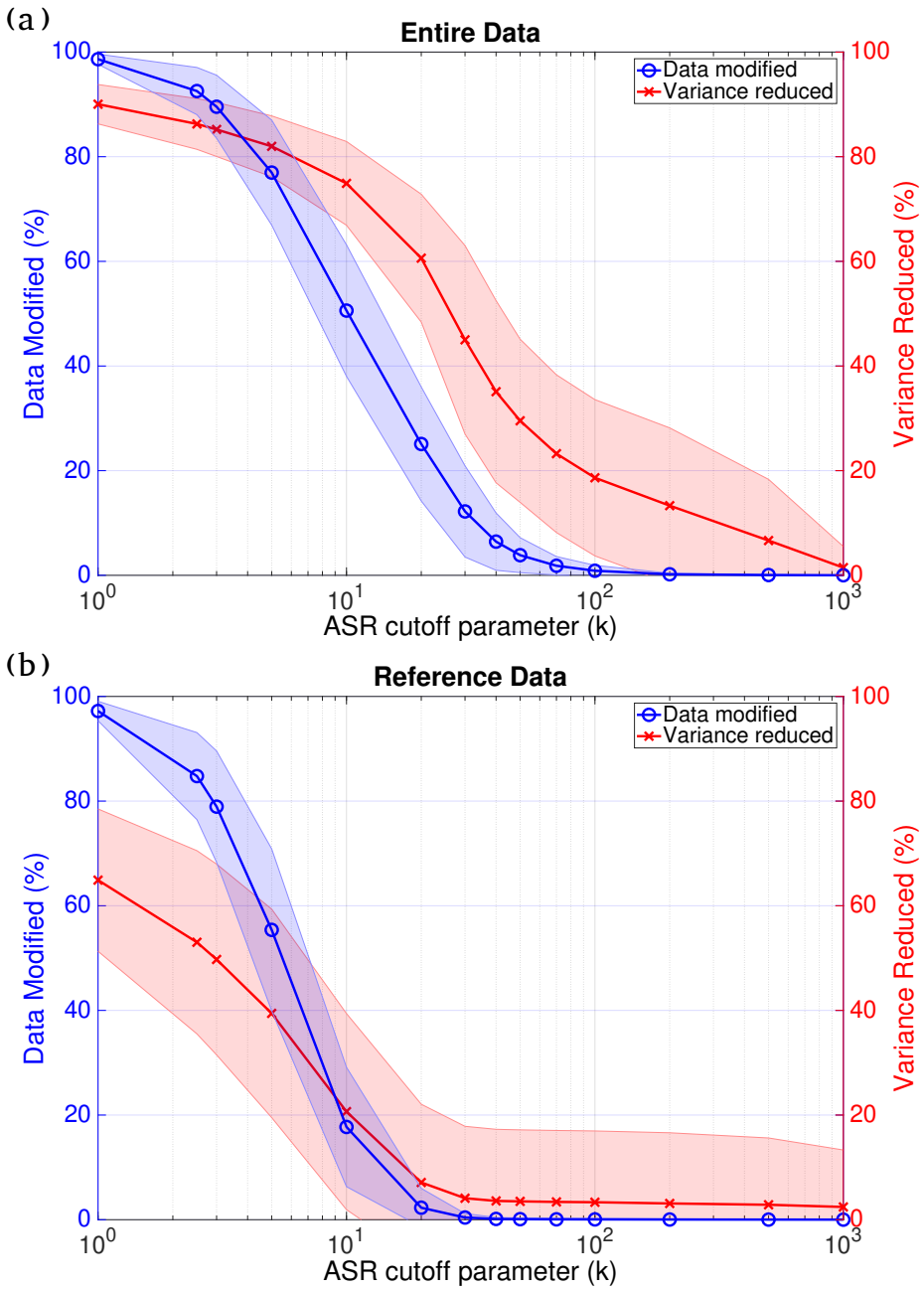

Figure 3.2. The percentage of data modified (blue) and variance reduced (red) by ASR with different cutoff parameters with respect to the same data with no ASR cleaning. The shaded area shows one standard deviation across twenty EEG recordings. Fig. 3.2(a) shows the result on the entire data and Fig. 3.2(b) only shows the result on the reference data, which ASR used to determine the value of thresholds.

classes were preserved. When using k between 5 and 7, less than 60% of Eye ICs and 50% of Muscle and Other ICs remained, but also removed 15%–25% Dipolar brain ICs.

3.5.3 Source power reduction by ASR cleaning

To further quantify how different types of signals were removed by ASR, we calculated the ICA decomposition of EEG data without ASR cleaning and applied the learned spatial filters (i.e. the IC scalp maps) to the same data after ASR-cleaning data, and computed their source activities retained after ASR cleaning with different cutoff parameters.

Although eye-blink ICs (IC7) and eye movements ICs (IC3) were still present as reported in Section 3.5.2, Fig. 3.5(a) shows that their power was reduced to the same level as those of brain sources when k is between 5 and 7. Moreover, Fig. 3.5(b) shows that, when k is between 5 and 7, eye-related and likely-artifact ICs only retained 5% of their power after ASR cleaning. On the other hand, 70% and 90% of the power of the brain-related ICs (IC2, IC4, IC5) were retained when $k = 5$–7 and 30, respectively. Even though IC1 was preserved when $k = 1$ in Fig. 3.3, ASR removed 65% and 40% of IC1's power with $k = 5$ and $k = 30$, respectively.

These single-subject results were also seen across subjects. Fig. 3.5(c) plots the source power of each of the 3 classes (Brain, Eye, and Muscle) averaged over all ICs in the same class across all twenty EEG recordings. To prevent the shaded area from exceeding the range of percentages (0 through 100), the shaded area shows 10% through 90% quantiles instead of standard deviation. The source power of ICs in the Eye class was 10 times larger than those in the Dipolar brain class in the data without ASR cleaning. However, when $k \leq 10$, the source power of the Eye class and Dipolar brain class were comparable. On the other hand, the source power of ICs in the Muscle class was comparable to the source power of ICs in the Dipolar brain class in data without ASR cleaning when k was large, but became 9 times smaller when $k \leq 100$.

Fig. 3.5(d) shows that when $k = 100$, ASR removed on average 30% of the source power of ICs in the Muscle class, and 10%-90% quantiles show that ASR's effectiveness varied drastically across Muscle ICs. When $k \geq 30$, ASR retained 90% of the power of Dipolar brain ICs, while

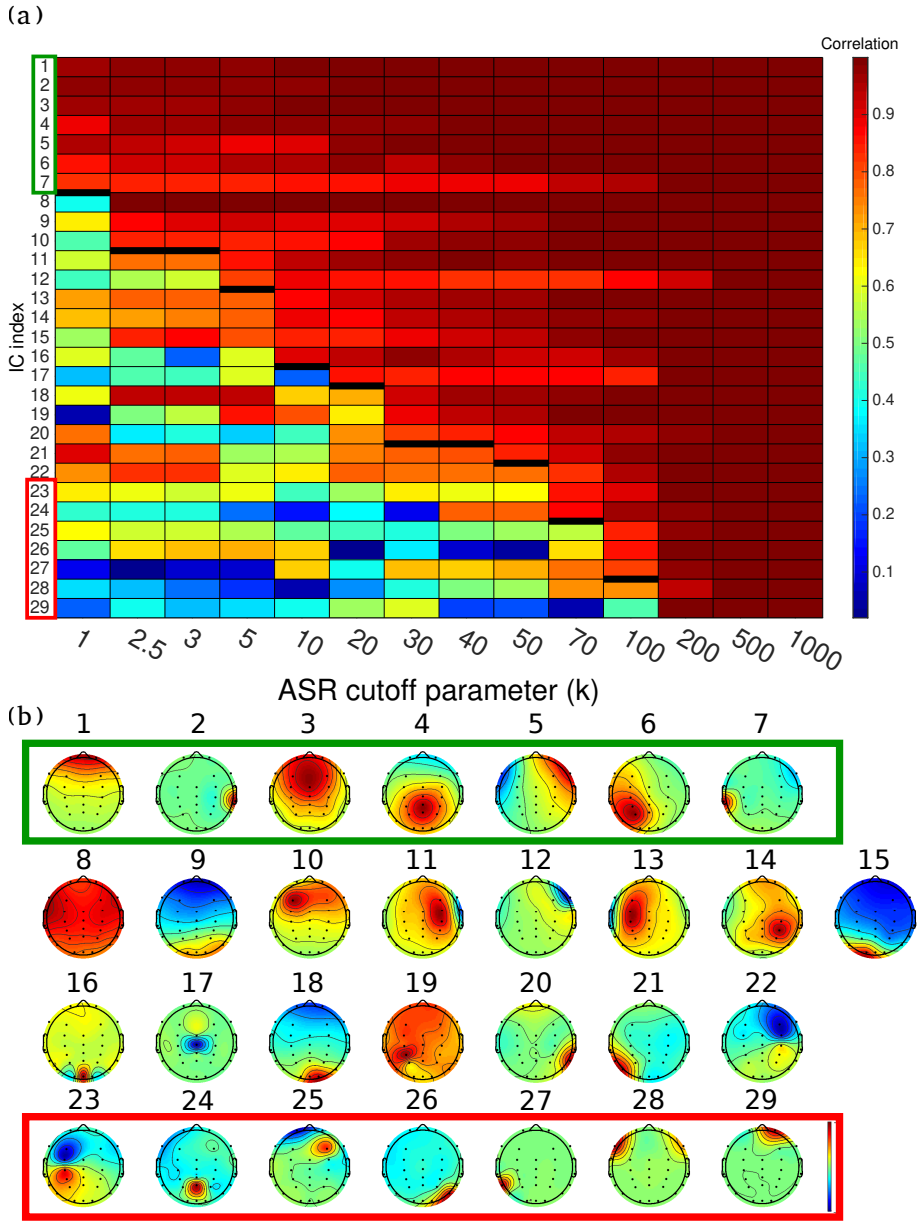

Figure 3.3. (a) Component-wise correlation coefficients of the best-matched ICs between ICA decomposition of the EEG data with ASR cleaning across different ASR cutoff parameters and without ASR cleaning. The IC index is sorted by whether ICs' correlation coefficient is higher than 0.8, indicated by the black lines. (b) The scalp maps, i.e. the spatial distribution of each source activities over the scalp channels, of the no-ASR-cleaned ICA decomposition (template ICs). The green box indicates preserved ICs after ASR cleaning while the red box indicates ICs which disappeared when $k \leq 50$.

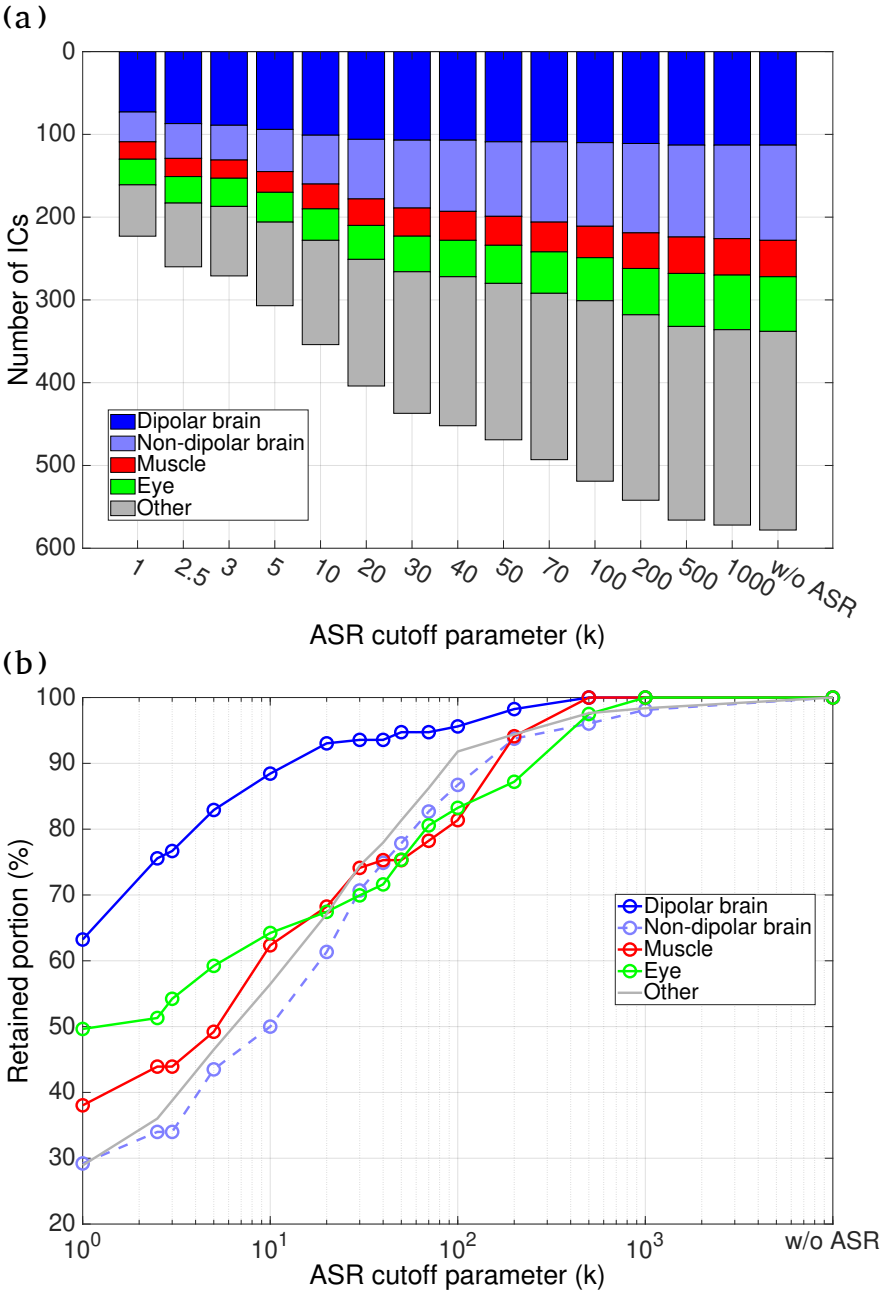

Figure 3.4. (a) Number of preserved ICs which have correlation coefficient higher than 0.8. The classification result of Dipolar brain, Non-dipolar brain, Muscle, and Eye are shown in blue, light blue, red, and green respectively. ICs outside these four classes are labeled as Other in gray. ICs with residual variance $< 5\%$ were labeled as dipolar. (b) The percentage of preserved ICs within classes.

only retaining 50% of the power of Eye and Muscle ICs. When using ASR with k between 5 and 7, the retained power in Eye and Muscle ICs were 10% and 30% respectively. However, 40% of the source power of Dipolar brain ICs was removed as well.

3.5.4 Improvement of ICA decomposition

Fig. 3.3 and 3.4 provide qualitative and quantitative results of which ICs survive ASR cleaning. To further assess ASR's effect, Fig. 3.6(a) reports the total number of dipolar ICs present after different levels of ASR cleaning. When $k \leq 50$, the ICA decomposition of the ASR-cleaned data found significantly more dipolar sources, which indicates a better decomposition according to [88]. Furthermore, Fig. 3.6(b) shows that the number of Dipolar brain sources increased, on average, by 10% when $k = 5$ and 5% when $k = 20$.

3.6 Discussion

Artifact Subspace Reconstruction is an automatic, online-capable artifact removal method that has been increasingly used in EEG pre-processing. However, ASR has not been properly validated and the optimal user-defined cutoff parameter is unknown. This study aims to systematically evaluate and quantitatively assessed the effectiveness of ASR on real EEG data using ICA decomposition with the following measures: (1) percentage of data modification versus variance reduction, (2) percentage of reference data that are affected, (3) how many artifact ICs remain and how much their powers are reduced, and (4) how many brain ICs are preserved and their source activities are affected.

The empirical results show that the effectiveness of ASR heavily depends on the choice of its cutoff parameter k. As shown in Fig. 3.2(a), a mild threshold ($k = 100$) could remove sparse (1% of data) yet large-amplitude artifacts (20% of the variance). When k was 20, ASR started to affect the reference data, indicating that even the clean data ASR used to determine thresholds were modified. With the previously suggested values (k between 5 and 7) [73], ASR modified 70% of data and removed up to 80% of the variance, which may affect brain signals and distort

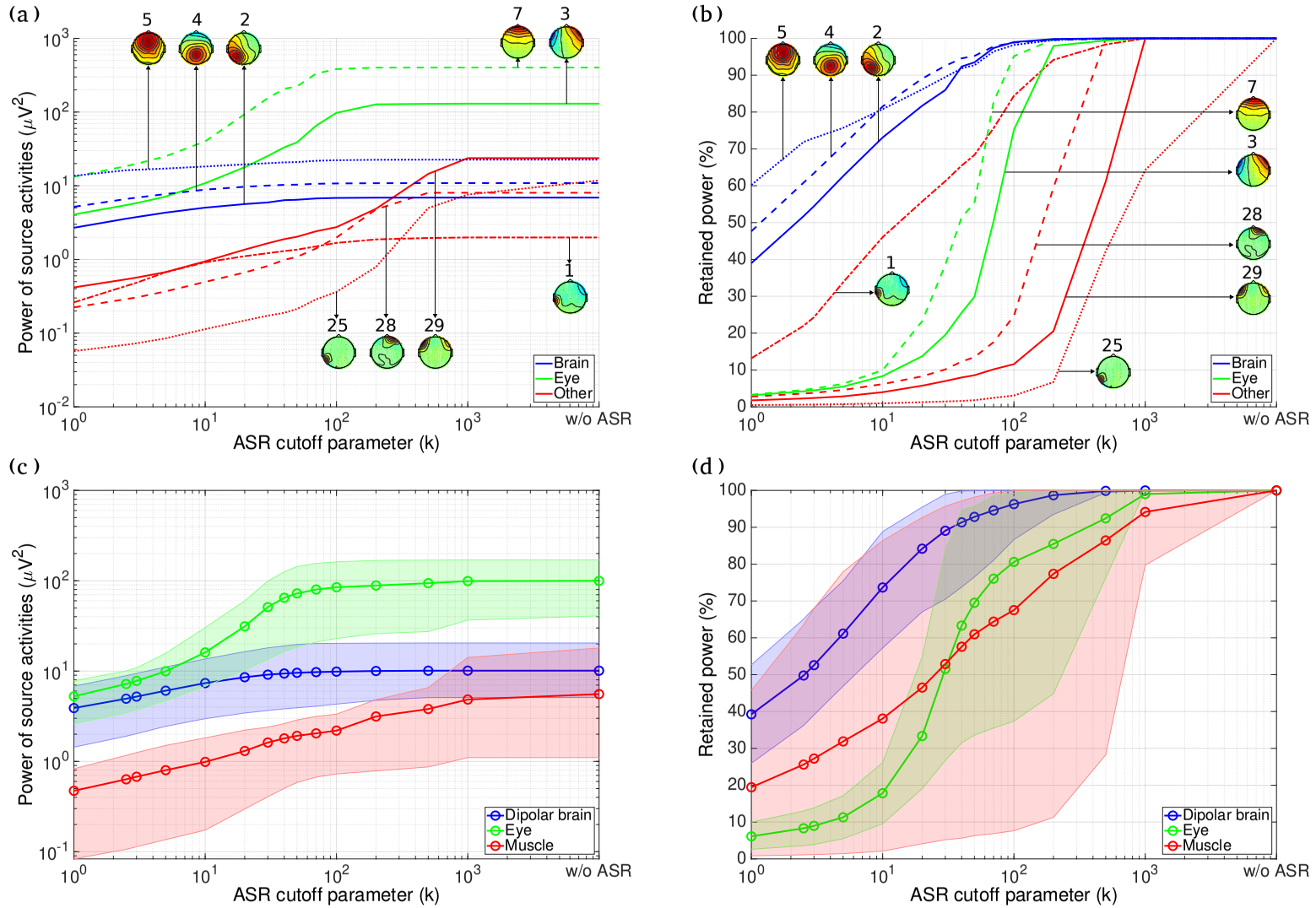

Figure 3.5. (a) The power of source activities of selected ICs in Fig. 3.3 when different ASR cutoff parameters were applied. These ICs were classified as Brain (blue), Eye (green), and Other artifact (red) by visual inspection. The scalp maps and indexes of those ICs are also shown. (b) The percentage of retained source activity power of the same selected ICs from (a). (c) The average power of source activities of ICs from all subjects. The ICs were classified by the *ICLabel* classifier and the resultant Brain class (blue), Eye class (green), and Muscle class (red). The shaded areas represent 10% through 90% quantiles. (d) The percentage of retained power of source activities of the same classified ICs from (c).

