Title: Practical Emotion Recognition using Wearable Brain and Physiological Sensors

Author(s): Thammasan, Nattapong

Citation

Issue Date

Text Version: ETD

URL: https://doi.org/10.18910/69712

DOI: 10.18910/69712

Note

Osaka University Knowledge Archive: OUKA
https://ir.library.osaka-u.ac.jp/repo/ouka/all/

Osaka University
Practical Emotion Recognition using Wearable Brain and Physiological Sensors

Submitted to
Graduate School of Information Science and Technology
Osaka University

January 2018

Nattapong THAMMASAN
Research Output

Journal Publications


Publications in Peer-reviewed International Conference Proceedings


4. Nattapong Thammasan, Megumi Iwano, Koichi Moriyama, Ken-ichi Fukui, Kornraphop Kawintiranon, Yanika Buatong, Shigenori Inagaki Tetsuichi Wazawa,


Publications in Non-Peer-reviewed International Conference Proceedings


Publications in Non-Peer-reviewed Domestic Conference Proceedings


Abstract

Detecting user emotion will play an important role in bridging the gap between human and computers. Recently, brain and physiological signals have been employed to detect emotional cues of human subjects with the assumption that bodily signals would provide information of intrinsic emotions better than conventional approaches including computer vision and speech analysis. However, conventional devices for brain imaging and physiological recording featured in previous works tend to be obtrusive since they were originally designed for use in clinical and controlled environments. This drawback limits the practicability of emotion recognition systems. Recently, a variety of wearable brain and physiological sensors have been developed which demonstrate potential in the emotion detection domain but come with the significant challenges regarding signal quality and stability. Hence, the objective of this study is to improve the practicability of emotion recognition system by using wearable brain and physiological sensors without significantly degrading the performance. In particular, this study has two main focuses. Firstly, the study employs multiple wearable sensors, including electroencephalographic (EEG) headset, chest-attachable electrocardiogram (ECG) patch, and wrist-worn galvanic skin response (GSR) band, with the aim of improving the robustness of the system by implementing efficient multimodal integration. Hereby, this study proposes making use of the reliability information of each modality, quantified by signal quality and accelerometer data, to regulate the information ensemble. The empirical results from experiments with 30 subjects performing music-listening tasks demonstrate that the context-aware system significantly outperforms traditional approaches in arousal and valence classification. Secondly, this study addresses limitations of existing systems with regards to accommodating new users by minimizing the amount of calibration data required to make use of the system. Conventional generalized systems designed to detect emotions tend to suffer from degraded performance due to inter-subject variability in bodily signals, especially with regards to EEG. This necessitates collection of calibration data recordings which can be time-consuming, annoying and reducing the practicability of the entire system. To mitigate this shortcoming, an emerging technique called transfer learning is adopted. This technique can reduce calibration data requirements by allowing the use of information collected from other subjects to build a model for a new subject. This reduced data requirement streamlines the process of adding new users to the system, and empirical results also demonstrate the method’s potential for enabling subject-independent emotion recognition. The proposed method may shed light on developing more practical emotion recognition systems for real-world applications.
Acknowledgement

The dissertation would not have been possible without supports and helps of my professors, family, and friends. I would like to express my sincere gratitude first and foremost to Professor Masayuki Numao for his constant and thoughtful guidance, encouragement, beneficial feedback, and various types of supports that allowed me to learn on innumerable occasions. I would also like to express my very great appreciation to Associate Professor Koichi Moriyama for his gracious advice, for stimulating a number of fruitful discussions, and for his persistent help on manuscript revision. I would also like to thank Associate Professor Ken-ichi Fukui for his continually comments and enlightenment with regards to my research and for a variety of opportunities in my academic endeavor. My gratitude also goes out to the other member of the committee for their great comments and contributions toward my doctoral degree.

I would also like to thank the wonderful ex-members of Numao laboratory, Dr. Roberto Legaspi, Dr. Rafael Cabredo, Dr. Paul Salvador Inventado, Dr. Danaipat Sodkomkham, Mr. Vanus Vachiratamporn, and Mr. Yu Yamano, who have shared their intellectual opinions at the beginning of my research. I am also grateful to other Numao laboratory students who have been substantially sharing their thoughts on my research, preparing demonstrations, and having fun moments together, especially Juan Lorenzo Hagad who has been closely working with me in many collaborative works. I would like to acknowledge the laboratory secretaries Ms. Misuzu Yuuki, Ms. Megumi Tanabe, Ms. Azusa Hirabayashi, Ms. Mika Kusakabe, Ms. Mitsuyo Otsuka, and Ms. Akiko Yamamoto for their accommodating and assisting me with all of the university affairs.

I would not have had the honor of being a student of Osaka University without the generous scholarship of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. In addition, I would like to express my gratitude and appreciation to the Humanware Innovation Program at Osaka University, the Program for Leading Graduate Schools (LGS) of MEXT. Since the beginning of my graduate study, this program has been financially supporting my research, providing opportunities to learn on transferable skills, helping my connection establishment with LGS students across the country, opening new worlds for me by a number of domestic and oversea study-tour programs, and sponsoring travel grants for domestic and international conference participation. Most importantly, the program also funds an oversea internship program that allowed me to conduct excellent collaborative research with Swartz Center for Computational Neuroscience (SCCN), a Center of the Institute for Neural Computation, the University of California San Diego. I would like to express my sincere appreciation to Dr. Makoto Miyakoshi, who took care for me during the visiting
period, enlightened me on valuable knowledge regarding computational neuroscience, closely worked with me and gave comments on works, revised manuscript, and helped me in a number of occasions both inside and outside working place. I also owe a debt of gratitude to Dr. Scott Makeig, the director of the center for his fabulous comments and accepting my application to work in a top-class research institute.

I would also like to thank Center of Innovation (COI) Program, Osaka University for a plenty of unforgettable opportunities allowing me to make research presentation in front of tremendous audiences and learn from feedbacks and for supporting me to take parts of the close-loop music composition project based on emotion recognized by wearable brain sensor via discussion and system demonstration. COI program also provides me numerous opportunities to use wearable devices obtained from its collaborative project with imec-Netherlands in this research. I also thank Mr. Bernard Grundlehner at imec who has been collaboratively working with me on sensor developing and customization, providing me various suggestions, and taking care of me during my imec site-visit.

I would like to express my appreciation to my friends, Dr. Adisorn Ratanayotha for his introductory tutorial on neuroscience, Ms. Salita Sombatsiri for her deep discussion regarding electrophysiological signal processing, Dr. Songpol Chaunchaiyakul and Dr. Kamalas Udomlamlert for their helpful suggestions on the academic career. I would also thank my friends at Osaka University for sharing good moments that could fabulously refresh me. In particular, I would like to thank my experimental subjects who had devoted their time from their busy schedules and contributed to this research. I am forever grateful to my family who has always been supportive and encouraging me throughout my study. I also thank my beloved who has been there for me whenever I feel discouragement no matter how difficult it was and helped me make important decisions.

I also thank Professor Boonserm Kijsirikul who opened and recommended me this pathway to be a student at Osaka University as well as many professors at Chulalongkorn University too innumerable to list here for their guidance and supports ever since I was an undergraduate student. A very special gratitude goes out to Dr. Peerapon Vateekul at Chulalongkorn University who had been sharing a memorable time with me during his research-exchange visit at Osaka University in summer 2014 and has been working on collaborative research with me since then. His ideas, comments, and refinement are valuable for shaping the experimental protocol of this research. I also thank internship students from Thailand who worked under my mentorship and shared fruitful ideas and results. Last but by no means least, I would like to thank countless inspiring scientists and persons who have been indirectly motivating me to overcome many obstructions during my study.
Contents

List of Figures iii
List of Tables v

1 Introduction 1
1.1 Background to the Research ......................... 1
1.2 Overview of Related Works .......................... 3
1.3 Significance of the Research ....................... 4
1.4 Organization ........................................ 5