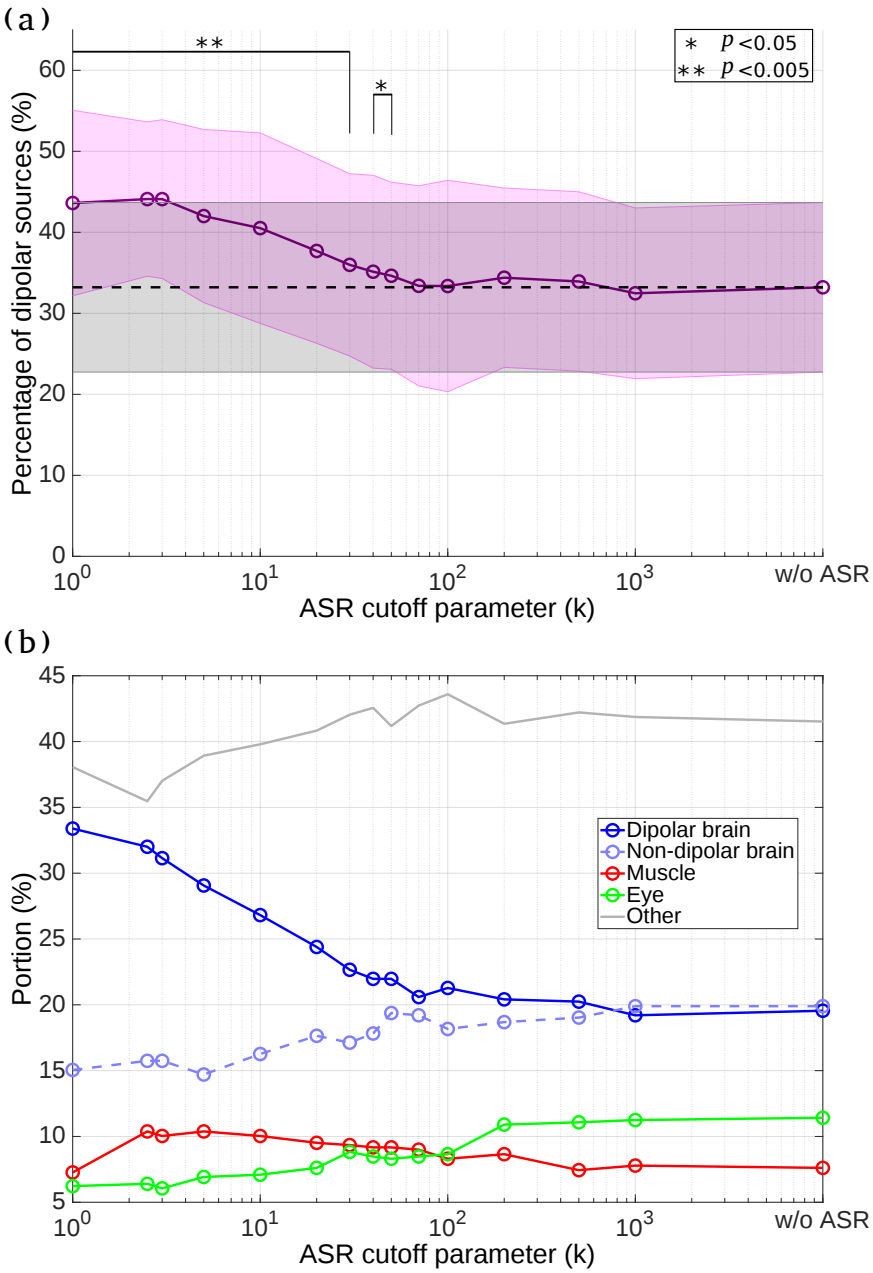

Figure 3.6. (a) The average percentage of all dipolar sources in ICA decomposition of EEG data with (solid line) and without (dashed line) ASR cleaning. The shaded area represents one standard deviation across subjects. The statistical significance between the number of dipolar sources with and without ASR is calculated by bootstrap. (b) The percentage of ICs in each IC class after ASR was applied with different cutoff parameters. The classification results into the categories of Dipolar brain, Non-dipolar brain, Muscle, and Eye are shown in blue, light blue, red, and green respectively. ICs outside these four classes are labeled as Other in gray. ICs with residual variance $< 5\%$ were considered dipolar.

experiment results.

To assess the types of signals removed by ASR, we decomposed the ASR-cleaned EEG signals using ICA and classified the ICs as Brain, Eye, and Muscle-related sources. We found that more Muscle, Eye, and Other ICs disappeared than the Dipolar brain ICs did after ASR cleaning (Fig. 3.4(b)). When $k \leq 20$, more Dipolar brain ICs were affected and the ratio of removing artifact ICs versus Dipolar brain ICs deteriorated. Although some Muscle and Eye ICs were still present after ASR cleaning, their powers were strongly reduced (Fig. 3.5). The retained power from the Eye ICs went from 80% ($k = 100$) to below 20% ($k = 10$); the retained power from the Muscle ICs went from 93% ($k = 1000$) to 40% ($k = 10$). When $k \geq 30$, 90% of power from the Brain ICs was still preserved, but the retained power decreased to 50%–60% with k between 5 and 7.

Given the above observations, the recommended ASR cutoff parameter k is between 20 and 30. ASR with a conservative threshold $k = 30$ removed 25% of the Eye and Muscle ICs and reduced almost 50% of the power of the Eye and Muscle activities while only affecting less than 10% of the Dipolar brain ICs and removing only 10% of their power. ASR with a lower threshold of $k = 10$ further removed 15%–30% of Eye and Muscle activities, but at the cost of reducing 15% more Brain signal power. The previously suggested value of k between 5 and 7 is too aggressive in removing both artifact and brain signals and is not recommended.

Interestingly, ICA decompositions of the ASR-cleaned data found more Dipolar Brain sources when a smaller cutoff parameter was applied. Because ICA is sensitive to large-amplitude artifacts, applying ASR before ICA can increase the quality of an ICA decomposition, as shown by the increase in the number of dipolar EEG sources found (Fig. 6). In addition, we examined whether the newly found Dipolar Brain sources were "novel" — different from the ones found in ICA without ASR cleaning — by comparing the correlation of their spatial topographies. We found that several ICs from the "Non-dipolar Brain" and "Other" classes became "Dipolar Brain" classes after ASR cleaning while most of the Dipolar Brain ICs remained the same. We conceive that spatially diffused or mixed signals originally explained by one IC could be further

decomposed into a few more localized ICs when ASR cleaning was applied. Furthermore, we examined the percentage of ICs with dipole locations outside versus inside the MNI head model. The empirical result shows that only when aggressive ASR thresholds (equal or below 5) were used, the number of ICs outside of the head model reduced, implying more artifact components were removed.

Compared to other existing artifact removal methods, the benefit of ASR is that it can automatically adapt its thresholds based on the statistics of the EEG data. Moreover, ASR can remove transient, large-amplitude artifacts which ICA-based methods are usually incapable of dealing with. In fact, the combined use of ASR and ICA might be even more effective in removing different types of EEG artifacts [92].

Even though ASR effectively removes large-amplitude artifacts, its limitations and potential solutions should be considered. Firstly, without aggressive cutoff parameters, ASR might be limited in removing regularly-occurring yet large-amplitude artifacts, such as eye blinks and temporalis muscle activities, for these signals would be inevitably included in the automatically-extracted reference data. As a possible remedy, especially for eye-related artifacts, one can use a user-defined, artifact-free data segment as reference data, or use a subsequent ICA-based artifact removal methods to complement ASR cleaning, as proposed in [92]. Secondly, ASR could also remove brain signals if they are not present in the reference data, which can happen in an online application. One potential solution is incorporating a clean portion of the incoming data to reference data and having an adaptive ASR threshold.

Another limitation of ASR is that it cannot be applied to single-channel EEG recordings and its performance might be impaired when the number of channels is small. Unlike EEMD- and SSA-based methods, which can exploit the temporal or spectral information in single-channel EEG, ASR requires the covariance of multi-channel signals to obtain artifact rejection threshold and reconstruct the clean signals. The effect of channel count on ASR performance is an important topic that applies to all component-based methods like ICA. From our empirical observation, ASR could achieve effective artifact removal for standard 20-channel EEG.

ASR can also be a powerful tool for online, near-real-time, automatic artifact removal, which has been implemented and disseminated in BCILAB [93] and in the Real-time EEG Source-mapping Toolbox (REST) [94]. Some important considerations of using ASR in the online application include: (1) online application of ASR requires user-defined reference data; we suggest users record an artifact-free section before the experiment. (2) Bad channels (e.g. with high impedance) should be removed before applying ASR. (3) The computation bottleneck of ASR is the eigenvalue decomposition of the sample covariance matrix, which has the complexity of $O(n^w)$ where n is the number of channels and $2 < w < 3$ depending on the implementation of the algorithm. In case of increased channel counts, a longer time window can be used to achieve real-time capability. More details of the online aspect of ASR can be found in [92].

3.7 Conclusion

This study demonstrates that Artifact Subspace Reconstruction is an effective automatic artifact removal approach, quantifies ASR's effectiveness in removing different types of signals as shown using Independent Component Analysis, and provides insights into the optimal choice of ASR's cutoff parameter.

Our empirical results suggest using a cutoff parameter between 20 and 30 rather than the previously suggested and default values between 5 and 7 [73] where brain activities were excessively removed. This study also found that ASR improves the quality of ICA decomposition as evidenced by an increased number of dipolar independent components.

Offline and online versions of the ASR code have been made available in EEGLAB and BCILAB. With an appropriate choice of the cutoff parameter, ASR can be a powerful artifact removal approach for subsequent data analysis such as ICA and its online capability enables real-time artifact rejection for brain-computer interfaces and clinical applications.

3.8 Acknowledgement

Chapter 3, in full, has been published as Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini, and Tzyy-Ping Jung, ”Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings”, in IEEE Transactions on Biomedical Engineering, 2019. The dissertation author was the first investigator of this paper. The authors gratefully acknowledge the valuable suggestions and comments from Makoto Miyakoshi and the author of ASR, Christian Kothe.

Chapter 4

Electroencephalography in Mobile Virtual Environments

4.1 Introduction

The fixation-related potential (FRP) technique is part of a broader class of event-related EEG methods whereby segments of time-series EEG associated with distinct classes of experimentally relevant events are aligned and averaged. In the basic ERP paradigm, time-locking events usually coincide with the presentation of some type of stimulus, such as a sound, an image, a video, and so forth. In the case of FRPs, the time-locking event is the onset of visual fixation.

This method has been used extensively to study reading and visual search, among other human abilities, in part because it allows for more ecologically valid measures of neurocognitive processing. In classic ERP language paradigms, for instance, the individual words comprising a text are presented serially on a computer monitor – one item at a time – allowing experimenters to time-lock to brain activity elicited by each individual word. However, because the FRP technique accomplishes time-locking through information about a reader's gaze rather than the onset of a stimulus, it is possible to present full blocks of text at once in FRP paradigms, approximating more natural reading conditions. This approach allows researchers to study the effect of parafoveal words (e.g., the words immediately to the right of a currently fixated foveal word) on foveal word processing [95, 96]. Analogously, FRP-based visual search paradigms allow researchers to dissociate brain activities elicited during the natural scanning of targets

versus distractors within a single complex visual scene [97].

Crucially, successfully deriving FRPs from time-series EEG depends on the accurate and precise demarcation of fixations from time-series gaze points. A significant challenge stems from the fuzziness of boundaries between saccades and fixations. Efforts to parse these two fundamental components of gaze can lead to variable outcomes depending on the parameters and structure of the parsing algorithm and noise in the signal. Many FRP protocols include the use of chin rests or other restraints to curtail the intrusion of head movement, which can result in data loss or distortion due to shifting of the head outside of the headbox in the case of remote eye trackers or slippage in the case of mobile systems [97]. Further, head movement can complicate the parsing of fixations and saccades due to vestibular-ocular reflex (VOR), which allows an individual to maintain their gaze on a fixed object under conditions of head movement using compensatory eye movements. Under conditions of such VOR-based responses, the change of the position of the eye-in-head during fixation can actually exceed the change of eye position that occurs during small saccades [98], rendering dispersion-based classifications algorithms – which rely on changes in the location of gaze points – ineffective. For these reasons, many eye-tracking studies take careful precautions to minimize head movement, often confining their tracking area to the 2D display on a computer monitor.

As head-mounted virtual reality (VR) with integrated mobile eye tracking has become more widely available, new possibilities for studying unconstrained, free viewing in 3D space have emerged. In head-mounted VR, visual scenes are presented binocularly via a flat screen system embedded inside a goggle-style face covered with the stereoptic simulation of depth cues. By continuously tracking the position and orientation of the user's head, the device can update the visual scene in response to the user's movements. Further, in addition to the user's head position and orientation, the distance between the user and objects and other features of the game world is known at all times, making it possible to compute gaze points as 3D vectors. A gaze intersection point (GIP) is the nearest point in the virtual environment that is intersected by the vector-based gaze ray. Thus, in virtual paradigms, it is possible to parse fixations according to

either traditional methods that measure changes in visual angle between eye position coordinates or alternatively, by relying on GIP-based approaches.

The present study evaluates the feasibility of a new event-related measure derived by time-locking to EEG data that coincide with periods when the user's gaze intersects with an object of interest – the so-called gaze-related potential (GRP). We simultaneously recorded EEG and eye-tracking data while healthy adults performed a simple target detection task in an immersive virtual scene. In VR, participants were asked to fixate on a central marker and then either to saccade to a target appearing in one of four locations in the near periphery (5 degrees from the central marker) or make a head turn and a saccade to targets in one of four locations in the far periphery (30 degrees from the central marker). Embedded in each target was either a standard cue (occurring on 80% of trials) or a deviant cue (occurring on 20% of trials), yielding a classic oddball paradigm. Participants were instructed to classify cues by pulling the trigger of their controller.

Our research goals were twofold. First, we aimed to characterize the interaction between head and eye movement when participants directed their gaze to far peripheral targets. Secondly, we aimed to compare the effectiveness of FRP versus GIP time-locking approaches in detecting visual P300 oddball effects under conditions of head movement. To our knowledge, this study is one of the first of its kind to examine gaze-based approaches to EEG analysis in 3D, immersive VR with free head movement.

4.2 Experiment and data acquisition

4.2.1 Data acquisition

In this study, we used a head-mounted VR system, an HTC VIVE Pro Eye, HTC Corp., New Taipei, Taiwan), to display the visual oddball task in a virtual environment and track the movements of the head and eyes. In the head-mounted virtual reality system, the head movements, including head locations and head rotations, were recorded using the inertial measurement unit

(IMU), and the eye tracking data, including pupil directions and gaze intersection points (GIPs), were recorded using the integrated camera-based eye trackers. Both head movements and eye tracking data were recorded at a 90Hz sampling rate. We elaborate on the definition of GIP in section 4.2.2. We recorded the EEG signals simultaneously using a 22-channels Smarting semi-dry headset (mBrainTrain LLC, Belgrade, Serbia) with salt-water-soaked sponge electrodes and a 500Hz sampling rate for the first 6 subjects, and a 64-channel BioSemi wet headset (BioSemi B.V., Amsterdam, Netherlands) with gel electrodes and a 512Hz sampling rate for the other 8 subjects. Both headsets have their electrodes placed on the scalp following the international 10-20 system [84]. We also recorded subjects' responses using an HTC VIVE controller. The virtual environment was built in the Unity game engine.

4.2.2 Gaze Intersection Point (GIP)

Gaze Intersection Point (GIP) indicates where and what a subject is looking at by reporting the intersection points of the gaze vector and the virtual environment. For each eye, the eye tracker describes eye movements by the location of the pupil. To take the blink into consideration, the eye tracker uses an eye openness index to indicate eye tracking reliability. Next, the eye tracker defines a 3D pupil vector that starts at the center of the modeled eyeball (perfect sphere) and points outward through the pupil for both eyes independently. The pupil vector reports -1, pointing inward when the eye is closed. In normal situations, these two pupil vectors collide when the subject looks at objects. Gaze vector is defined as the vector pointing from the middle of two eyes toward the colliding point. Finally, the eye tracker extends the gaze vector until it hits objects or boundaries in a virtual environment, and defines the intersection point as GIP. Fig. 4.1 illustrates the definition of GIP. Since the source code is unavailable, we hypothesize the eye tracker will use the pupil vector with a higher eye openness index than the gaze vector.

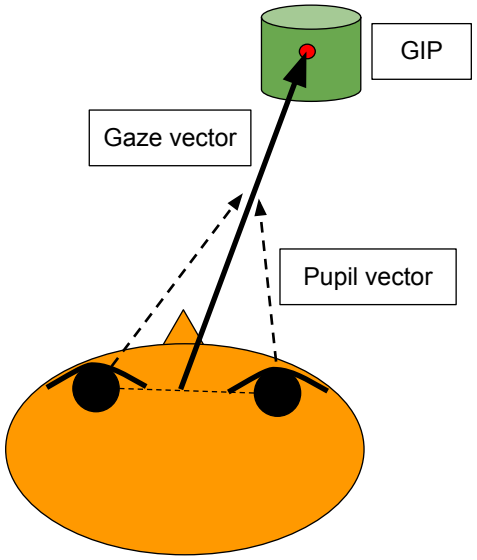

Figure 4.1. Illustration of Gaze intersection point.

4.2.3 Virtual environment

Two scenarios in the virtual environment correspond to two conditions in the experiments. For both scenarios, there is a blue cube displayed in front of the subject at the height of eye level which serves as a central fixation point for the experiment. Two possible concentric rings of target positions were arranged around the central fixation point perpendicular to the subject, with a low-excursion ring with a radius equal to 0.0875 meters and a larger, high-excursion ring with a radius equal to 1.1548 meters. Depending on the condition, visual targets would appear on either of these rings at 90° coordinate intervals: up, right, down, and left. Fig. 4.2 shows the representation of the visual field of the two conditions. For the low-excursion ring condition, subjects were asked to stand 1 meter away from the central fixation point, creating a visual angle of 5 degrees between the central fixation point and target cubes. For the high-excursion ring condition, subjects were asked to stand 2 meters away from the central fixation point, creating a visual angle of 30 degrees between the central fixation point and target cubes. All of the cubes in the experiment have a width of 11.04 cm.

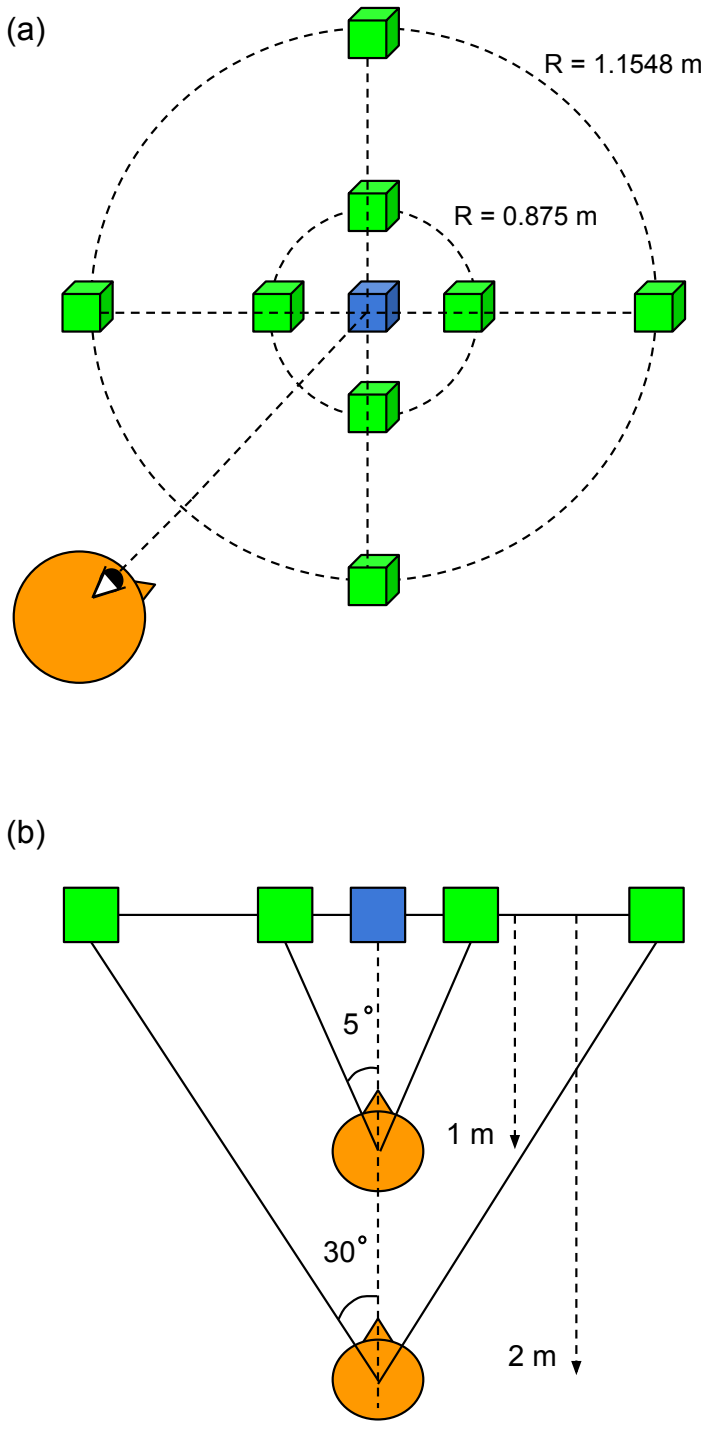

Figure 4.2. The representation of the visual field in the virtual environment. Fig. 4.2(a) shows the locations of the central fixation point (blue cube) and possible target positions (green cubes) for both low-excursion and high-excursion rings. Fig. 4.2(b) shows the top view of the representation and indicates the visual angles corresponding to the subjects (orange head).