2 Designing Experiment on Dynamic Music-Emotion Recognition 6
2.1 Critical Research Questions .......................... 7
  2.1.1 Temporal continuity of emotion recognition ........ 8
  2.1.2 Familiarity effects ................................ 9
  2.1.3 Features and Window Size ......................... 16
  2.1.4 Application of State-of-the-art machine learning techniques . . 18
  2.1.5 Annotation Smoothing ............................. 19
  2.1.6 Incorporating information from musical features ........ 20
2.2 Experimental Protocol ................................ 22
  2.2.1 Subject Recruiting and Questionnaires ............. 23
  2.2.2 Musical Stimuli .................................... 23
  2.2.3 Music-listening Task ............................... 27
  2.2.4 Brain and Physiological Recording ................. 28
2.3 Summary .............................................. 29

3 Emotion Recognition based on Electroencephalography 32
3.1 Questionnaires Results ................................ 32
3.2 Signal Preprocessing .................................. 34
  3.2.1 Band-pass Filters .................................. 34
  3.2.2 Artifact Rejection using ASR method ................ 35
  3.2.3 Rejecting Epochs in Resting-State Data ............ 36
  3.2.4 Decomposition by ICA and IC Evaluation .......... 36
  3.2.5 EEG signals Reconstruction ....................... 39
3.3 Feature Extraction .................................... 39
  3.3.1 Time Domain Features ............................. 40
3.3.2 Frequency Domain Features ........................................... 41
3.3.3 Time-Frequency Domain Features ................................. 41
3.3.4 Cross-Channel Features .............................................. 42
3.4 Feature Adjustment by Subtracting the Baseline .................... 43
3.5 Emotion Classification .................................................. 45
3.6 Emotion Classification Result ......................................... 46
3.7 Enhancing Performance by Feature Selection ....................... 47
3.8 Summary ................................................................. 49

4 Multimodal Adaptive Emotion Recognition .......................... 51
4.1 Data Acquisition and Signal Preprocessing ......................... 51
4.2 Feature Extraction ..................................................... 52
4.3 Signal Quality Quantification ......................................... 53
4.4 Decision-level Multimodal Fusion .................................... 55
4.5 Results ................................................................. 57
4.6 Summary ................................................................. 59

5 Toward Subject-independent Emotion Recognition using Transfer Learning ........................................... 60
5.1 Methodologies .......................................................... 61
5.1.1 Acquiring Classifiers ............................................... 61
5.1.2 Assessing the Extent of Similarity ................................ 62
5.1.3 Experimental settings ............................................. 65
5.2 Results ................................................................. 66
5.3 Summary ................................................................. 69

6 Discussion and Conclusion .............................................. 70
6.1 Discussion ............................................................... 70
6.1.1 Access to testing data .............................................. 70
6.1.2 Comparison to previous works using conventional devices ... 71
6.1.3 Limitation .......................................................... 72
6.2 Conclusion ............................................................. 72
6.3 Recommendation for future studies .................................. 73

A Supplementary Data ...................................................... 82
List of Figures

2.1 Average arousal and valence classification accuracy for all subjects in the conducted experiment ................................................................. 8
2.2 Arousal and valence annotations from subject 4 and their estimation by the model constructed with SVM and FD values from all instances .......... 9
2.3 A topological plot of the variation of average PSD values across subjects produced by songs with high and low music familiarity from the conducted experiment and DEAP dataset ........................................ 13
2.4 Functional connectivity with significant difference values owing to music familiarity from the conducted experiment and DEAP dataset .......... 14
2.5 Arousal and valence classification accuracies above the chance levels for high familiarity (familiar songs), low familiarity (unfamiliar songs), and combined (all songs) data groups from the conducted experiment and from the DEAP dataset ................................................................. 16
2.6 Averages across subjects of the accuracy over chance level in valence and arousal classification with FD, PSD, and DWT features. ................. 17
2.7 An example of resulting smoothed annotation data compared with the original annotation data of a song by Subject 12 ................................. 20
2.8 Performance of continuous valence and arousal recognition with and without applying smoothing technique reported by MCCs, Pearson correlations, and MSEs. .......................................................... 21
2.9 Averaged emotion classification MCCs across subjects with varying weighting factors ............................................................ 22
2.10 Experimental Protocol ................................................................. 27
2.11 A screenshot of the developed software ........................................... 28
2.12 Wearable sensors used in the experiment ......................................... 29
3.1 Results of Edinburgh Handedness Inventory ..................................... 33
3.2 Results of Profile of Mood States (POMS) ....................................... 33
3.3 Results of Emotion Awareness Questionnaire (EAQ) ......................... 34
3.4 Signal Preprocessing Protocol ...................................................... 35
3.5 Visualization of the properties of brain IC, eye-movement IC, and muscular-activity IC ................................................................. 37
3.6 Correlation of PSD-band features and arousal/valence ratings comparing between features before and after adjustment by subtracting the baseline 43
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>Correlation of time domain features and arousal/valence ratings com-</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>paring between features before and after adjustment by subtracting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the baseline</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>Valence and Arousal Classification Accuracies of each feature com-</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>paring between using features before and after adjustment by sub-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tracting the baseline</td>
<td></td>
</tr>
<tr>
<td>3.9</td>
<td>Emotion classification accuracies when varying the number of re-</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>maining features separated by the used features</td>
<td></td>
</tr>
<tr>
<td>3.10</td>
<td>Valence and Arousal Classification Accuracies of each feature com-</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>paring between using all feature set and applying feature se-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lection</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Visualization of EEG instability index of an exemplified data drawn</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>from Subject 11</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Emotion recognition accuracies of each unimodality and each method</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>of multimodal fusion</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Valence and arousal classification accuracies comparing between EE-</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>G unimodality and multimodality</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Distance matrices comparing the difference between subjects using va-</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>rious methods</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>Emotion classification accuracies of subject-independent classifica-</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>tion using each method</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>The difference between valence/arousal classification accuracies and</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>the corresponding chance levels when using transfer learning tech-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nique to learn from 5-similar subjects</td>
<td></td>
</tr>
</tbody>
</table>
List of Tables

2.1 Musical stimuli used in the experimentss ........................................ 24
2.2 Wearable device specification ......................................................... 31
3.1 Frequency Band Ranges and Decomposition Levels of EEG signals Recorded at a sample frequency of 256 Hz ........................................ 41
4.1 Summary of unimodal and multimodal methods for emotion classification 57
5.1 Summary of methods used for subject-independent emotion classification. 67
A.1 Edinburgh Handedness Inventory ................................................. 82
A.2 Profile of Mood States (POMS) – Short Form .................................. 83
A.3 Emotion Awareness Questionnaire (EAQ) ...................................... 84
A.4 Additional information of musical stimuli ..................................... 85
Chapter 1

Introduction

1.1 Background to the Research

Emotion is an essential element of human nature and behavior and plays a crucial role in the interaction between each other and external world. Meanwhile, computer was originally designed to be an intelligent machine capable of assisting human in logical, rational, and predictable ways, but it failed to interact with human intelligently owing to the lack of understanding human emotion. Accordingly, recognizing human emotion could be a key to bridge the gap between human and computer in natural communication resulting in the advent of affective computing research field around two decades ago [73]. Although scientists have diverse opinions about emotion, researchers in affective computing agree that a change of external and internal state triggers the emergence of emotion that is reflected by the alternation of external appearances including facial expression, voice, gesture, posture, and other physiological conditions [45], and consequently propose frameworks that give computer the ability to recognize human emotion by applying computational intelligence and machine learning techniques to capture the patterns from external manifestation of emotion [14].

Recognizing emotion from external appearance such as facial or voice expression can lead to inaccurate inference as that information might not always correspond to a specific emotion as subject might suppress or pretend to exhibit false emotion [45]. In contrast, accessing bodily responses, such as scalp potential, heart rate, skin conductivity, that are closer to emotion-related cognitive process might be better capable of providing intrinsic information regarding emotion. Accordingly, affective computing researchers have embraced the brain and physiological measurements as the merely or supplementary methods to estimate emotional state of human. However, information
from single modality might not always sufficient to recognize emotion effectively, hence current research is targeted to integrate information from multiple modalities to improve the reliability of emotion estimation. Nevertheless, this introduces a new challenge for recognizing emotion using brain and physiological signals since increasing the number of sensors leads to spending a longer time in device set-up and the reduced comfortability as the devices tend to be obtrusive since they were originally designed for medical use or in controlled environments. This shortcoming limits the practicability of emotion recognition system constructed in previous works.