4.2.4 Experiment design

To characterize the relationship between the eye and head movements, we designed a visual oddball task with cued eye and/or head movements inside a virtual environment. The oddball task is commonly used in ERP studies [99]. In an oddball task, a sequence of repetitive stimuli (standard stimuli) with a few different stimuli (deviant stimuli) randomly inserted is presented to subjects. Once subjects receive deviant stimuli, their brain will generate an ERP named P300 [100]. In our experiment, we used P300 as the ground truth brain signal.

Our experiment has two conditions: (1) eye-shifting condition using the low-excursion ring, where subjects were asked to use their eyes to position their gaze without head movements. (2) head-turning condition using the high-excursion ring, where subjects can naturally position their gaze with both eye and head movements.

At the beginning of each trial, subjects were asked to stand 1 or 2 meters away from the central fixation point (blue cube) for eye-shifting and head-turning conditions respectively with their gaze fixated on the blue cube. After 2 seconds, a green target cube popped up at one of the four possible positions. Subjects were asked to position their eye gaze on the target cube as soon as possible and fixate their eye gaze on it. A triangle or a circle is printed on the target cube, representing deviant and standard events respectively. Subjects were asked to pull the trigger on the controller whenever they could justify that the shape of the target cube is a circle. There are 20 deviant events and 80 standard events for both conditions. The target cubes disappeared after 2 seconds and subjects were asked to position their eye gaze back to the central fixation point to finish the trials. The conditions are presented in order of eye-shifting and then head-turning with a break between them. Fig. 4.3 shows the pipeline of one trial in the visual oddball task.

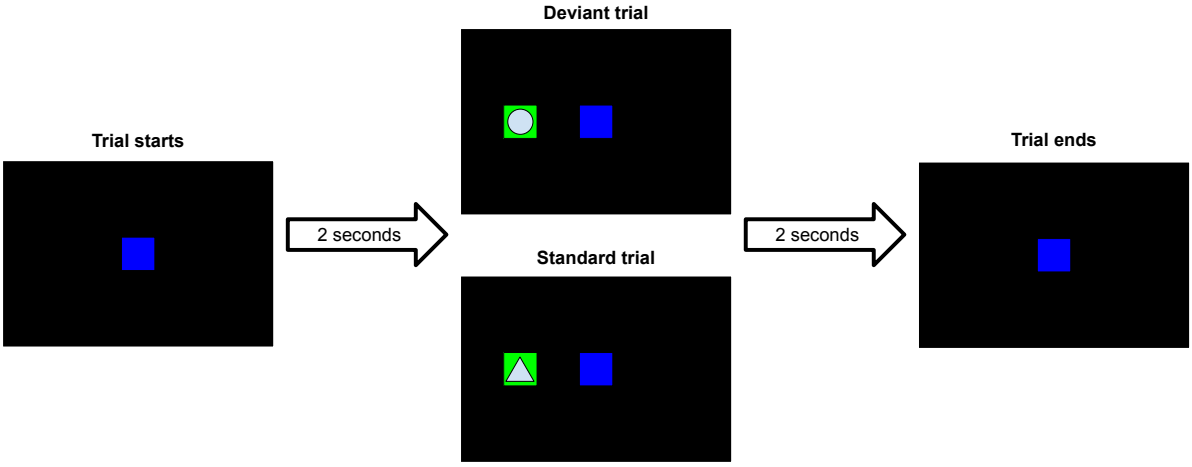

Figure 4.3. The pipeline of one trial in the visual oddball task.

4.3 Data preprocessing and methods

4.3.1 Eye tracking and behavioral data

The eye movements were calculated from the normalized pupil directions, defined as a vector pointing from the center of the eyeball to the pupil. To reduce the noise in eye movements, we preprocessed the data following the Tobii pipeline [101]. First, we determined the unreliable data based on eye openness, ranging from 0 to 1. Once the eye openness is smaller than 0.1, we labeled the period of data as gaps. If the latency of a gap is shorter than 75 ms, we linearly reconstruct the eye movements within the gap. Otherwise, the gap is defined as a blink or data missing when it has a latency between 75 ms to 150 ms or longer than 150 ms respectively. Then, We smoothed the rest of the data with a 40 ms window to get a robust estimation of eye movements. We also smoothed the head movements and GIPs with a 40 ms window, and no further preprocessing is applied.

4.3.2 EEG data

To remove high-frequency noise, we clean the EEG data using a band-pass FIR filter (1Hz - 50Hz). Then, we re-referenced the data to mastoids. Since two EEG recording systems have different sampling rates, we downsample the data to 250Hz. Next, we implemented

`clean_rawdata`, an EEGLAB plugin function [85], to remove channels with negligible activities (flat line threshold: 5), strong line noise (noisy line threshold: 4), or a poor correlation with adjacent channels (correlation threshold: 0.7). Once bad channels were removed, we applied Artifact Subspace Reconstruction (ASR) with a burst threshold of 10 using the same function [102]. Lastly, we performed Independent Component Analysis [54] and rejected artifact-like components labeled by `ICLabel` [80] (probability threshold: 0.8).

4.3.3 Dispersion- and Velocity-based fixation detection

Since eye movements have plenty of types, such as blink, fixations, saccades, and smooth pursuits, there is a wide variety of eye movement classifiers [103]. In this study, we focus on fixation detection to investigate the relationship between fixation onset timing and the oddball response of the brain and behavior. Most of the popular fixation detection algorithms use a threshold-based method on dispersion and/or the velocity of eye movements [104, 105].

We built a fixation detection algorithm based on the Tobii I-VT fixation filter [101]. First, we calculated the angular dispersion and the angular speed of eye movements. Next, we applied a threshold to determine the before-merged fixations. The optimal type and value of threshold choosing vary across different studies and tasks. [106, 107, 108] We implemented an angular-speed-based threshold of 30 degrees per second. After we calculated the before-merged fixations, we merged the adjacent fixations if the latency between them is shorter than 75 ms and the angle difference between them is smaller than 1 degree. Finally, we removed the fixations shorter than 200 ms after merging.

4.3.4 Event-related potential with different time locking

We extracted three types of event-related potential (ERP) with different event onset and epoch lengths. First, we extract the stimulus-locked ERP with an epoch length of 200 ms before and 2000 ms after stimulus onset. Next, we extract the GIP-locked ERP using the GIP onset, when GIP first collides with target cubes after stimulus onset. The GIP-locked epoch length is set

to 1000 ms before and after GIP onset. Finally, we extract the fixation-locked ERP using the first fixation onset after stimulus onset. The fixation-locked epoch length is equal to the GIP-locked epoch length. We used the conventional stimulus-locked ERP as a benchmark to compare with other types of ERP. All the ERPs in the same trials removed the same baseline (200 ms before stimulus onset). We further removed bad epochs if the latency between GIP onset and stimulus onset is shorter than 100 ms or longer than 1000 ms.

4.3.5 1D representation of 3D eye tracking and behavioral data

To investigate the dynamics of head movements and eye movements under head-turning conditions, we first projected 3D eye tracking and behavioral data onto the vectors pointing from the subject toward the target cubes for each trial independently. Next, we calculated the angle for head-turning directions, pupil vectors, and gaze vectors. The angle is created by the vectors before and after stimulus onset. At the beginning of each trial, all the angles remain zero (facing and looking at the central fixation point). When subjects turn their head toward the target cube, a positive angle is created. The definition of positive angle is the same for pupil vectors and gaze vectors. Finally, we calculated the angular velocity using a 40-ms sliding window. We also recorded the distance between GIP and the target cube to check if the distance reaches its minimum at GIP onset.

4.4 Results

4.4.1 Relationship among behaviors

Fig. 4.4 and Fig. 4.5 show the angle and angular velocity of 1D-projected behavioral data. The results of eye-shifting and head-turning conditions are shown in the left and right columns respectively. The events are time-locked to stimulus onset, GIP onset, and fixation onset from the top to the bottom. Since we found no significant difference between the results of standard trials and deviant trials, we presented the results from standard trials only. Fig. 4.4

shows subjects' gaze rested at 4 and 23 degrees after GIP onset in the eye-shifting and head-turning conditions respectively. We can also verify that subjects had no head movements in the eye-shifting condition according to Fig. 4.4 (a, b, c), and the distance between GIP and target cubes reaches the minimum at GIP onset for both conditions.

In Fig. 4.4 (a, d), we observe that subjects initiated their movements around 100 ms after stimulus onset for both conditions. Fig. 4.4 (d) further shows that subjects first use their eyes to locate the target cube, then turn their head to face toward the target cube. In Fig. 4.4 (e, f), we see the angle of eye movement starts to decrease while the angle of head movement increases after GIP onset. On the other hand, both angles of the eye and head movement stop changing after fixation onset. We can visualize this phenomenon more obviously in Fig. 4.5 (e, f). The sign of angular velocity of the pupil vector and head-turning direction are opposite after GIP onset in Fig. 4.5 (e), and the angular velocity drops to nearly zero after fixation on set in Fig. 4.5 (f).

4.4.2 Event onset and latency

Fig. 4.6 shows the probability distribution of GIP onset, fixation onset, and response onset. The results in eye-shifting and head-turning conditions are shown in Fig. 4.6 (a) and (b) respectively. Since there is no significant difference between standard and deviant trials, we only present the results from deviant trials for clarity. All three event onsets happen significantly later in the head-turning condition compared to the eye-shifting condition.

To investigate how gaze affects response time, we compare the probability distribution of the latency between events in Fig. 4.7. The results of eye-shifting and head-turning conditions are shown in the left and right columns respectively. The first row (a, d) shows the latency between fixation onset and GIP onset. The second row (b, e) shows the latency between response onset and GIP onset. The third row (c, f) shows the latency between response onset and fixation onset.

In the eye-shifting condition, the mean of the latency between fixation and GIP onset is -15 ms as shown in Fig. 4.7 (c) and both events have a followed response onset of around 300 ms. In the head-turning condition, the mean of the latency between fixation and GIP onset

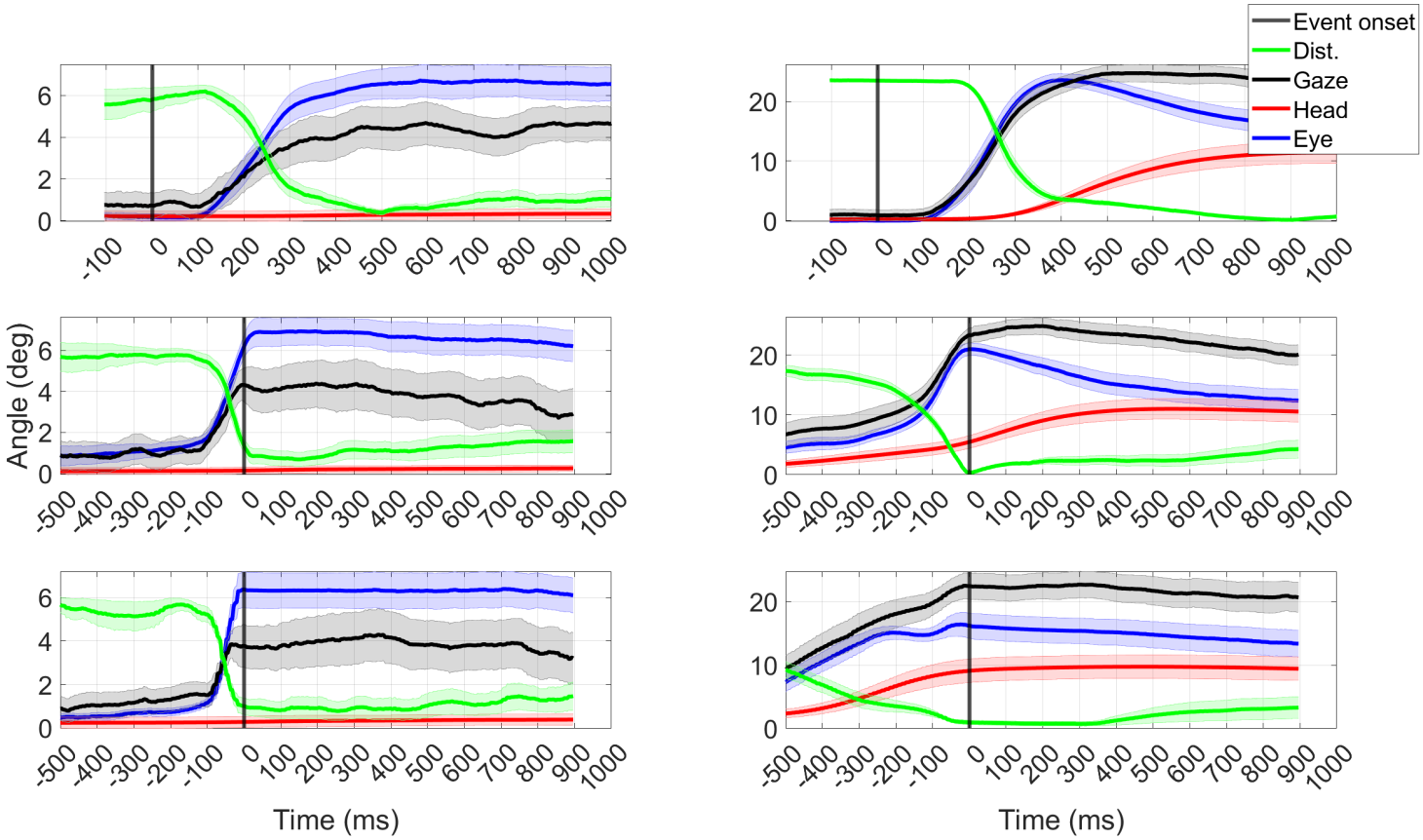

Figure 4.4. The angular dispersion of behaviors with different time locking. The results under eye-shifting and head-turning conditions are shown in Fig. 4.4 (a, b, c) and (d, e, f) respectively. The first row (a, d) shows the behaviors time locking to stimulus onset. The second row (b, e) shows the behaviors time locking to GIP onset. The third row (c, f) shows the behaviors time locking to fixation onset. The eye movement, head movement, and gaze are labeled as Eye (blue), Head (Red), and Gaze (Black) respectively. The distance between GIP and the target cube is labeled as Dist. (Green) and scaled according to eye movement to fit in the figures. The solid lines indicate the mean and the shaded area indicates the standard error of the mean across subjects.

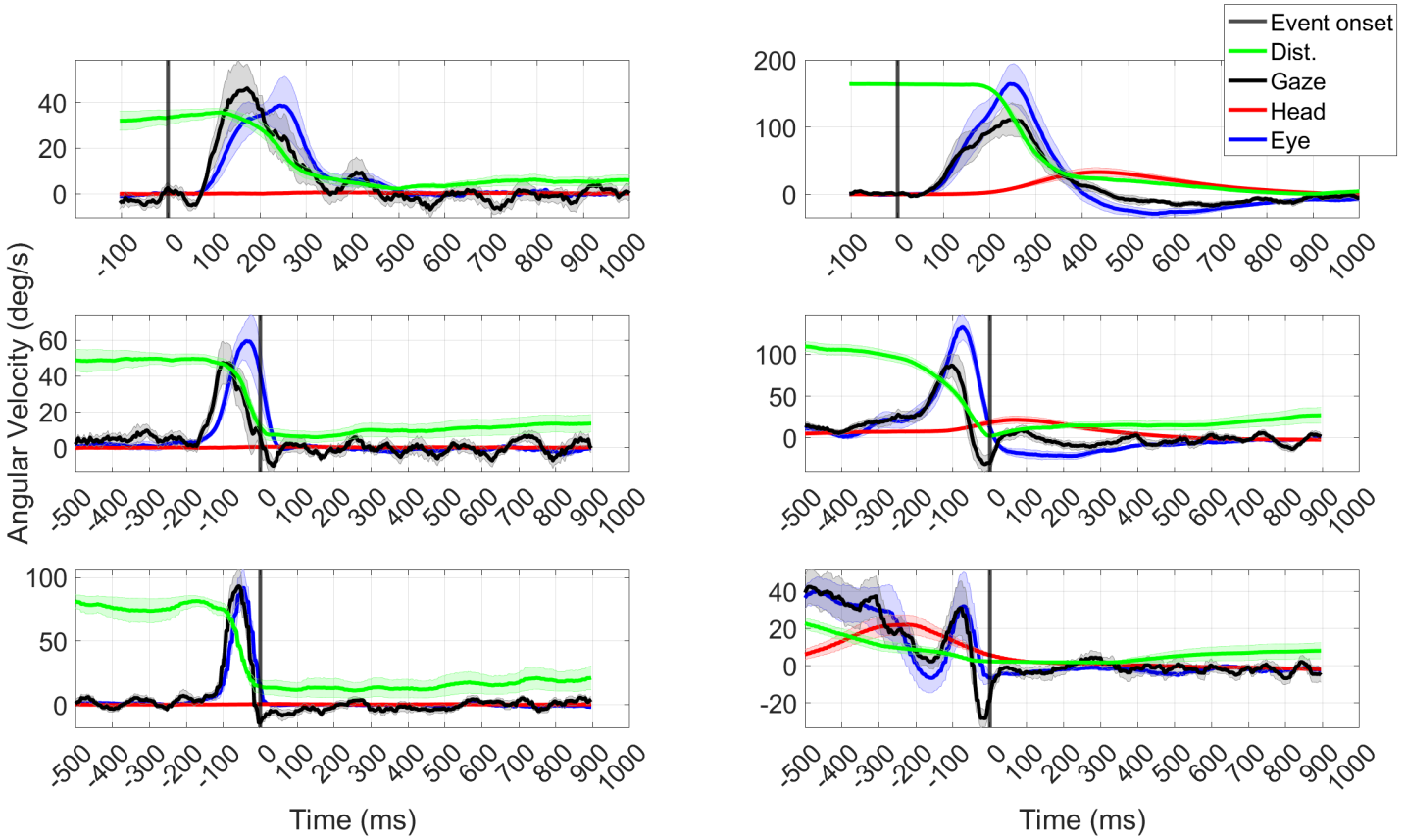

Figure 4.5. The angular dispersion of behaviors with different time locking. The results under eye-shifting and head-turning conditions are shown in Fig. 4.5 (a, b, c) and (d, e, f) respectively. The first row (Fig. 4.5 (a, d)) shows the behaviors time-locked to stimulus onset. The second row (Fig. 4.5 (b, e)) shows the behaviors time-locked to GIP onset. The third row (Fig. 4.5 (c, f)) shows the behaviors time-locked to fixation onset. The eye movement, head movement, and gaze are labeled as Eye (blue), Head (Red), and Gaze (Black) respectively. The distance between GIP and the target cube is labeled as Dist. (Green) and scaled according to eye movement to fit in the figures. The solid lines indicate the mean and the shaded area indicates the standard error of the mean across subjects.

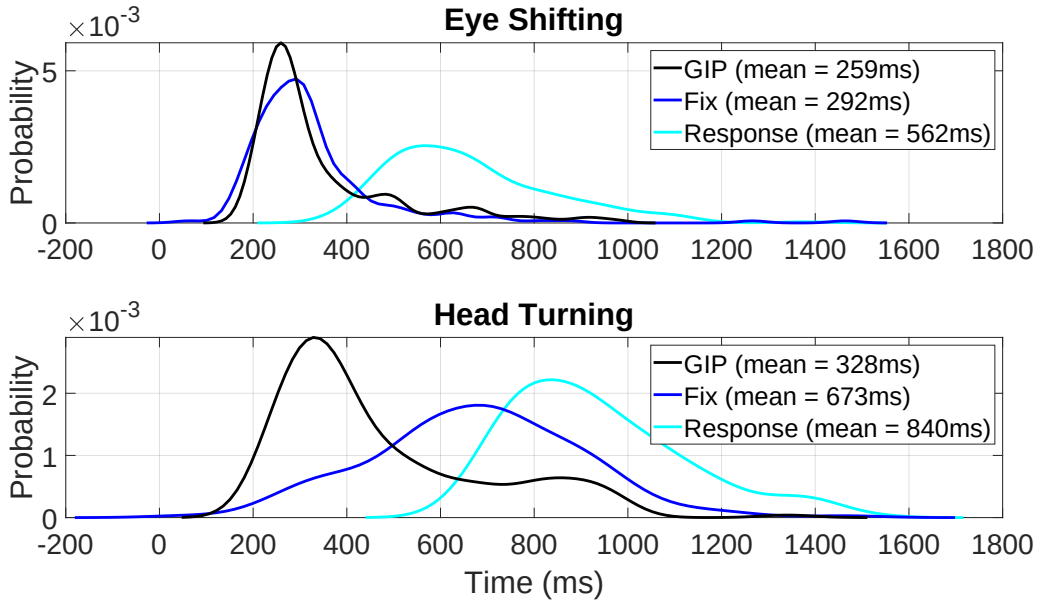

Figure 4.6. The probability distribution of GIP onset (black), fixation onset (blue), and response onset (cyan) in eye-shifting and head-turning conditions.

increases to 200 ms. The latency between response and GIP onset significantly increases to 464 ms, while the latency between response and fixation onset significantly decreases to 264 ms in the two-sample t-test.

4.4.3 Event-related potential with different time-locks

Fig. 4.8 and Fig. 4.9 show the event-related potential (ERP) time locking to stimulus onset, GIP onset, and fixation onset on Cz and O2 channels respectively. The results of eye-shifting and head-turning conditions are shown in the left and right columns respectively. The first row (a, d) shows the ERP time locking to stimulus onset. The second row (b, e) shows the ERP time locking to GIP onset. The third row (c, f) shows the ERP time locking to fixation onset. All the statistical significance described in this section is calculated using a two-tails paired-wise t-test with alpha equal to 0.05.

Fig. 4.8 (a) shows there is a stimulus-related potential between 100 ms and 300 ms for both standard and deviant trials in eye-shifting conditions. The amplitude of deviant trials is significantly lower around 400 ms and significantly higher around 800 ms after event onset

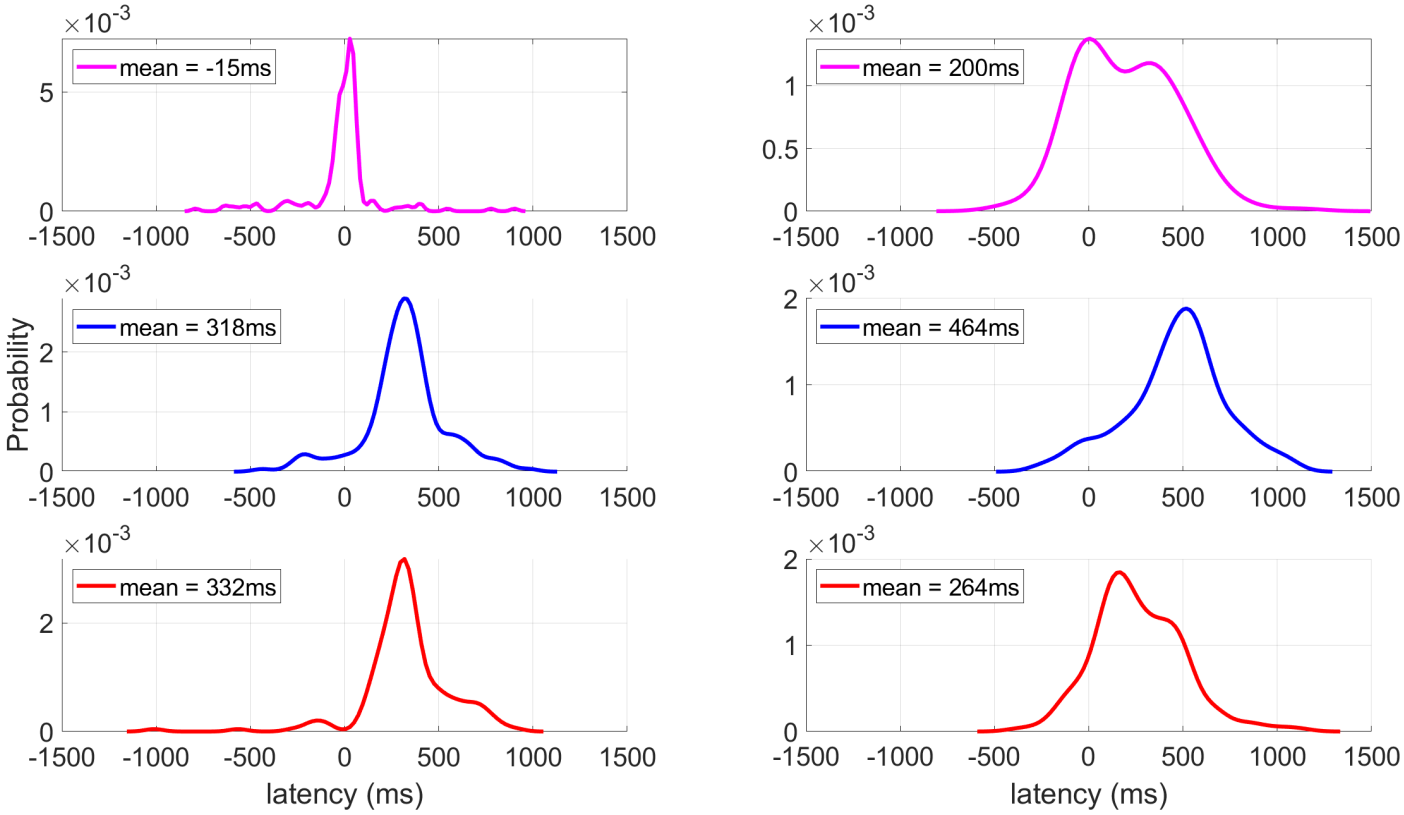

Figure 4.7. The probability distribution of event latency. The results in eye-shifting and head-turning conditions are shown in Fig. 4.7 (a, b, c) and (d, e, f) respectively. The first row (a, d) shows the latency between fixation and GIP onset. The second row (b, e) shows the latency between response and GIP onset. The third row (c, f) shows the latency between response and fixation onset.

compared to the standard trials. In Fig. 4.8 (b) and (c), the amplitude of deviant trials is significantly lower 200 ms before the event onset and significantly higher around 500 ms after event onset when time-locked to GIP and fixation. Out of our expectation, though the amplitude of deviant trials seems to be higher around 300 ms after event onset in Fig. 4.8 (b) and (c), the difference between standard and deviant trials does not reach statistical significance. In the head-turning condition, there is no clear stimulus-related nor GIP-related potential in Fig. 4.8 (d) and (e). However, Fig. 4.8 (f) shows the amplitude of deviant trials is significantly higher around 380 ms after fixation onset.

We also investigate if we can capture fixation-related potential (FRP), a more well-studied ERP, on O2 channel. Fig. 4.9 (a, b, c) shows there is clear stimulus-related, GIP-related, and fixation-related potential in the eye-shifting condition. The FRP shown in Fig. 4.9 (b) is similar to those found in previous studies which a negative drop followed by a positive peak around 100 ms after fixation onset [109]. However, the peak of FRP shown in Fig. 4.9 (c) is not as sharp as in (b). Moreover, instead of having a single peak, there seems to be a consecutive peak of around 200 ms. In head-turning condition, we can still observe the initial stimulus-related potential between 100 ms and 200 ms in Fig. 4.9 (d), and GIP-related potential in Fig. 4.9 (e). However, the signals are much noisier and no FRP is found in Fig. 4.9 (f).

One important thing is that after applying Holm-Bonferroni correction [110] for multiple comparisons, there is no significant difference between standard and deviant trials at the Cz and O2 channels in all conditions.

4.5 Discussion

Eye movements are an important indicator of human attention and interest and have been widely used in the fields of behavioral research, marketing, and neuroscience. One of the most informative features of eye movement is fixations, which have previously been shown to provide precise timing references for neuroscience research under lab conditions where the subjects

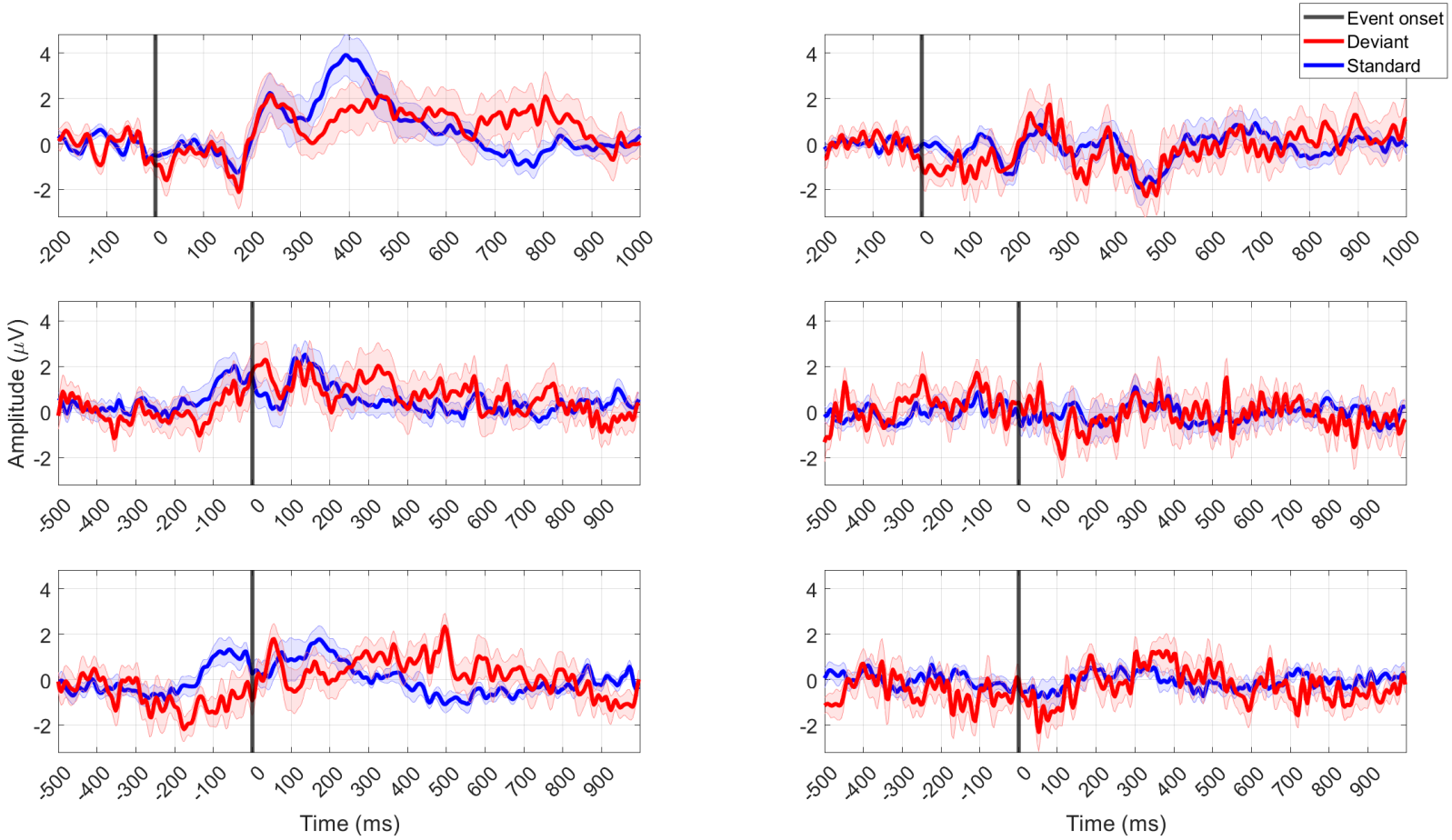

Figure 4.8. The event-related potential with different time-locks at the Cz channel. Fig. 4.8 shows the ERPs under eye-shifting and head-turning conditions in Fig. 4.8 (a, b, c) and (d, e, f) respectively. The first row (a, d) shows the ERP time-locks to stimulus onset. The second row (b, e) shows the ERP time-locks to GIP onset. The third row (c, f) shows the EPR time-locks to fixation onset. The standard trial ERP and deviant trial ERP are shown in blue and red respectively. The solid lines indicate the mean of ERP and the shaded area indicates the standard error of the mean across subjects.

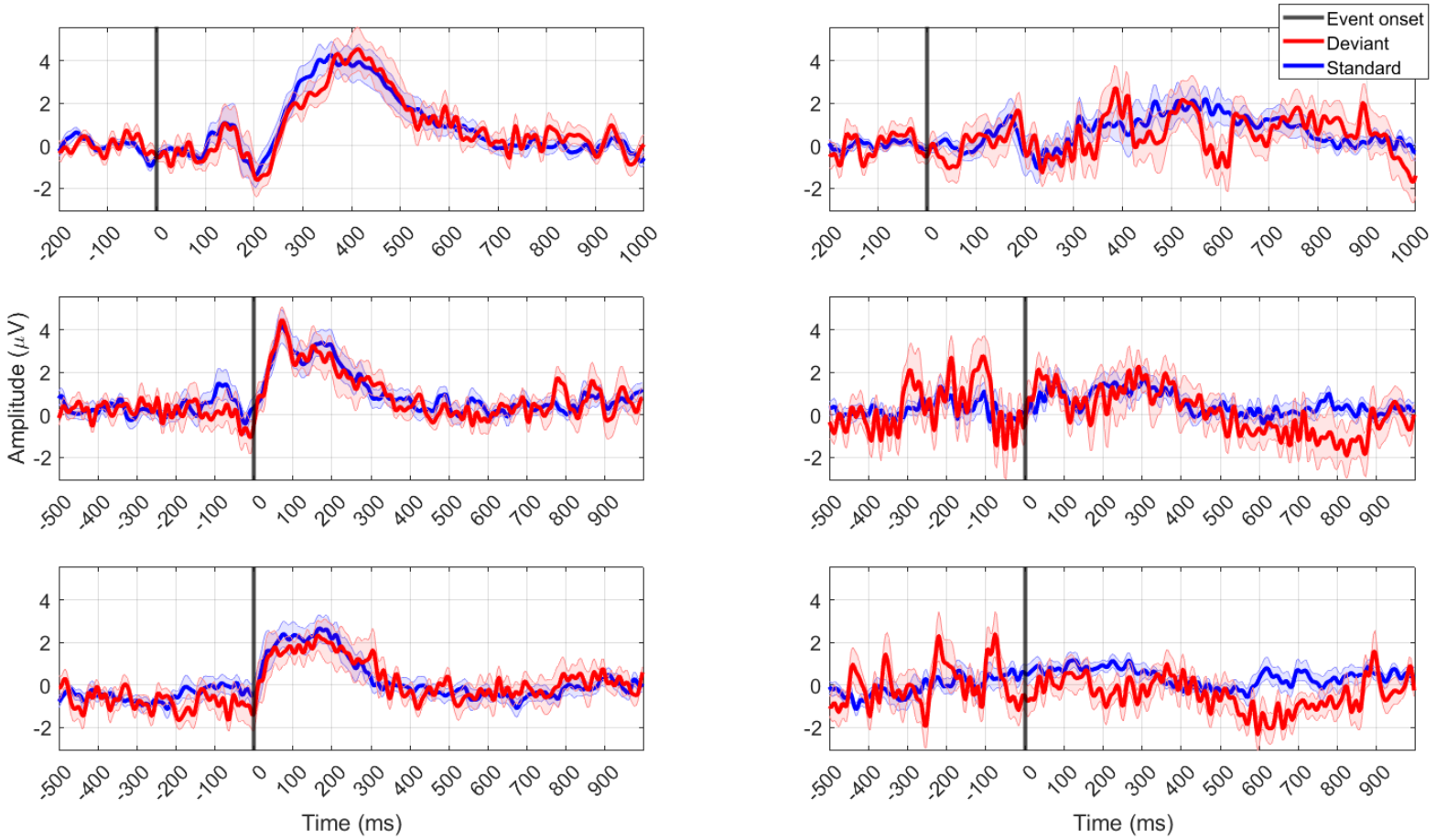

Figure 4.9. The event-related potential with different time-locks on O2 channel. Fig. 4.8 shows the ERPs under eye-shifting and head-turning conditions in Fig. 4.8 (a, b, c) and (d, e, f) respectively. The first row (a, d) shows the ERP time-locks to stimulus onset. The second row (b, e) shows the ERP time-locks to GIP onset. The third row (c, f) shows the EPR time-locks to fixation onset. The standard trial ERP and deviant trial ERP are shown in blue and red respectively. The solid lines indicate the mean of ERP and the shaded area indicates the standard error of the mean across subjects.

have no head movement. However, when it comes to more realistic conditions involving natural, unconstrained head and body movement, methods for parsing saccades and fixations and using these segmentations to analyze other modalities of simultaneously recorded physiological data are still in their infancy. This study has two aims. First, characterize the interaction between head and eye movement when participants directed their gaze to the far peripheral targets. Secondly, compare the effectiveness of FRP versus GIP time-locking approaches in detecting visual P300 oddball effects under conditions of head movement.

To characterize the interaction between head and eye movement, we first have to verify our eye-tracking system and algorithm are reliable. We observed three pieces of evidence to support the reliability in Fig. 4.4 and Fig. 4.5. First, the minimum distance between the target cube and GIP happens at GIP onset. Second, the subjects' gaze rested around the desired visual angles after GIP onset. Third, no head movement is found in the eye-shifting condition and no eye movement is found after fixation onset. Subjects' gaze does not rest exactly at the desired angle because they stop moving their gaze once they reach the edge of the target cubes.

In the eye-shifting condition, there is an offset between the angle of the eye and gaze as shown in Fig. 4.4 (a, b, c). Theoretically speaking, these two values should be the same. There are two possible reasons: (1) the center of the eye is closer to the center fixation point than the center of the head, causing the eye angle to be larger than the head angle. (2) the eye tracking system is less reliable when the pupil approaches the edge of the eyeball, causing the estimation of the eye angle to be larger than the actual value. In the head-turning condition, we see the eye angle starts decreasing and the head angle keeps increasing as shown in Fig. 4.4 (d, e, f). At the same time, the gaze angle decreases at a smaller rate compared to the eye angle. This phenomenon indicates subjects fixated their gaze while adjusting their head direction to point toward the target cube. We hypothesize that compensatory eye movement in conjunction with head movement is highly related to vestibular-ocular reflex (VOR), which allows an individual to maintain their gaze on a fixed object under conditions of head movement. In Fig. 4.5 (e) and (f), we observe that the gaze angular velocity has a negative value right before GIP and fixation

onset, indicating the gaze has a sharp break with an overshoot when the gaze approaches near the target cube. In both conditions, we observe subjects initiating their movements around 100 ms after stimulus onset as shown in Fig. 4.4 (a, d), indicating that human reaction time to process visual information from peripheral vision is independent of the visual angle.

One thing out of expectation is there is no behavioral difference between standard and deviant trials as mentioned in Section 4.4.1 and 4.4.2. To verify if peripheral vision can detect oddball events before fixation onset, we investigated the event onset time and the latency between different events. In Fig. 4.6, we see that all event onsets in the head-turning condition happen later than in the eye-shifting condition with a wilder variance. The variance of onset time might be due to the control ability of the eyeball and neck toward different directions as different muscle groups are involved. In Fig. 4.7, we see that the latency between fixation and GIP onset is nearly zero. The response latency corresponding to fixation and GIP onset is almost the same in the eye-shifting condition. The reason why the mean of the latency between fixation and GIP onset is -15 ms might be due to the outliers since the median of the latency is 15 ms. When comparing eye-shifting and head-turning conditions, we found the latency between response and GIP is higher in the head-turning condition while the latency between response and fixation is lower compared to in the eye-shifting condition. The increased latency between response and GIP indicates that the ability to process visual information decreases when the head turns even if the gaze is fixating on the target cube. However, this result does not mean that there is no information being processed as the latency between response and fixation decreases, indicating that it takes a shorter time to justify if there is an oddball event after fixation.

Lastly, we compare the ERP time locking to stimulus, GIP, and fixation onset on Cz and O2 channels to observe the oddball response and FRP. First, we investigated whether we can observe the P300 in an oddball task on Cz channel in Fig. 4.8. We found there is a stimulus-related potential between 100 ms and 300 ms on Cz in the eye-shifting condition but not in the head-turning condition as shown in Fig. 4.8. This result indicates the stimulus-related potential might relate to controlling eye movements. In the eye-shifting condition, instead of the P300

reported in previous studies, we found the amplitude of deviant trials is significantly lower around 400 ms but significantly higher around 800 ms in stimulus-locked ERP. Also, in Fig. 4.8 (b) and (c), we found this negative amplitude appears right before the GIP and fixation onset. Combining these observations and the results in Fig. 4.6, we hypothesize that negative GIP-related and fixation-related potentials are related to peripheral visual information processing of oddball events. Though there seems to be a positive amplitude of deviant trials as shown in Fig. 4.6 (b) and (c), the difference does not pass the statistical test. However, we still found the amplitude of deviant trials is significantly higher at 500 ms which is caused by the response-related potential. In the head-turning condition, the ERP is noisier and the only significant difference we observed is the significantly higher amplitude of deviant trials around 380 ms in the fixation-locked potential caused by the response-related potential.

Next, we investigated whether we can observe FRPs on the O2 channel in Fig. 4.9. In the stimulus-locked potential, we observe a positive peak at 150 ms and a negative peak at 200 ms in both conditions, indicating this ERP is related to the initiation of eye movements. In the eye-shifting condition, we can observe FRPs at 80 ms followed by a consecutive ERP at 150 ms in GIP-locked and fixation-locked potentials. The FRP in GIP-locked potential is sharper and more similar to the FRP reported in previous studies. On the other hand, the FRP in fixation-locked potential seems to be covered by the consecutive ERP. One possible origin of the consecutive ERP might be from the sudden change of brightness in the focal region when the subjects' gaze is moving from a black background to green target cubes. In the head-turning condition, the signals are noisier because the O2 channel is close to the neck muscle. However, we can still observe FRP in GIP-locked potential with a drop of peak amplitude from 4 μV to 2 μV. There is no FRP found in fixation-locked potential. Based on these observations, we hypothesize the FRPs reported in previous studies might be generated when the visual stimulus moves into the focal region instead of the stop of eye movements. Further investigation and new experiments are needed in the future.