Fortunately, recent years have witnessed the growth of the development of wearable sensors and systems that enable the recordings of brain and physiological signals in effective and more naturalistic ways. Although these sensors are primarily designed for physiological monitoring that could help in medical diagnosis and treatment, they might serve as practical devices enabling emotion recognition in real-world scenarios but this potential has not been studied in previous works. Hence, the objective of this research is to demonstrate making use of wearable brain and physiological sensors to recognize human emotion with the aim to improve the practicability of system and achieve satisfactory accuracy. The empirical results from experiments with 30 subjects using electroencephalographic (EEG) headset, chest-attachable electrocardiogram (ECG) patch, and wrist-worn galvanic skin response (GSR) band are presented. In addition, two main concerns in multimodal emotion recognition based on bodily signals are addressed.

First, multimodal emotion recognition cannot always guarantee the increased performance. In particular, when a single modality is contaminated by noise, blindly combining information sources from every modality might degrade the performance of system rather than enhance it. In contrast, rejecting signal from one modality could adversely affect the synchrony and accordance with the other modalities. Instead of discarding information, evaluating the reliability of each modality and using it in regulating information integration might help mitigate the problem. In this research, the context-aware system is proposed to answer the extremely important question on which modality to rely in information ensemble. Hereby, the context is defined as the stability of each modality quantified by signal quality and accelerometer data. This novel technique was found outperforming the straightforward majority-voting and unimodal approaches.

Second, inter-subject variability in bodily signals impedes the success in constructing the generalized emotion recognition system and necessitates the collection of calibration
data from a newly-coming subject which can be a time-consuming and annoying process. More recently, transfer learning has been proposed to alleviate the issue by allowing the use of information from another subject to build a model for a new subject. This research proposes a novel technique of measuring subject transferability which determines how proper a subject could be for including into the population of source subjects that would be used to build a model for a target subject. The superior performance over the traditional method that discards subject transferability yet blindly makes use of all available source subjects is presented.

While emotion can be induced by a variety of stimuli, music is considered an extraordinary material to elicit strong emotions and evoke a wide variety of emotions [40]. Besides, studying neural and physiological correlates with music could enable the application of music in therapy and uplifting emotional state. Music is thus used as stimuli in this research. Furthermore, although scientists have various ways to represent human emotion, the commonly-used arousal-valence model [81] is employed.

1.2 Overview of Related Works

There are several neural pathways involving in the experience and expression of emotion, from cerebral cortex in central nervous system (CNS) to autonomic nervous system (ANS) [5]. The limbic system was found to be an underlying process of experiencing emotion and includes a number of small neural structures including amygdala, hypothalamus, mamillary body, cortex nearby corpus callosum (mainly the cingulate gyrus), and hippocampus [50]. In addition, the interaction between neocortex and the limbic system also governs the behavioral expression of social emotion. Consequently, scalp electrodes of EEG were expected to be capable of capturing emotion-related cortical activities in neocortex and the brain signatures resulted from emotional processes. Recently, researchers have reported EEG correlates with emotion in patterns of frontal EEG asymmetry [87], error negativity [13], and power spectra [66, 32], although the inconsistency between discoveries of each study still exists. Since then, there has been the explosion of attention in applying computational approach to automatically map particular patterns in EEG signals and the human emotion to obtain the model for emotion recognition [2]. EEG has the advantage of being a non-invasive neuroimaging tool with excellent temporal resolution. Compared to other non-invasive methods including functional magnetic resonance imaging (fMRI) and magnetoencephalogram (MEG), EEG has lower cost and
does not limit the movement of subjects posing the possibility to employ in real-world situation. By these reasons, the past decade has witnessed the enormous growth of emotion recognition research utilizing EEG signals, stimulating the emergence of affective brain-computer interface (aBCI), which is a highly multidisciplinary research attempting to measure affective states from neurophysiological responses and apply the resulting information to improve the interaction with computer [66]. A variety of EEG feature extraction [34] and machine learning techniques [57, 39] have been adopted with the aim to recognize emotion more accurately. However, this research field has encountered various challenges, especially robustness of system and subject variability, limiting the use merely in controlled environments.