4.6 Conclusion

This study characterizes the interaction between head and eye movement during gaze relocating and investigates the effect of different time locking on ERP. Our results indicate that the reaction time for humans to process visual information from peripheral vision is independent of the visual angle, and the visual information can be processed before the visual stimulus moves into the focal region. Moreover, this study provides strong evidence of vestibular-ocular reflex under the head-turning condition. Though there is no significant difference related to P300 in the eye-shifting condition, the amplitude of deviant trials seems to be higher than the standard trials in GIP-locked and fixation-locked potential. On the other hand, the P300 in head-turning condition has been observed in fixation-locked potential. Our results also indicate the FRP reported in previous studies might be generated when the visual stimulus moves into the focal region instead of the stop of eye movements. However, new experiments and further investigation are required to justify this claim. This study demonstrates the relationship between the eye and head movements and the importance of choosing time locking in ERP analysis under natural moving conditions, providing a solid foundation for future studies in a real-world scenario.

4.7 Acknowledgement

Chapter 4, in full, is being prepared for submission for publication as Chi-Yuan Chang, Weichen Lu, Cory Stevenson, Russell Cohen Hoffing, Steven Thurman, Tzyy-Ping Jung and Ying Choon Wu, "Free Moving Gaze-related Electroencephalograms Electroencephalography in Mobile Virtual Environments". The dissertation author was the first investigator of this paper. This research was supported by grant #1734883 from the National Science Foundation and #W911NF2120126 and #W911NF2020088 from the Army Research Laboratory. The authors gratefully acknowledge the technical support on data recordings from Nicole Wells and Xin Wen.

Chapter 5

Deployment of mental stress detection

5.1 Introduction

In the previous chapters, we evaluated the performance of Artifact Subspace Reconstruction (ASR) and demonstrated its ability to signal reconstruction in the VR environment. In this chapter, we implemented the artifact removal methods in our stress detection algorithms and investigated two important issues of deployment: the ability of real-time detection and the number of required recording channels. The algorithms have to be online and capable of detecting stress continuously in daily life. Moreover, the number of recording channels has to be as less as possible to accommodate comfortableness and the manufacturing cost. In this study, we first compared the performance of our stress detection algorithms before and after applying different artifact removal pipelines. Second, we proposed an IC projection method to remove eye-related artifacts without running ICA. Third, we investigated the effect of recording channel reduction on our stress detection algorithms. Finally, we evaluated the online ability of our stress detection algorithms using sliding windows and a majority vote.

5.2 Dataset, feature selection, and model selection

The dataset used in this study is recorded in a classroom experiment in which students take lectures while their EEG and ECG were recorded. We extracted the first 5 mins of eye open resting and collected 92 sessions from all the recordings. Next, we labeled each session into

an increased and normal stress level group based on their DASS-21 questionnaire results. In the end, there are 21 sessions in the increased stress level group and 71 sessions in the normal stress level group. For all the sessions, we calculated their PSD for each channel and extracted the frequency bins 3Hz to 7Hz as features used in stress detection. There are two models used to detect stress levels in this study. First, we implement LDA with equal probability prior. We selected PSD features from Fz, FCz, and Cz and gathered 15 features to train our LDA. Second, we implement a 4-layers fully-connected NN. We train our NN with the PSD features from all the channels which is 150 features.

5.3 Artifact removal pipeline

In chapter 2, we demonstrated that bandpass filtering preprocessing is not able to remove artifacts and thus reduces the performance of our stress detection algorithms. In this chapter, we further preprocessed the data before training our algorithms. The first step after bandpass filtering from 1Hz to 50Hz is to remove the bad channels which contain negligible activity, noisy signals, or have a poor correlation with adjacent channels. Next, we implemented ASR to reconstruct the portion where EEG signals are contaminated by muscle-related or motion-related artifacts. Finally, we implemented ICA and rejected ICs with a probability higher than 0.8 of being artifact-related ICs labeled by `ICLabel`. Example of the rejected eye-related and muscle-related ICs are shown in Fig. 5.1 and Fig. 5.2 respectively. In the following sections, we demonstrated the effect of our artifact removal pipeline on the performance of our stress detection algorithms. Fig. 5.3 shows the entire algorithms of this study.

5.4 Practical issues of deployment

To deploy stress detection in the real world, several problems must be considered. In chapter 3, we demonstrated ASR's powerful ability to remove common artifacts caused by muscle and motion in the real world. However, there are two critical problems that remain. First,

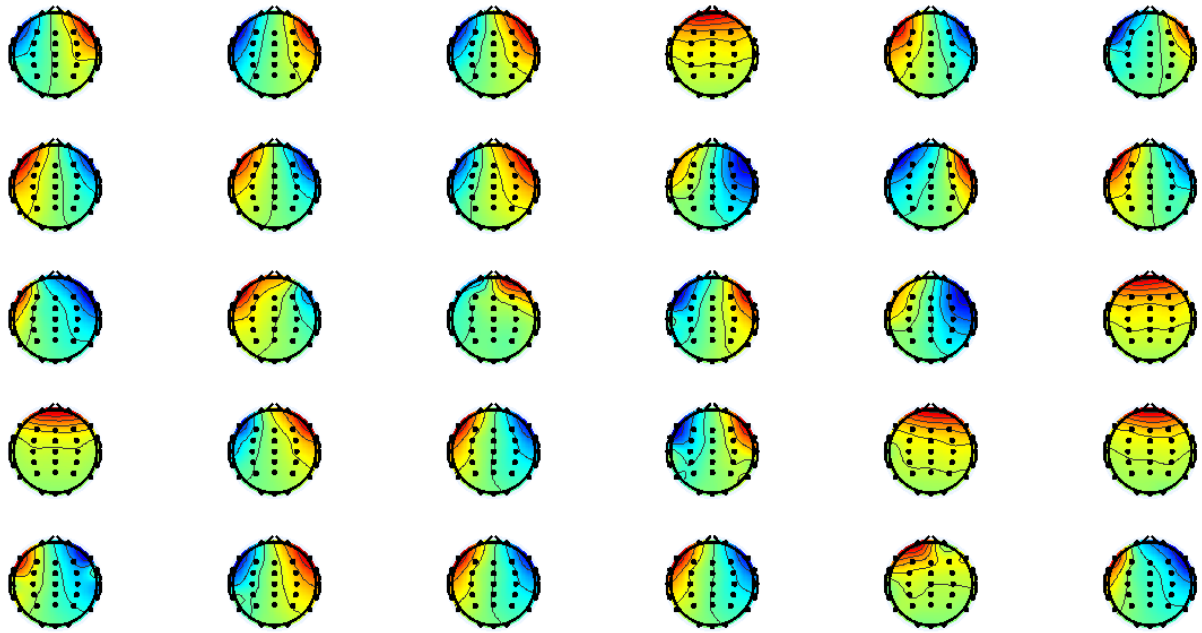

Figure 5.1. Example of the rejected eye-related ICs.

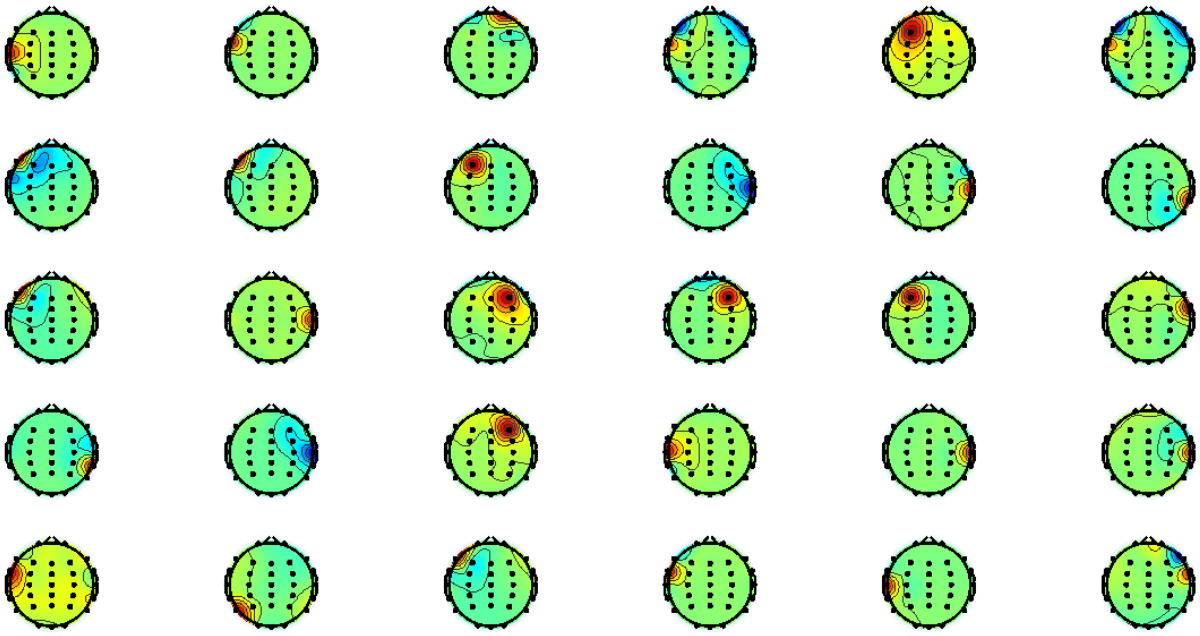

Figure 5.2. Example of the rejected muscle-related ICs.

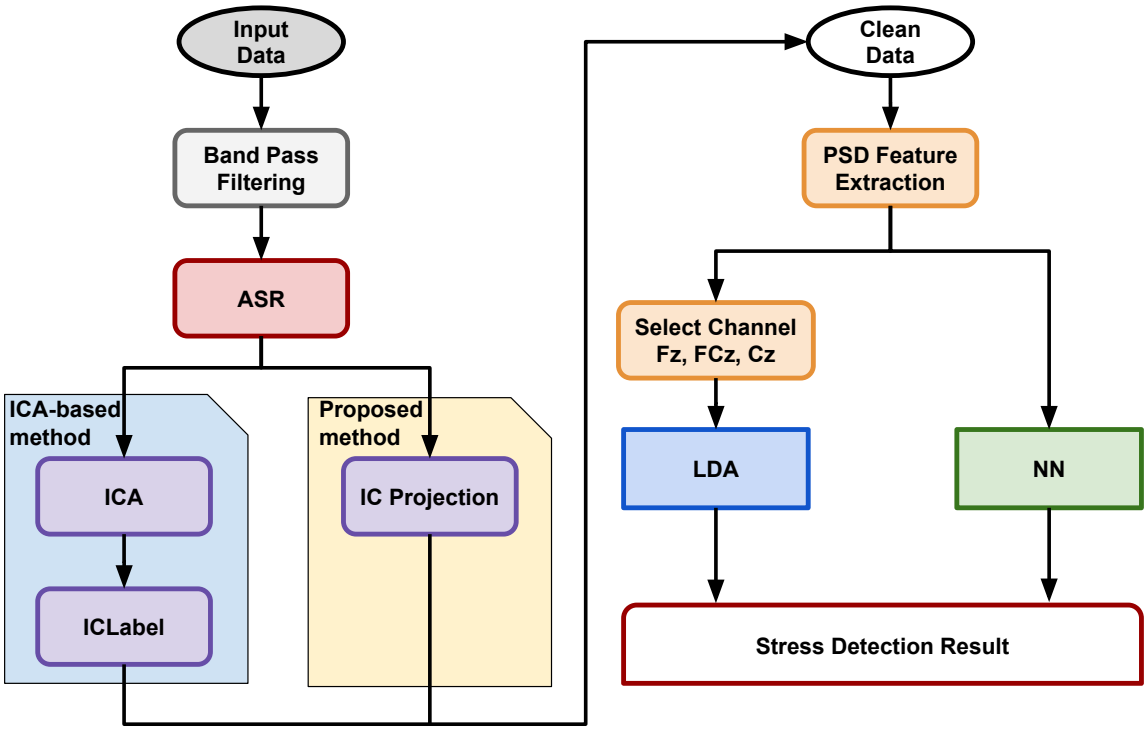

Figure 5.3. The stress detection algorithms of this study.

we would like to monitor our stress levels in real time. Unlike the laboratory scenarios, asking the users to wait 5 minutes before getting the results is unrealistic. Hence, online capability is one of the most important indicators when we evaluated our stress detection. Second, we would like to reduce the number of recording channels needed in our stress detection. EEG setup can be time-consuming and labor-intensive. Also, the manufacturing cost can significantly increase when including one additional channel to the recording system. Therefore, we investigated the effect of recording channel reduction on our stress detection algorithms.

5.4.1 Real-time eye-related artifact removal

Independent component analysis can identify artifact-related ICs, especially eye-related ICs, and remove them from the EEG signals [54, 63, 111]. However, the computational time of ICA is too long to utilize in real-time scenarios. In this study, we proposed an IC projection method to remove eye-related ICs without actually running ICA. We only focus on removing

eye-related ICs because of two reasons. First, eye activities contaminate EEG signals in the frequency range from 0Hz to 8Hz which overlaps with the frequency range of our PSD features [112]. Second, compared to the sparse and complicated muscle activities, two kinds of the most severe eye artifacts, blink and horizontal saccade, can be well summarized by ICA into blink-related and saccade-related ICs as shown in Fig. 5.1. To validate whether the majority of sessions contain blink-related and saccade-related ICs, we performed ICA for each session independently and counted the number of eye-related ICs found in each session.

There are three steps in our proposed IC projection method. First, we performed ICA on one single subject and recorded its ICA unmixing matrix W and its ICA mixing matrix W^{-1}. Next, we removed the blink-related IC and saccade-related IC from the mixing matrix W^{-1} by assigning the corresponding columns to 0 and recorded this new mixing matrix M^*. If the number of recording channels between the template and the data being processed is different, we approximated the new mixing and unmixing matrix by preserving the rows in M^* and the columns in W according to the channel labels. Finally, we combined the unmixing matrix and the new mixing matrix together and created an IC decomposition template P. This template can remove the eye-related IC activities using matrix multiplication instead of actually running ICA.

To achieve online capability, we replaced the ICA-based method in the artifact removal pipeline with the proposed method in Section 5.4.2 and Section 5.4.3. We also evaluated whether the proposed method can maintain its performance when using a template recorded by 30 channels to process sessions recorded by 11 channels.

$$y = Wx$$

$$x = W^{-1}Wx$$

$$x_{without\ eye} = M^*Wx$$

$$x_{without\ eye} = Px$$

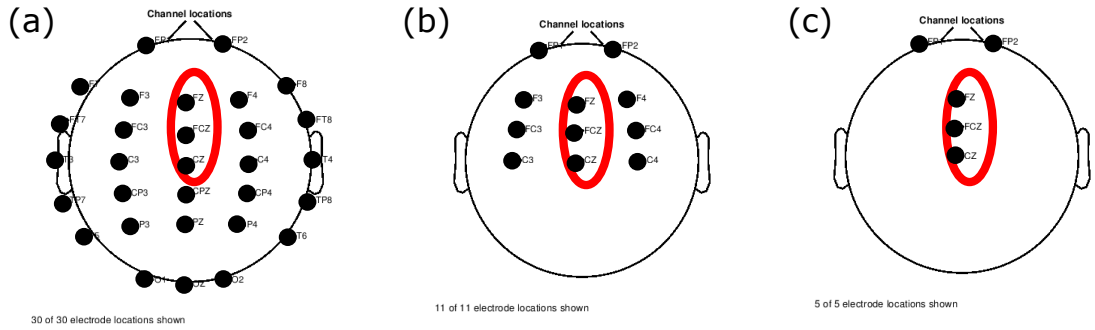

Figure 5.4. Montage of channel reduction. (a) shows all 30 channels. (b) shows the subselected 11 channels. (c) shows the subselected 5 channels. The red circle indicates the location of 3 feature channels.

5.4.2 Reduction of recording channels

To investigate the effect of recording channel reduction, we compared three sets of recording channels and their stress detection performance. Besides the original 30 recording channels, we selected 11 and 5 channels in the frontal region out of the 30 channels. Fig. 5.4 shows the location of the three channels. We selected these channels because we would like to capture the eye-related ICs in the IC decomposition template. After the channel reduction, we run the artifact removal pipeline with the proposed IC projection method and train our stress detection algorithms with LDA and NN. One thing to clarify is that the number of nodes in the input layer of NN changes from 150 to 55 and 25, corresponding to 11 channels and 5 channels, to adapt to the change in the number of recording channels.

5.4.3 Online stress detection

In previous sections, we built an online capable artifact removal pipeline and reduced the required recording channels. In this section, we investigated two important parameters in online use cases: the length of the sliding window for the PSD estimation and the step size to move the sliding window forward. First, we segmented a session into multiple epochs using different sliding window lengths and step sizes. Next, we trained our LDA stress detection model using the rest of the sessions and applied the model to the epochs. For each window, we had a

prediction about whether the session are in the increased stress level group or the normal stress level group. To summarize these predictions, we used a majority vote to assign the session to the group with a larger number of predictions. After we repeated this process for all the sessions, we calculated the balanced accuracy for each pair of sliding window lengths and step sizes.

5.5 Results

5.5.1 Stress detection performance with and without artifact removal and SMOTE

Fig. 5.5 demonstrated the EEG signals before and after our artifact removal pipeline. In Fig. 5.5 (a), we can see there are several eye blinks affecting the channels in the frontal region, and there is a muscle-related artifact caused by jaw clenching. In Fig. 5.5 (b), the eye blinks have been removed by our artifact removal pipeline and the contamination from the jaw clenching has been diminished.

Fig. 5.6 shows the effect of our artifact removal pipeline on the PSD features of selected channels. On the top row, the PSD after bandpass filtering does not have a significant difference between the increase and normal stress level groups in the feature frequency range of all the feature channels. On the other hand, the PSD after artifact removal showing on the bottom row shows significant differences in all three channels between the frequency range of 4Hz to 10Hz. The scale of the PSD also drops from 20 dB to 10dB after artifact removal.

The comparisons of stress detection performance before and after artifact removal are shown in Table 5.1 and Table 5.2 with respect to the LDA model and NN model. In both models, we see the performance gradually increase after each artifact removal step and the best performance appears after applying ASR with offline ICA. Our proposed IC projection method also increases the stress detection performance after applying ASR in both models. Moreover, there is only a 1% difference between the LOO balanced accuracy of the proposed method and the offline ICA in the LDA model. According to the observation in Fig. 5.6, we

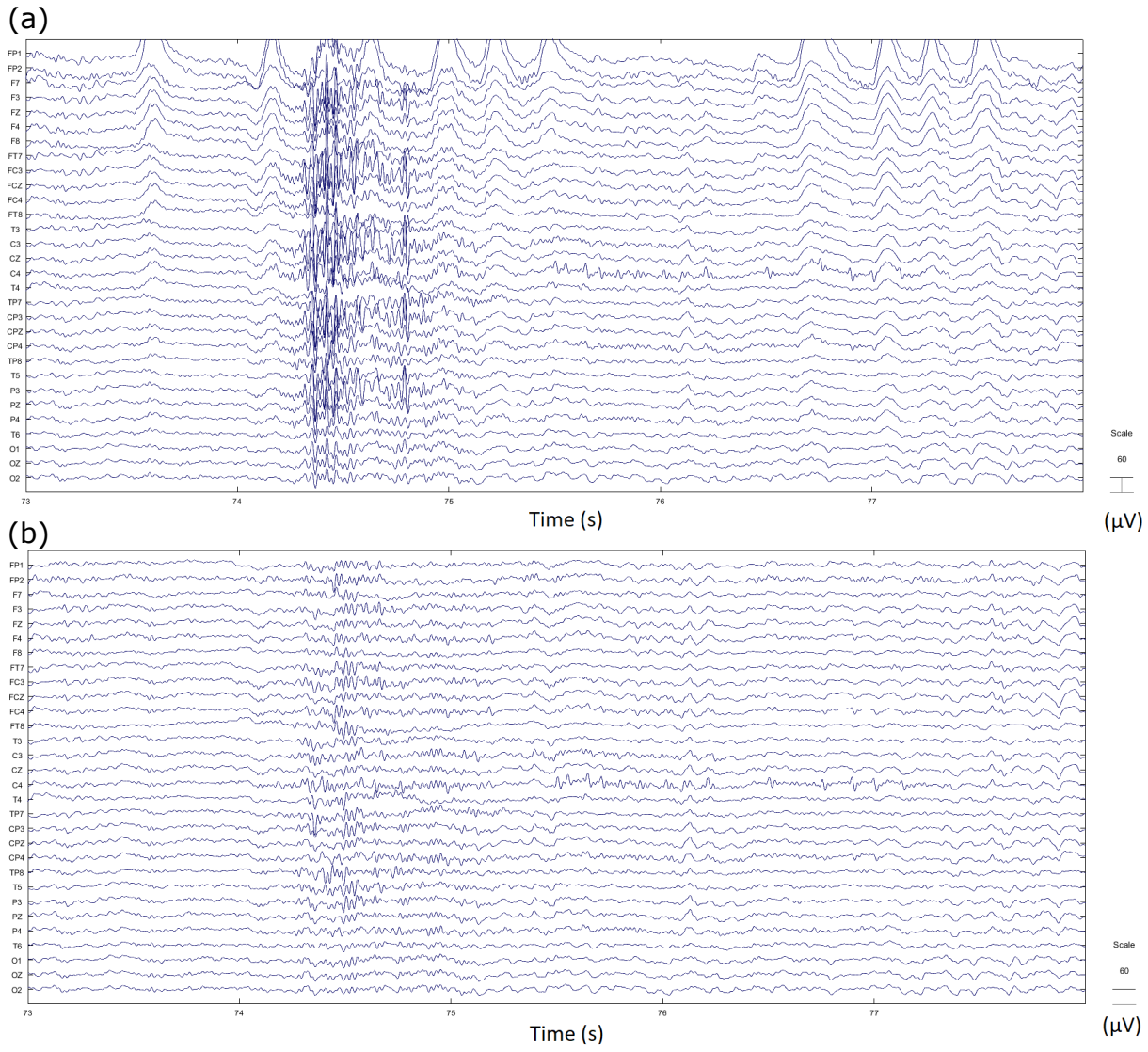

Figure 5.5. Example of the performance of artifact removal. (a) shows the bandpass filtered only data. (b) shows the data after applying our artifact removal pipeline.

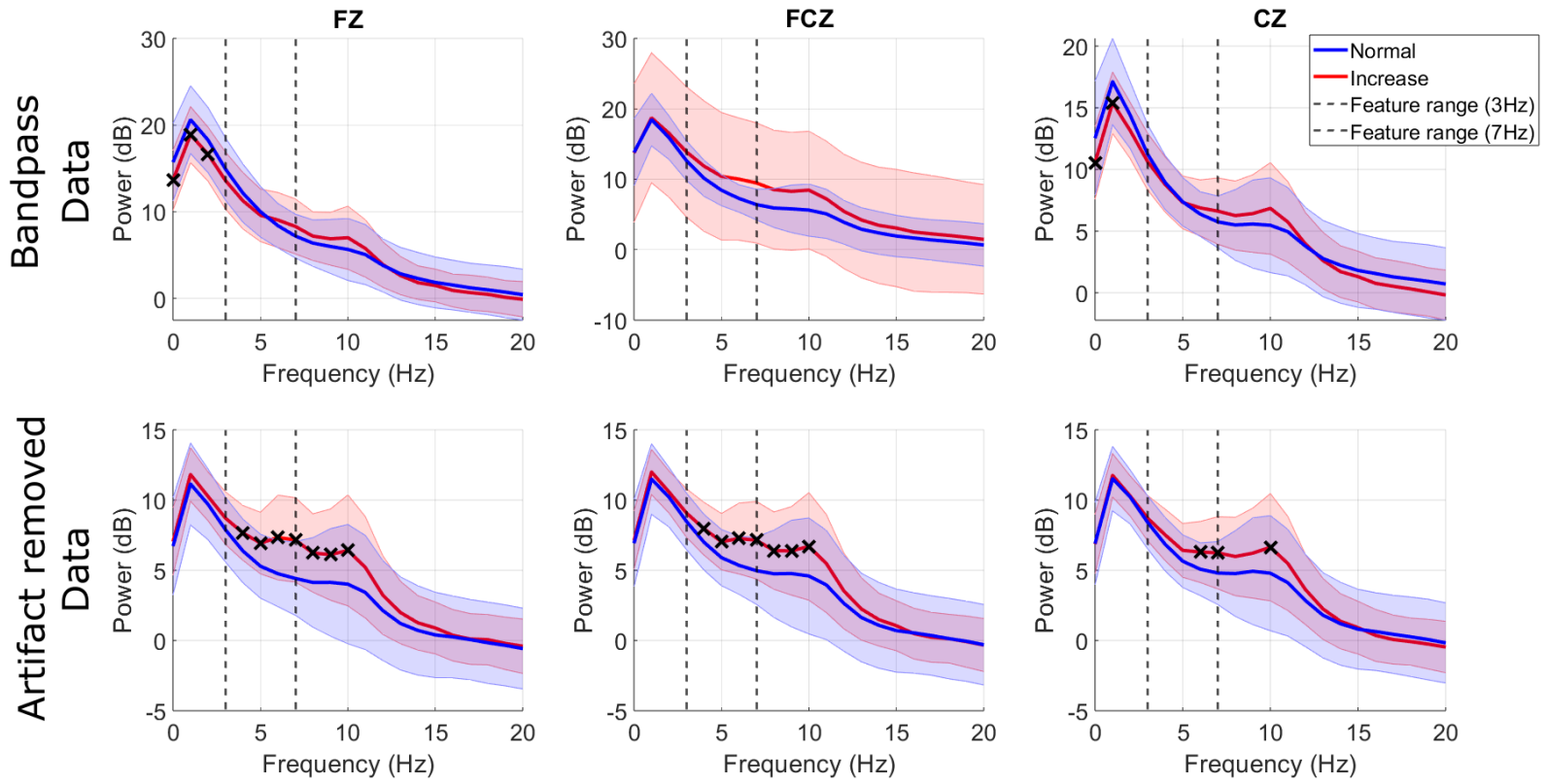

Figure 5.6. PSD features of selected channels before and after artifact removal. The first row shows the PSD of bandpass-only data. The second row shows the PSD of artifact-removed data. The results of the normal and increased stress level groups are shown in blue and red respectively. The solid lines indicate the mean and the shaded area indicates the standard deviation across sessions. The cross dots show the significant difference between the two groups using a bootstrap test ($p < 0.05$). The dashed line indicates the frequency range used in the LDA analysis.

Table 5.1. The effects of artifact removal on LDA stress detection. BP refers to bandpass filtering. Projection refers to our proposed IC projection method. sig. feat. refers to the significantly different frequency features.

LDA	LOO Acc.	LOO Balanced Acc.	F1
BP Only	68.48%	62.81%	0.43
BP + ASR	72.83%	70.66%	0.53
BP + ASR + ICA	**79.35%**	**78.24%**	**0.63**
BP + ASR + Projection	80.43%	77.26%	0.63
BP + ASR + ICA (sig. feat.)	78.26%	72.50%	0.57

Table 5.2. The effects of artifact removal on 4-layers NN stress detection. BP refers to bandpass filtering. Projection refers to our proposed IC projection method. LDA features refer to using the same features as LDA stress detection.

NN	LOO Acc.	LOO Balanced Acc.	F1
BP Only	75.00%	56.98%	0.3
BP + ASR	77.17%	66.77%	0.49
BP + ASR + ICA	**83.70%**	**74.35%**	**0.62**
BP + ASR + Projection	80.43%	68.88%	0.53
BP + ASR + ICA (LDA features)	77.17%	61.74%	0.4

also compared the performance of LDA stress detection with a fixed frequency range (3Hz to 7Hz) and significantly different frequency bins in each channel (shown as a cross in Fig. 5.6) features. The balanced accuracy of using significantly different frequency bin features is 6% worse than using the fixed frequency range features. To compare the two models, we trained our NN stress detection using the LDA features with the architecture described in Fig. 2.2 (b). Although the NN stress detection accuracy is only 2% worse than the LDA stress detection, the balanced accuracy of the NN stress detection is 18% worse than the LDA one.

Table 5.3 shows the effect of SMOTE on the artifact-removed data. Though the LOO balanced accuracy increases in both models with SMOTE, the impact of SMOTE on increased balanced accuracy drops from 5% to 1% compared to the bandpass-filtered-only data. There is an exception when we applied SMOTE on the NN model trained with LDA features. The balanced accuracy boosted up from 61% to 72% which is close to the performance of the NN model trained with features from all recording channels (74%).

Table 5.3. The effects of SMOTE on LDA and NN stress detection. LDA features refer to using the same features as LDA stress detection.

	LOO Acc.	LOO Balanced Acc.	F1
LDA (w/o SMOTE)	79.35%	78.24%	0.63
LDA (w/ SMOTE)	81.52%	79.64%	0.65
NN (w/o SMOTE)	83.70%	74.35%	0.62
NN (w/ SMOTE)	81.52%	76.29%	0.62
NN (LDA features w/o SMOTE)	77.17%	61.74%	0.4
NN (LDA features w/ SMOTE)	78.26%	72.50%	0.57

5.5.2 Consistency of eye-related ICs across different recordings

To evaluate the consistency of eye-related ICs across different recordings, Figure. 5.7 shows the portion of sessions with different number of eye-related ICs found. We found there are 45.7% of sessions contain 2 eye-related ICs and 42.4% of sessions contain 3 eye-related ICs. Moreover, only 2.2% of sessions contain no eye-related ICs. Fig. 5.8 shows an example of the three eye-related ICs found within a session. In this example, we once again observed blink-related IC and saccade-related IC.

Next, we examined whether we can apply the proposed method on a session recorded by few recording channels using a template recorded by a higher amount of channels. Table 5.4 compared the effect of using different templates on the performance of our LDA stress detection. We found there are a 13% accuracy drop and an 18% balanced accuracy drop in the 11-channels LDA stress detection after replacing the 11-channels template with the truncated 30-channels template.

5.5.3 Effect of recording channels reduction on stress detection performance

After we evaluated the performance of our proposed IC projection method, we further compared the effect of recording channel reduction on the data after artifact removal with IC projection. Table 5.5 and Table 5.6 show the results of the LDA model and NN model respectively. All the NN models have been preprocessed by SMOTE. In Table 5.5, we found the performance

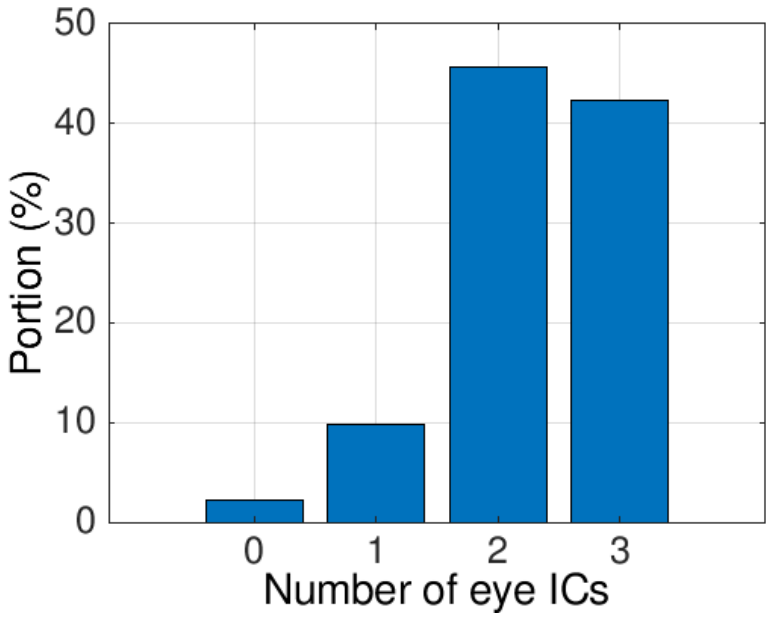

Figure 5.7. Histogram of the number of eye-related ICs found in each session. The histogram has been normalized by the number of sessions.

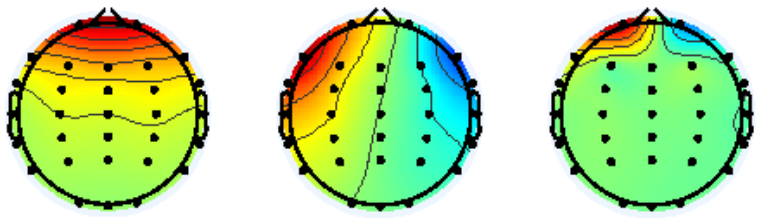

Figure 5.8. Example of the three eye-related ICs found within a session. The left and center ICs are corresponding to blink-related IC and saccade-related IC respectively.

Table 5.4. The effect of different templates on the performance of LDA stress detection. BP refers to bandpass filtering. Proj. refers to our proposed IC projection method followed by the template used.

LDA	Preprocessing	LOO Acc.	LOO Balanced Acc.	F1
	BP Only	68.48%	62.81%	0.43
30 Channels	BP+ASR+ICA	79.35%	78.24%	0.63
	Proj., 30-channels template	80.43%	77.26%	0.63
11 Channels	Proj., 11-channels template	80.43%	78.94%	0.64
	Proj., 30-channels template	63.04%	60.97%	0.41

Table 5.5. The effects of channel reduction on LDA stress detection. The number of channels is first reduced to decided numbers and then the data is preprocessed by the proposed projection method. All the models used frequency bins 3Hz to 7Hz on Fz, FCz, Cz as features.

LDA	LOO Acc.	LOO Balanced Acc.	F1
30 Channels	80.43%	77.26%	0.63
11 Channels	**80.43%**	**78.94%**	**0.64**
5 Channels	67.39%	60.43%	0.40

of LDA stress detection slightly increases 1% after reducing the number of recording channels from 30 to 11. However, once we reduced the recording channels to 5, the balanced accuracy drastically drops from 77% to 60%. Table 5.6 shows the performance of stress detection with the NN model trained with features from all recording channels gradually drops as the number of recording channels decreases. On the other hand, similar to the results of the LDA model, the performance of NN stress detection using LDA features slightly increases 1% when reducing the number of recording channels from 30 to 11, while the balanced accuracy drops from 72% to 54% when reducing the number of recording channels to 5.

Fig. 5.9 shows the example of the template IC decomposition calculated from different numbers of recording channels. Fig. 5.9 (a) and (b) show that the blink-related and saccade-related ICs, highlighted in red squares, can be found in the template IC decompositions of 30 and 11 recording channels. However, there is no obvious eye-related IC in the template IC decomposition of 5 recording channels.

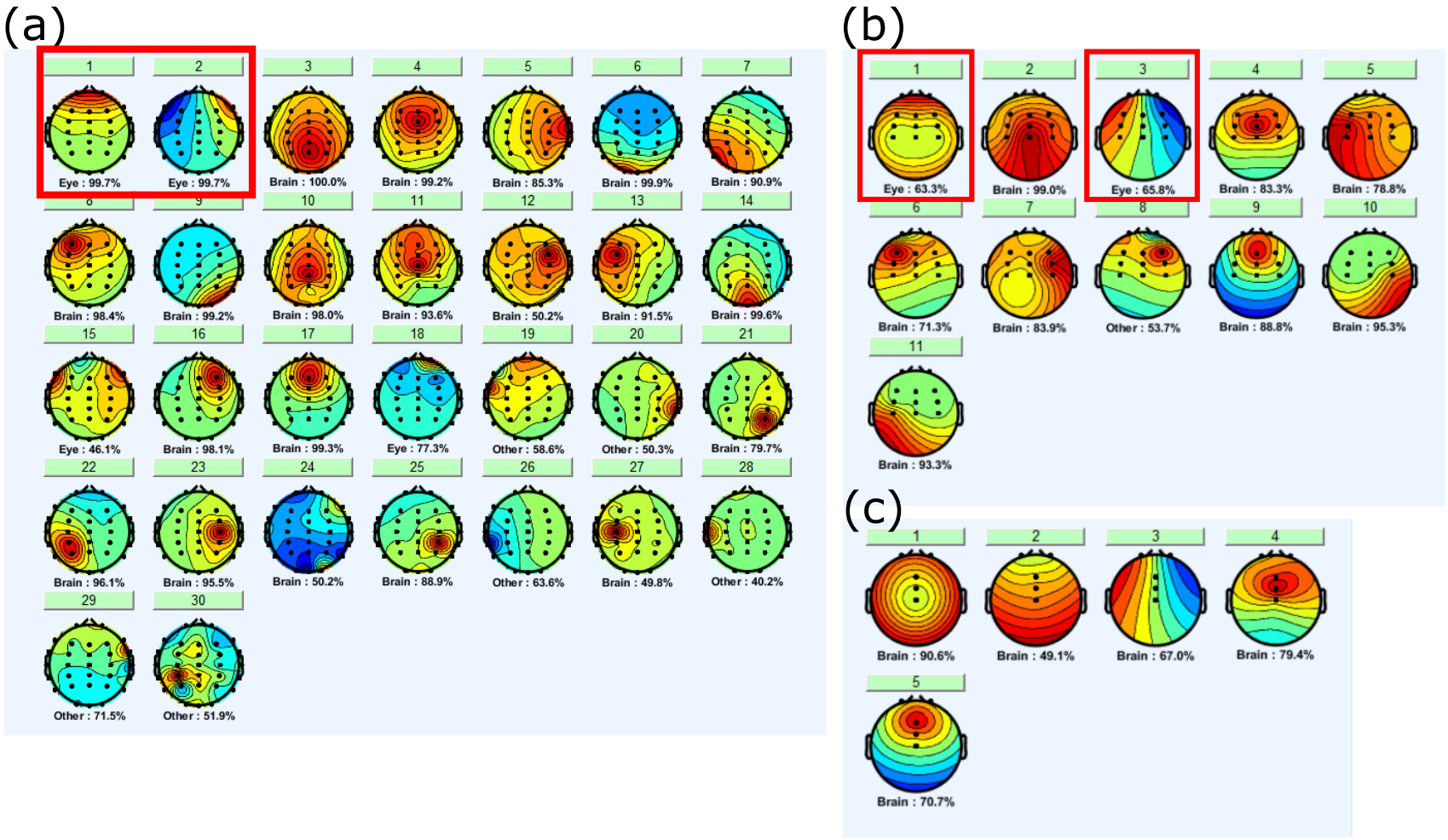

Figure 5.9. Example of the IC decomposition template calculated from different numbers of recording channels. The components are classified by `ICLabel` and eye components for blink and saccade are marked by the red square.

Table 5.6. The effects of channel reduction on NN stress detection. The number of channels is first reduced to decided numbers and then the data is preprocessed by the proposed projection method. The sizes of the input layer in the NN models trained with features from all channels are 150, 55, and 25 corresponding to the 30-channel, 11-channel, and 5-channel models. All the models have been preprocessed by SMOTE.

NN	LOO Acc.	LOO Balanced Acc.	F1
30-channels model	78.26%	67.47%	0.50
11-channels model	75.00%	65.36%	0.47
5-channels model	72.83%	62.27%	0.42
LDA features, 30 channels	78.26%	72.50%	0.57
LDA features, 11 channels	**79.35%**	**73.21%**	**0.58**
LDA features, 5 channels	68.48%	54.43%	0.29

5.5.4 Subject variability

Besides using leave-one-session-out validation to evaluate our model, we also performed leave-one-subject-out validation on the stress detection algorithms using LDA to evaluate if our models are biased to some specific subjects. We summarized the performance of stress detection with the LDA model under different numbers of channels and different preprocessing using leave-one-session-out validation and leave-one-subject-out validation in Table 5.7 and Table 5.8 respectively. The mean operation in Table 5.8 is performed across all subjects.

To exclude the potential biases caused by subjects with only one label, we selected the subjects who have both normal and increased stress level sessions and compared the balanced accuracy before and after artifact removal with 11 recording channels as shown in Fig. 5.10. We found the balanced accuracy of subject 3, 4, and 9 does not change after artifact removal. The balanced accuracy after artifact removal increases for subject 11 and 17 while decreases for subject 16. The mean of the balanced accuracy is 63.18% before artifact removal and 76.37% after artifact removal.

5.5.5 Online stress detection

In this section, we evaluated the online capability of our 11-channels LDA stress detection with the IC projection method. We evaluated this model because it achieves comparable

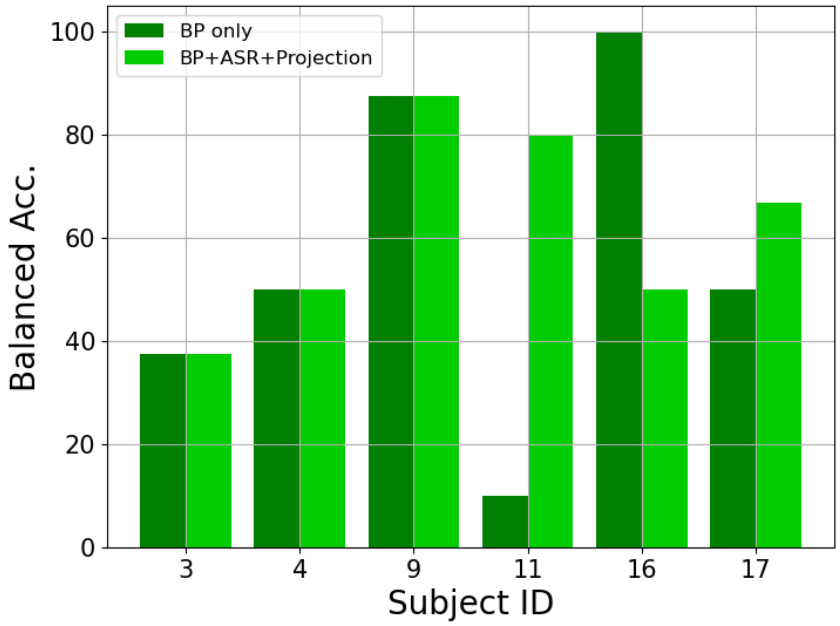

Figure 5.10. Balanced accuracy of each subject before and after artifact removal with 11 recording channels. The figure only includes the subjects who have both normal and increased stress level sessions. The dark green bars indicate the balanced accuracy before artifact removal. The light green bars indicate the balanced accuracy after artifact removal.

performance to the model processed by offline ICA with fewer recording channels. Fig. 5.11 shows the balanced accuracy of 11-channels LDA stress detection with different pairs of sliding window length and step size. When using a window length equal to 1 second, the balanced accuracy of LDA stress detection is around 55%. When the window length increases to 20 seconds, the balanced accuracy reaches the best performance of 78%. When the window length equals 5 mins, the balanced accuracy is 75% and there is a 4% decrease compared to offline results shown in Table 5.5. Out of our expectation, the balanced accuracy drops when the window length is between 40 seconds to 3 mins.

5.6 Discussion

Mental stress can greatly affect one's cognitive performance. Hence, detecting stress in daily life can benefit clinical treatments and task monitoring. However, unlike in the laboratory environment, there are several considerable issues in the real world. In this study, we addressed

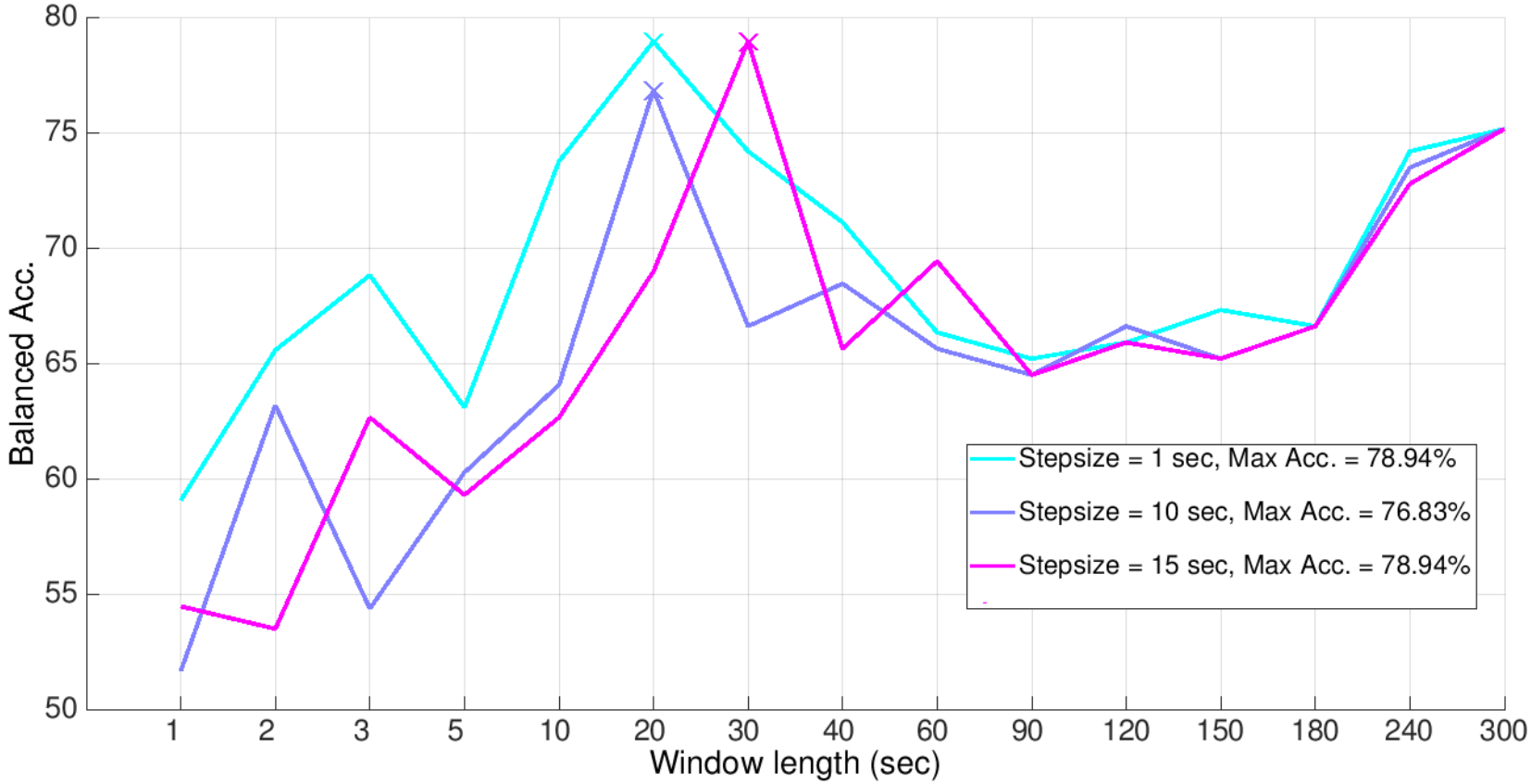

Figure 5.11. Online evaluation with different window and step size. The balanced accuracy is calculated from LOO results and the prediction of each session is made by a majority vote across all the sliding windows. The cyan, blue, and purple lines show the results of step size 1 sec, 10 sec, and 15 sec respectively. The window length with maximum balanced accuracy is marked by a cross in all step sizes.

Table 5.7. Leave-one-session-out validation of LDA stress detection. All the models used frequency bins 3Hz to 7Hz on Fz, FCz, Cz as features. The Proj. refers to the proposed IC projection method.

LDA	Preprocessing	LOO Acc.	LOO Balanced Acc.	F1
	BP Only	68.48%	62.81%	0.43
30 Channels	BP+ASR	72.83%	70.66%	0.53
	BP+ASR+ICA	79.35%	78.24%	0.63
	BP+ASR+Proj.	80.43%	77.26%	0.63
11 Channels	BP+ASR	75.00%	75.42%	0.58
	BP+ASR+Proj.	**80.43%**	**78.94%**	**0.64**
5 Channels	BP+ASR	70.65%	64.22%	0.45
	BP+ASR+Proj.	67.39%	60.43%	0.40

Table 5.8. Leave-one-subject-out validation of LDA stress detection. All the models used frequency bins 3Hz to 7Hz on Fz, FCz, Cz as features. The mean operation is performed across all subjects. The Proj. refers to the proposed IC projection method.

LDA	Preprocessing	Mean Acc.	Mean Balanced Acc.
	BP Only	64.84%	63.18%
30 Channels	BP+ASR	68.17%	66.51%
	BP+ASR+ICA	72.63%	72.93%
	BP+ASR+Proj.	71.17%	71.27%
11 Channels	BP+ASR	71.35%	71.95%
	BP+ASR+Proj.	**77.05%**	**76.37%**
5 Channels	BP+ASR	62.07%	61.19%
	BP+ASR+Proj.	63.84%	63.75%

three important problems of deploying stress detection in the real world: (1) robust EEG recordings using artifact removal, (2) online capability using the proposed IC projection method, and (3) investigation of recording channel reduction.

The results show that our artifact removal pipeline can effectively remove artifacts (Fig. 5.5) and improve the performance of stress detection (Table 5.1, 5.2). Since the artifacts have been removed, we can estimate the brain signals PSD more robustly and capture the significant difference in our PSD features between the increase and normal stress level groups, especially in the range of 3Hz to 7Hz. Moreover, the scale of the PSD after artifact removal aligns with the empirical observations. We used a bootstrap test as our statistical test because the

number of sessions are different between groups and the variances of each group are different, especially for the PSD without artifact removal on FCz as shown in Fig. 5.6.

Each step in our artifact removal pipeline gradually increases the performance of stress detection as shown in Table 5.1 and Table 5.2. Moreover, our proposed IC projection method reaches a comparable balanced accuracy compared to the offline ICA. These results indicate that contamination from eye artifacts is more severe than muscle artifacts in stress detection. We also extract the significantly different frequency bins for each channel as features as shown in Fig. 5.6 and trained the LDA stress detection. However, the balanced accuracy of using these significantly different frequency bin features is worse than using the fixed frequency range features. We hypothesize that though there is a significant difference, the wide variance between 8Hz to 10Hz in both groups make LDA difficult to find a linear combination to minimize the within-group variance and thus reduce the accuracy.

One of the biggest assumption of our proposed IC projection method is the consistency of eye-related ICs across recordings. The empirical result shown in Fig. 5.1 supports the consistency of eye-related ICs as the ICs can be approximately categorized into blink-related IC and saccade-related IC by visual inspection. Moreover, Fig. 5.7 shows that nearly 90% of the sessions in this dataset contain 2 or more eye-related ICs, indicating that blink-related and saccade-related ICs appear in most of the sessions. Besides blink-related and saccade-related ICs, we found some sessions have an additional IC which might be related to small lateral eye movements or frown as shown in Fig. 5.8. On the other hand, the muscle-related ICs are more diverse as shown in Fig. 5.2. Because muscles consist of numerous fibers, it is challenging for ICA to summarize a body movement in a single IC. In addition, the signals generated by muscles also have greater power than the signals generated by brain, resulting extra demands for ICs to explain muscle activities.

There are two problems required further investigations on this method. First, we hypothesize that the IC decomposition template are consistent across datasets and thus can be calculated from any dataset. However, we have not yet validated this hypothesis using other

datasets. Second, we hypothesize the proposed method can be applied on the datasets with number of recording channels less than the template's and larger than 2. Since the proposed method is a component-based artifact removal method, it requires at least 2 components to explain blink and saccade independently. Nevertheless, when we applied 30-channels template on data recorded by 11 channels, the performance of the stress detection drops. One possible reason is the approximation error caused by truncating the mixing and unmixing matrix. The negative influence becomes more and more severe as the difference between the template and the number of recording channels increases as shown in Table 5.4.

Since our dataset is unbalanced, we evaluated whether SMOTE can increase the performance of stress detection. The results show that SMOTE increases stress detection performance with both LDA and NN models, especially for the NN model trained with LDA features. When training the NN model with LDA features without SMOTE, we can observe a high accuracy with low balanced accuracy, indicating the model is strongly biased to the normal stress level group due to the imbalanced classes problem. With SMOTE feature preprocessing, the balanced accuracy of the NN model trained with the LDA features increases to the level of the balanced accuracy of the NN model trained with features from all channels. This result indicates that the selected channels (Fz, FCz, and Cz) contain most of the information related to stress level changes. However, even with the improvement after SMOTE, the balanced accuracy of the NN model is still worse than the LDA model. One of the most possible reasons is the overfitting due to training a model with large trainable parameters using a relatively small number of data.

In the following analysis, we only performed SMOTE on the NN stress detection algorithms because the performance improvement of LDA stress detection algorithms after SMOTE is less significant on the artifact removed data compared to the bandpass filtered only data as shown in Table 2.1 (1% improvement versus 5% improvement). One possible explanation for why the improvement decreases is that the features of the bandpass filtered data in feature space are relatively sparse compared to the artifact removed data features, and thus can benefit more from SMOTE on feature distribution estimation. Because the features of the artifact-removed

data are already condensed, the benefit of SMOTE on distribution estimation becomes less significant.

To reduce the time for EEG setup and manufacturing costs, we would like to include as few recording channels as possible. Since we only selected three feature channels and ICA can capture eye activities with channels near the frontal region, it is not necessary to cover the whole scalp with recording channels. All the models have been preprocessed with the proposed IC projection method. Our results show that in both LDA stress detection and NN stress detection with LDA features can use only 11 recording channels to achieve the performance of using 30 recording channels (Table 5.5, Table 5.6). The results once again strongly support our statement about the selected channels (Fz, FCz, and Cz). However, our IC projection method is a component-based artifact removal method, and its performance drops once the number of recording channels becomes too small. Our empirical results show that when the number of recording channels drops to 5, ICA can not well summarize eye activities into blink-related and saccade-related ICs. Alternatively, ICA obtained a mixture IC of both brain and eye activities as shown in IC2 and IC3 in Fig. 5.9 (c). Furthermore, if we force to use this IC decomposition as our template and reject the most possible eye-related ICs (IC2 and IC3 in this case), there are only three ICs left for us to reconstruct the EEG signals and a huge portion of the brain signals is rejected together with the low-quality eye-related ICs. We can clearly see the performances decline for all the models when using only 5 recording channels (Table 5.5, Table 5.6).

When comparing the performance of NN stress detection with features from all the recording channels and with LDA features, we found the performance with LDA features is better. This result indicates that channel selection captures the information related to stress level changes and prevents NN models from learning brain signals irrelevant to stress level changes. On the other hand, when using 5 recording channels, the performance of NN stress detection with LDA features is worse than with features from all 5 recording channels because the NN stress detection with LDA features further removed the brain-related information from the already damaged signals by selecting 3 channels.

One concern of our stress detection algorithms is, instead of classifying based on the stress level of each session, the algorithms might actually classifying based on the characteristics of each individual subject. To examine this concern, we first compared the leave-one-session-out validation and leave-one-subject-out validation results of our LDA stress detection algorithms. In both validations, the performance of LDA stress detection increases after each step in our artifact removal pipeline when using 11 and 30 recording channels while the performance decreases when using 5 recording channels. This observation slightly reduce the concern. However, to further diminish the concern, we investigate the balanced accuracy of the subjects with both increased stress level and normal stress level sessions in leave-one-subject-out validation as shown in Fig. 5.10. Though half of the subjects do not show increase or decrease in the balanced accuracy, we found there are more subjects having increased balanced accuracy than having decreased balanced accuracy, and the mean of the balanced accuracy across subjects also increases after artifact removal. Moreover, the mean balanced accuracy after artifact removal is comparable to the leave-one-session-out balanced accuracy (76.39% versus 78.94%). These results support that our classifications are based on the stress level rather than subject variability.

Finally, based on the results shown in previous sections, we selected the 11-channel LDA stress detection with the IC projection method and evaluated its online capability by investigating the effect of different pairs of sliding window length and step size on the balanced accuracy. We found the difference in the performance of stress detection with different step sizes is subtle while the choice of window length dominated the performance. If the window length is shorter than 20 seconds, the shorter the window length is, the worse the performance is. Moreover, when the window length is shorter than the step size, the algorithm ignores the data points between the window length and step size, creates a down-sampling effect, and thus decreases the performance of stress detection. On the other hand, when the window length equals 5 mins, we found the balanced accuracy slightly drops compared to the offline results. It may be because some sessions have their eyes open resting longer than 5 mins. We observed the best performances of different step sizes appear when using a window length of around 20 seconds, indicating the quality of

PSD estimation is good enough to detect stress level changes.

One thing that caught our attention is that, instead of a monotonic increase, the balanced accuracy drops before coming back to the expected performance when using a 5-min window length. We hypothesize that mental stress changes do not affect EEG signals constantly but with a repeated cycle. Therefore, the mental stress changes in a window might be diluted by the long window length. However, getting the benchmarks for stress detection in an online scenario is very difficult to test this hypothesis.

5.7 Conclusion

This study proposes a robust, online capable, and less recording channel required mental stress detection algorithm for mental stress monitoring in the real world. We found after our artifact removal pipeline, the PSD is significantly higher in the increased stress level group between frequency bins 4Hz to 10Hz on Fz, FCz, and Cz channels. To remove eye activities in real-time, we propose an IC projection method and reduce the required computational time while maintaining the performance compared to using offline ICA. This study also explores the influences of different models, feature preprocessing, and the number of recording channels. Our empirical results show that with 15 features from Fz, FCz, and Cz channels, the proposed LDA stress detection algorithm can achieve a 77% balanced accuracy in an imbalanced binary stress detection problem. Moreover, only 11 required recording channels are located near the frontal region in the proposed LDA stress detection algorithm. The findings of this study could lead to the deployment of daily mental stress monitoring in the future.

5.8 Acknowledgement

Chapter 5 is an ongoing project at SCCN. The dissertation author was the first investigator of this project. The authors gratefully acknowledge the support from Microsoft Research and the help from Chieh Hsu, Darin Tsui, and Siwan Wang.

Chapter 6

Summary of Contributions

This dissertation aims to build a BCI to monitor mental stress in the real world.

In Chapter 2, we built two stress detection algorithms using LDA and NN with PSD features. We demonstrated how artifacts can contaminate EEG signals and degenerate the performance of stress detection algorithms.

In Chapter 3, we evaluated the performance of ASR, an automated, real-time-capable artifact removal method. Although ASR has existed since 2014, there is no systematic evaluation and the proper choice of the key parameter remains unclear. With our results, people can better understand the mechanism of ASR and have a guideline for choosing the parameters while using it.

In Chapter 4, we further evaluated ASR's brain signal reconstruction ability and explored human behaviors in a visual-tracking experiment conducted in a virtual environment. Though we evaluated ASR in the previous chapter, the ability of brain signal reconstruction is still unclear due to a lack of ground truth. Hence, we introduced a well-studied brain signal using a visual oddball task. Moreover, we investigated the relationship between eye movements and head movements and demonstrated the importance of choosing different time points for analysis in mobile scenarios.

In Chapter 5, we improved the online capability and reduced the deployment difficulty of our stress detection BCI. Monitoring mental stress in daily life required the pipeline of stress

detection algorithms to be online capable. In addition, the recording channels on BCI prefer to be as less as possible due to setup difficulty and manufacturing costs. We first proposed an IC projection method to remove eye activities in real time. Next, we investigated the effect of recording channel reduction on our stress detection algorithms. We found our stress detection algorithm using LDA can have a 77% balanced accuracy in an online scenario with only 11 recording channels placed in the frontal region.

With all of these advancements, recording from BCI can be more robust and the deployment of mental stress monitoring in the real world is more feasible.

Bibliography

[1] Sheldon Cohen, Ronald C Kessler, and Lynn Underwood Gordon. *Measuring stress: A guide for health and social scientists*. Oxford University Press on Demand, 1997.

[2] Giorgos Giannakakis, Dimitris Grigoriadis, Katerina Giannakaki, Olympia Simantiraki, Alexandros Roniotis, and Manolis Tsiknakis. "Review on psychological stress detection using biosignals". In: *IEEE Transactions on Affective Computing* 13.1 (2019), pp. 440–460.

[3] Shashank P Behere, Richa Yadav, and Prakash B Behere. "A comparative study of stress among students of medicine, engineering, and nursing". In: *Indian journal of psychological medicine* 33.2 (2011), pp. 145–148.

[4] Gerhard Andersson and Lucy Yardley. "Time-series Analysis of the Relationship between Dizziness and Stress". In: *Scandinavian Journal of Psychology* 41.1 (2000), pp. 49–54.

[5] JPA Delaney and DA Brodie. "Effects of short-term psychological stress on the time and frequency domains of heart-rate variability". In: *Perceptual and motor skills* 91.2 (2000), pp. 515–524.

[6] Ulf Lundberg, Ingela Elfsberg Dohns, Bo Melin, Leif Sandsjö, Gunnar Palmerud, Roland Kadefors, Maria Ekström, and Deirdre Parr. "Psychophysiological stress responses, muscle tension, and neck and shoulder pain among supermarket cashiers." In: *Journal of occupational health psychology* 4.3 (1999), p. 245.

[7] Suzanne C Segerstrom and Gregory E Miller. "Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry." In: *Psychological bulletin* 130.4 (2004), p. 601.

[8] Anilkumar B Jarali and G Radhakrishnan. "Stress, obesity and selected health problems among professionals". In: *Asian Journal of Nursing Education and Research* 3.3 (2013), p. 3.

[9] Thomas G Pickering. "Mental stress as a causal factor in the development of hypertension and cardiovascular disease". In: *Current hypertension reports* 3.3 (2001), pp. 249–254.

[10] Stuart Checkley. "The neuroendocrinology of depression and chronic stress". In: *British medical bulletin* 52.3 (1996), pp. 597–617.

[11] Gwendolyn C Dieleman, Anja C Huizink, Joke HM Tulen, Elisabeth MWJ Utens, Hanneke E Creemers, Jan van der Ende, and Frank C Verhulst. “Alterations in HPA-axis and autonomic nervous system functioning in childhood anxiety disorders point to a chronic stress hypothesis”. In: *Psychoneuroendocrinology* 51 (2015), pp. 135–150.

[12] Dariusz Kotlega, Monika Gołab-Janowska, Marta Masztalewicz, Sylwester Ciećwież, and Przemysław Nowacki. “The emotional stress and risk of ischemic stroke”. In: *Neurologia i neurochirurgia polska* 50.4 (2016), pp. 265–270.

[13] David A Brent. “Risk factors for adolescent suicide and suicidal behavior: mental and substance abuse disorders, family environmental factors, and life stress”. In: *Suicide and Life-Threatening Behavior* 25 (1995), pp. 52–63.

[14] Jiban Debnath, Nitesh Chauahan, T Prakash, and Divakar Goli. “Stress,” a proxy killer” and role of adaptogen as antistress agent”. In: *Archives of Pharmacy Practice* 2.1 (2011), p. 6.

[15] Hans Selye. “Stress without distress”. In: *Psychopathology of human adaptation*. Springer, 1976, pp. 137–146.

[16] Robert Mearns Yerkes, John D Dodson, et al. “The relation of strength of stimulus to rapidity of habit-formation”. In: (1908).

[17] Mark A Staal. “Stress, cognition, and human performance: A literature review and conceptual framework”. In: (2004).

[18] J Graham Jones and Lew Hardy. “Stress and cognitive functioning in sport”. In: *Journal of Sports Sciences* 7.1 (1989), pp. 41–63.

[19] Nancy Cahir and Robin D Morris. “The psychology student stress questionnaire”. In: *Journal of clinical psychology* 47.3 (1991), pp. 414–417.

[20] Christian S Crandall, Jeanne J Preisler, and Julie Aussprung. “Measuring life event stress in the lives of college students: The Undergraduate Stress Questionnaire (USQ)”. In: *Journal of behavioral medicine* 15.6 (1992), pp. 627–662.

[21] Susan Levenstein, Cosimo Prantera, Vilma Varvo, Maria L Scribano, Eva Berto, Carlo Luzi, and Arnaldo Andreoli. “Development of the Perceived Stress Questionnaire: a new tool for psychosomatic research”. In: *Journal of psychosomatic research* 37.1 (1993), pp. 19–32.

[22] Peter F Lovibond and Sydney H Lovibond. “The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories”. In: *Behaviour research and therapy* 33.3 (1995), pp. 335–343.

[23] Dag Rissén, Bo Melin, Leif Sandsjö, Ingela Dohns, and Ulf Lundberg. "Surface EMG and psychophysiological stress reactions in women during repetitive work". In: *European journal of applied physiology* 83.2 (2000), pp. 215–222.

[24] Nestor L Lopez-Duran, Robin Nusslock, Charles George, and Maria Kovacs. "Frontal EEG asymmetry moderates the effects of stressful life events on internalizing symptoms in children at familial risk for depression". In: *Psychophysiology* 49.4 (2012), pp. 510–521.

[25] Ginette C Blackhart, Jennifer A Minnix, and John P Kline. "Can EEG asymmetry patterns predict future development of anxiety and depression?: A preliminary study". In: *Biological psychology* 72.1 (2006), pp. 46–50.

[26] Delroy L Paulhus, Simine Vazire, et al. "The self-report method". In: *Handbook of research methods in personality psychology* 1.2007 (2007), pp. 224–239.

[27] John Newport Langley. *The autonomic nervous system*. W. Heffer, 1921.

[28] Nandita Sharma and Tom Gedeon. "Objective measures, sensors and computational techniques for stress recognition and classification: A survey". In: *Computer methods and programs in biomedicine* 108.3 (2012), pp. 1287–1301.

[29] Jackson Beatty, Brennis Lucero-Wagoner, et al. "The pupillary system". In: *Handbook of psychophysiology* 2.142-162 (2000).

[30] Shruti Gedam and Sanchita Paul. "A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques". In: *IEEE Access* 9 (2021), pp. 84045–84066. DOI: 10.1109/ACCESS.2021.3085502.

[31] Tyler S Grummett, Sean P Fitzgibbon, Trent W Lewis, Dylan DeLosAngeles, Emma M Whitham, Kenneth J Pope, and John O Willoughby. "Constitutive spectral EEG peaks in the gamma range: suppressed by sleep, reduced by mental activity and resistant to sensory stimulation". In: *Frontiers in human neuroscience* 8 (2014), p. 927.

[32] Fiorenzo Artoni, Chiara Fanciullacci, Federica Bertolucci, Alessandro Panarese, Scott Makeig, Silverstro Micera, and Carmelo Chisari. "Unidirectional brain to muscle connectivity reveals motor cortex control of leg muscles during stereotyped walking". In: *NeuroImage* 159 (2017), pp. 403–416.

[33] Hengameh Marzbani, Hamid Reza Marateb, and Marjan Mansourian. "Neurofeedback: a comprehensive review on system design, methodology and clinical applications". In: *Basic and clinical neuroscience* 7.2 (2016), p. 143.

[34] Georg E. Fabiani, Dennis J. McFarland, Jonathan R. Wolpaw, and Gert Pfurtscheller. "Conversion of EEG activity into cursor movement by a brain-computer interface (BCI)". In: *IEEE Trans. on Neural Syst. and Rehabil. Eng.* 12.3 (2004), pp. 331–338.

[35] MM Sani, H Norhazman, HA Omar, Norliza Zaini, and SA Ghani. "Support vector machine for classification of stress subjects using EEG signals". In: *2014 IEEE Conference on Systems, Process and Control (ICSPC 2014)*. IEEE. 2014, pp. 127–131.

[36] JF Alonso, S Romero, MR Ballester, RM Antonijoan, and MA Mañanas. "Stress assessment based on EEG univariate features and functional connectivity measures". In: *Physiological measurement* 36.7 (2015), p. 1351.

[37] FM Al-Shargie, Tong Boon Tang, Nasreen Badruddin, and Masashi Kiguchi. "Mental stress quantification using EEG signals". In: *International conference for innovation in biomedical engineering and life sciences*. Springer. 2015, pp. 15–19.

[38] Richard S Lewis, Nicole Y Weekes, and Tracy H Wang. "The effect of a naturalistic stressor on frontal EEG asymmetry, stress, and health". In: *Biological psychology* 75.3 (2007), pp. 239–247.

[39] Sanay Muhammad Umar Saeed, Syed Muhammad Anwar, Humaira Khalid, Muhammad Majid, and Ulas Bagci. "EEG based classification of long-term stress using psychological labeling". In: *Sensors* 20.7 (2020), p. 1886.

[40] Katarina Dedovic, Robert Renwick, Najmeh Khalili Mahani, Veronika Engert, Sonia J Lupien, and Jens C Pruessner. "The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain". In: *Journal of Psychiatry and Neuroscience* 30.5 (2005), pp. 319–325.

[41] Likun Xia, Aamir Saeed Malik, and Ahmad Rauf Subhani. "A physiological signal-based method for early mental-stress detection". In: *Biomedical Signal Processing and Control* 46 (2018), pp. 18–32.

[42] Jesus Minguillon, Eduardo Perez, Miguel Angel Lopez-Gordo, Francisco Pelayo, and Maria Jose Sanchez-Carrion. "Portable system for real-time detection of stress level". In: *Sensors* 18.8 (2018), p. 2504.

[43] Anum Asif, Muhammad Majid, and Syed Muhammad Anwar. "Human stress classification using EEG signals in response to music tracks". In: *Computers in biology and medicine* 107 (2019), pp. 182–196.

[44] Li-Wei Ko, Oleksii Komarov, W David Hairston, Tzyy-Ping Jung, and Chin-Teng Lin. "Sustained attention in real classroom settings: An EEG study". In: *Frontiers in human neuroscience* 11 (2017), p. 388.

[45] Irina I Goncharova, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. "EMG contamination of EEG: spectral and topographical characteristics". In: *Clinical neurophysiology* 114.9 (2003), pp. 1580–1593.

[46] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique". In: *Journal of artificial intelligence research* 16 (2002), pp. 321–357.

[47] Suresh Balakrishnama and Aravind Ganapathiraju. "Linear discriminant analysis-a brief tutorial". In: *Institute for Signal and information Processing* 18.1998 (1998), pp. 1–8.

[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[49] Şener Büyüköztürk and Ömay Çokluk-Bökeoğlu. "Discriminant function analysis: Concept and application". In: *Eurasian J Educ Res* 33 (2008), pp. 73–92.

[50] Omneya Attallah. "An effective mental stress state detection and evaluation system using minimum number of frontal brain electrodes". In: *Diagnostics* 10.5 (2020), p. 292.

[51] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[52] Dennis J McFarland, William A Sarnacki, Theresa M Vaughan, and Jonathan R Wolpaw. "Brain-computer interface (BCI) operation: signal and noise during early training sessions". In: *Clinical Neurophysiology* 116.1 (2005), pp. 56–62.

[53] Fabio Babiloni, Andrzej Cichocki, and Shangkai Gao. *Brain-computer interfaces: towards practical implementations and potential applications*. 2007.

[54] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin J. McKeown, Vicente Iragui, and Terrence J. Sejnowski. "Removing electroencephalographic artifacts by blind source separation". In: *Psychophysiology* 37.2 (2000), pp. 163–178.

[55] Jose Antonio Urigüen and Begona Garcia-Zapirain. "EEG artifact removal—state-of-the-art and guidelines". In: *J. of Neural Eng.* 12.3 (2015).

[56] Alain de Cheveigné. "Sparse time artifact removal". In: *Journal of neuroscience methods* 262 (2016), pp. 14–20.

[57] Mainak Jas, Denis Engemann, Yousra Bekhti, Federico Raimondo, and Alexandre Gramfort. "Autoreject: Automated artifact rejection for MEG and EEG data". In: *NeuroImage* 159 (2017), pp. 417–429.

[58] Borna Noureddin, Peter D Lawrence, and Gary E Birch. “Online removal of eye movement and blink EEG artifacts using a high-speed eye tracker”. In: *IEEE Transactions on Biomedical Engineering* 59.8 (2012), pp. 2103–2110.

[59] Atilla Kilicarslan, Robert G Grossman, and Jose Luis Contreras-Vidal. “A robust adaptive denoising framework for real-time artifact removal in scalp EEG measurements”. In: *Journal of neural engineering* 13.2 (2016), p. 026013.

[60] Thea Radüntz, Jon Scouten, Olaf Hochmuth, and Beate Meffert. “Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features”. In: *Journal of neural engineering* 14.4 (2017), p. 046004.

[61] Laura Frølich, Tobias S Andersen, and Morten Mørup. “Classification of independent components of EEG into multiple artifact classes”. In: *Psychophysiology* 52.1 (2015), pp. 32–45.

[62] Irene Winkler, Stephanie Brandl, Franziska Horn, Eric Waldburger, Carsten Allefeld, and Michael Tangermann. “Robust artifactual independent component classification for BCI practitioners”. In: *Journal of neural engineering* 11.3 (2014), p. 035013.

[63] Nima Bigdely-Shamlo, Ken Kreutz-Delgado, Christian Kothe, and Scott Makeig. “EyeCatch: Data-mining over half a million EEG independent components to construct a fully-automated eye-component detector”. In: *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE. 2013, pp. 5845–5848.

[64] Chi Zhang, Li Tong, Ying Zeng, Jingfang Jiang, Haibing Bu, Bin Yan, and Jianxin Li. “Automatic artifact removal from electroencephalogram data based on a priori artifact information”. In: *BioMed research international* 2015 (2015).

[65] Roberto Guarnieri, Marco Marino, Federico Barban, Marco Ganzetti, and Dante Mantini. “Online EEG artifact removal for BCI applications by adaptive spatial filtering”. In: *Journal of neural engineering* 15.5 (2018), p. 056009.

[66] Xun Chen, Hu Peng, Fengqiong Yu, and Kai Wang. “Independent vector analysis applied to remove muscle artifacts in EEG data”. In: *IEEE Transactions on Instrumentation and Measurement* 66.7 (2017), pp. 1770–1779.

[67] Soojin Lee, Martin J McKeown, Z Jane Wang, and Xun Chen. “Removal of high-voltage brain stimulation artifacts from simultaneous EEG recordings”. In: *IEEE Transactions on Biomedical Engineering* 66.1 (2019), pp. 50–60.

[68] Xun Chen, Xueyuan Xu, Aiping Liu, Martin J McKeown, and Z Jane Wang. “The use of multivariate EMD and CCA for denoising muscle artifacts from few-channel EEG recordings”. In: *IEEE Transactions on Instrumentation and Measurement* 67.2 (2018), pp. 359–370.

[69] Xun Chen, Qiang Chen, Yu Zhang, and Z Jane Wang. "A novel EEMD-CCA approach to removing muscle artifacts for pervasive EEG". In: *IEEE Sensors Journal* (2018).

[70] Ajay Kumar Maddirala and Rafi Ahamed Shaik. "Separation of sources from single-channel EEG signals using independent component analysis". In: *IEEE Transactions on Instrumentation and Measurement* 67.2 (2018), pp. 382–393.

[71] Christian Andreas Edgar Kothe and Tzyy-Ping Jung. "Artifact removal techniques with signal reconstruction". In: *U.S. Patent Application No. 14/895,440* (2014).

[72] Md Kafiul Islam, Amir Rastegarnia, and Zhi Yang. "Methods for artifact detection and removal from scalp EEG: a review". In: *Neurophysiologie Clinique/Clinical Neurophysiology* 46.4-5 (2016), pp. 287–305.

[73] Tim R. Mullen, Chreistian A.E. Kothe, M. Chi, A. Ojeda, T. Kerth, S. Makeig, T.-P. Jung, and G. Cauwenberghs. "Real-time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG". In: *IEEE Trans. on Biomed. Eng.* 62 (2015), pp. 2553–2567.

[74] Harshani Perera, Mohd Fairuz Shiratuddin, and Kok Wai Wong. "A review of electroencephalogram-based analysis and classification frameworks for dyslexia". In: *International Conference on Neural Information Processing*. Springer, 2016, pp. 626–635.

[75] Thomas C Bulea, Saurabh Prasad, Atilla Kilicarslan, and Jose L Contreras-Vidal. "Sitting and standing intention can be decoded from scalp EEG recorded prior to movement execution". In: *Frontiers in neuroscience* 8 (2014), p. 376.

[76] Sarah Blum, Nadine Jacobsen, Martin Georg Bleichner, and Stefan Debener. "A Riemannian modification of Artifact Subspace Reconstruction for EEG artifact handling". In: *Frontiers in Human Neuroscience* 13 (2019), p. 141.

[77] Laurel J Gabard-Durnam, Adriana S Mendez Leal, Carol L Wilkinson, and April R Levin. "The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): standardized processing software for developmental and high-artifact data". In: *Frontiers in neuroscience* 12 (2018), p. 97.

[78] Alejandro Ojeda, Marius Klug, Kenneth Kreutz-Delgado, Klaus Gramann, and Jyoti Mishra. "A Bayesian framework for unifying data cleaning, source separation and imaging of electroencephalographic signals". In: *bioRxiv* (2019), p. 559450.

[79] Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini, and Tzyy-Ping Jung. "Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal". In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2018, pp. 1242–1245.

[80] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website". In: *NeuroImage* 198 (2019), pp. 181–197.

[81] Jing Hu, Chun-sheng Wang, Min Wu, Yu-xiao Du, Yong He, and Jinhua She. “Removal of EOG and EMG artifacts from EEG using combination of functional link neural network and adaptive neural fuzzy inference system”. In: *Neurocomputing* 151 (2015), pp. 278–287.

[82] Junfeng Gao, Chongxun Zheng, and Pei Wang. “Online removal of muscle artifact from electroencephalogram signals based on canonical correlation analysis”. In: *Clinical EEG and neuroscience* 41.1 (2010), pp. 53–59.

[83] R.-S. Huang, T.-P. Jung, and S. Makeig. “Tonic changes in EEG power spectra during simulated driving”. In: *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* 5638 (2009), pp. 394–403.

[84] George H Klem, Hans Otto Lüders, HH Jasper, C Elger, et al. “The ten-twenty electrode system of the International Federation”. In: *Electroencephalogr Clin Neurophysiol* 52.3 (1999), pp. 3–6.

[85] Arnaud Delorme and Scott Makeig. “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis”. In: *J. of neurosci. methods* 134(1) (2004), pp. 9–12.

[86] Te-Won Lee, Mark Girolami, and Terrence J. Sejnowski. “Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources”. In: *Neural Comput.* 11.2 (1999), pp. 417–441.

[87] Harold W Kuhn. “The Hungarian method for the assignment problem”. In: *Naval research logistics quarterly* 2.1-2 (1955), pp. 83–97.

[88] A. Delorme, J. Palmer, J. ONton, R. Oostenveld, and S. Makeig. “Independent EEG Sources Are Dipolar”. In: *PLoS ONE* 7 (2012).

[89] Julie Onton, Marissa Westerfield, Jeanne Townsend, and Scott Makeig. “Imaging human EEG dynamics using independent component analysis”. In: *Neuroscience & biobehavioral reviews* 30.6 (2006), pp. 808–822.

[90] Alain Delorme. *Dipfit plug-in: Equivalent dipole source localization of independent components—DIPFIT validation study using the spherical head model*.

[91] CJ Henderson, SR Butler, and A Glass. “The localization of equivalent dipoles of EEG sources by the application of electrical field theory”. In: *Electroencephalography and clinical neurophysiology* 39.2 (1975), pp. 117–130.

[92] Luca Pion-Tonachini, Sheng-Hsiou Hsu, Chi-Yuan Chang, Tzyy-Ping Jung, and Scott Makeig. “Online Automatic Artifact Rejection using the Real-time EEG Source-mapping Toolbox (REST)”. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2018, pp. 106–109.

[93] Christian Andreas Kothe and Scott Makeig. “BCILAB: a platform for brain–computer interface development”. In: *Journal of neural engineering* 10.5 (2013), p. 056014.

[94] Luca Pion-Tonachini, Sheng-Hsiou Hsu, Scott Makeig, Tzyy-Ping Jung, and Gert Cauwenberghs. “Real-time EEG Source-mapping Toolbox (REST): Online ICA and source localization”. In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE. 2015, pp. 4114–4117.

[95] Federica Degno, Otto Loberg, Chuanli Zang, Manman Zhang, Nick Donnelly, and Simon P Liversedge. “Parafoveal previews and lexical frequency in natural reading: Evidence from eye movements and fixation-related potentials.” In: *Journal of Experimental Psychology: General* 148.3 (2019), p. 453.

[96] Federica Degno and Simon P Liversedge. “Eye movements and fixation-related potentials in reading: a review”. In: *Vision* 4.1 (2020), p. 11.

[97] Lisandro N Kaunitz, Juan E Kamienkowski, Alexander Varatharajah, Mariano Sigman, Rodrigo Quian Quiroga, and Matias J Ison. “Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task”. In: *NeuroImage* 89 (2014), pp. 297–305.

[98] Benjamin W Tatler, Dan Witzner Hansen, and Jeff B Pelz. “Eye movement recordings in natural settings”. In: *Eye Movement Research*. Springer, 2019, pp. 549–592.

[99] Nancy K Squires, Kenneth C Squires, and Steven A Hillyard. “Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man”. In: *Electroencephalography and clinical neurophysiology* 38.4 (1975), pp. 387–401.

[100] Terence W Picton. “The P300 wave of the human event-related potential”. In: *Journal of clinical neurophysiology* 9.4 (1992), pp. 456–479.

[101] Anneli Olsen. “The Tobii I-VT fixation filter”. In: *Tobii Technology* 21 (2012), pp. 4–19.

[102] Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini, and Tzyy-Ping Jung. “Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings”. In: *IEEE Transactions on Biomedical Engineering* 67.4 (2019), pp. 1114–1121.

[103] Oleg V. Komogortsev, Denise V. Gobert, Sampath Jayarathna, Do Hyong Koh, and Sandeep M. Gowda. “Standardization of Automated Analyses of Oculomotor Fixation and Saccadic Behaviors”. In: *IEEE Transactions on Biomedical Engineering* 57.11 (2010), pp. 2635–2645. DOI: 10.1109/TBME.2010.2057429.

[104] Susan M Munn, Leanne Stefano, and Jeff B Pelz. “Fixation-identification in dynamic scenes: Comparing an automated algorithm to manual coding”. In: *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. 2008, pp. 33–42.

[105] Dario D Salvucci and Joseph H Goldberg. “Identifying fixations and saccades in eye-tracking protocols”. In: *Proceedings of the 2000 symposium on Eye tracking research & applications*. 2000, pp. 71–78.

[106] Frederick Shic, Brian Scassellati, and Katarzyna Chawarska. “The incomplete fixation measure”. In: *Proceedings of the 2008 symposium on Eye tracking research & applications*. 2008, pp. 111–114.

[107] Pieter Blignaut. “Fixation identification: The optimum threshold for a dispersion algorithm”. In: *Attention, Perception, & Psychophysics* 71.4 (2009), pp. 881–895.

[108] Marcus Nyström and Kenneth Holmqvist. “An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data”. In: *Behavior research methods* 42.1 (2010), pp. 188–204.

[109] Juan E Kamienkowski, Matias J Ison, Rodrigo Quian Quiroga, and Mariano Sigman. “Fixation-related potentials in visual search: a combined EEG and eye tracking study”. In: *Journal of vision* 12.7 (2012), pp. 4–4.

[110] Sture Holm. “A simple sequentially rejective multiple test procedure”. In: *Scandinavian journal of statistics* (1979), pp. 65–70.

[111] Laura Frølich, Tobias S. Andersen, and Morten Mørup. “Automatic removal of eye movement and blink artifacts from EEG data using blind component separation”. In: *Psychophysiology* 41 (2004), pp. 313–325.

[112] Alois Schlögl, Claudia Keinrath, Doris Zimmermann, Reinhold Scherer, Robert Leeb, and Gert Pfurtscheller. “A fully automated correction method of EOG artifacts in EEG recordings”. In: *Clinical neurophysiology* 118.1 (2007), pp. 98–104.