Co-occurring with CNS responses, it has been discovered that emotion also triggers different activities of peripheral physiological responses in ANS [46, 27] including heart rate, skin conductance, respiration, and muscle tension [17]. Accordingly, there is also considerable interest in recognizing emotion from physiological signals [38, 47]. Similar to CNS responses, subject’s reaction in ANS can reveal intrinsic emotion and is not easily concealed by the subject thus enriches reliable information sources for estimating emotional states. However, the robustness of the system is still challenging.

It might be reasonable to assume that information from single modality might not be adequate for accurately recognizing emotion leading to the introduction of using multimodal approaches to leverage the robustness of the system [22, 106]. By this way, EEG signals have been used in conjunction with physiological signals [43, 92, 95] or even with another type of information such as facial expression [44], eye gaze [58], and audio signals [92]. As mentioned earlier, multimodal approaches, however, still confront with a number of challenges. It is not always necessary that each modality would provide complimentary information in ensemble process; a certain modality might contribute less by being exposed to noise, so blindly integrating multiple information sources cannot always enhance the performance but instead degrade the system. Unfortunately, the previous works have not adequately investigated the issue. Besides, using higher number of sensors also poses the formidable problem of the increased exhaustive set-up time.

1.3 Significance of the Research

By addressing the above concerns, this research contributes to emotion recognition research with the primary objective to increase the practicability of system. First, this
research presents an early study of employing wearable brain and physiological sensors to recognize emotion and the obtained results were adequately satisfactory. Second, the question which modality to rely on at particular time in multimodal approach is answered by introducing a new context-based adaptive fusion technique. In particular, the relative stability of the individual modality is used to adjust its contribution in decision-level fusion. Finally, the subjective variability that necessitates the long-recording of calibration data for a new-coming user is mitigated by applying a newly proposed transfer learning technique. In particular, the novel method of measuring the subject transferability, which determines the degree of appropriateness of data possessed by a subject to be included into the source space that is later used to build emotion recognition model for a new-coming user, is proposed.

The success of this research can shed light on emotion recognition research toward more practical system that can be potentially used in real-life. With special consideration on the emotion variation over time, the prospective system can provide emotion estimation in real-time manner enabling a variety of application both in medical (e.g., diagnosis, monitoring, or assistance) and non-medical (e.g., entertainment, marketing, or education) domains.

1.4 Organization

The rest of this dissertation is organized as follows: Chapter 2 is dedicated to designing experiment for emotion recognition in dynamic fashion based on the assumption that emotion is subjective and varies over the course of time. The previous studies answering critical research questions in experimental design and emotion recognition model construction are presented in the first half, whereas the experimental protocol that was resulted from deliberately taking into account the implication from those studies, is presented in the second half of the chapter. To demonstrate the usability of wearable EEG headset, Chapter 3 presents the study of emotion recognition by solely using EEG data, and the following chapter shows the results of fusion with another physiological sensor and the proposed adaptive context-aware approach is elucidated here. Then, Chapter 5 explains the proposed framework for subject-independent emotion recognition using transfer learning method and is followed by Chapter 6 that discusses and concludes this research.
Chapter 2

Designing Experiment on Dynamic Music-Emotion Recognition

The design of emotion recognition system based on bodily signals includes the creation of a model capable of inferring emotional states of user from sensor data. To construct a model, the collection of training data is a vital step before extracting informative features from them and then training classification or regression model with the resultant features in a supervised manner. Accordingly, the good quality of training data is extremely important to obtain well-performing emotion recognition system and it can be achieved by carefully designing experiment for data collection in psycho-physiologically appropriate manners. Despite the fact that music is capable of eliciting strong and various emotions, successfully using music to stimulate emotion for data collection experiment has become a grand challenge for music-emotion researchers for many years. Emotion in listening to music is contingent upon a variety of factors including properties of songs and subjective conditions such as age, gender, cultural background, music familiarity, or even pre-experiment mood [115]. Apart from that, training model to recognize emotion with the obtained experimental data still has a number of factors and parameters awaiting to be clarified.

The first half of this chapter is, therefore, devoted to the presentation of early contributive studies carried out by the author attempting to investigate the unanswered questions in music-emotion recognition. The insights learned from these studies largely contributed to the design of the experiment for data collection of this research which is described in the second half of this chapter.
2.1 Critical Research Questions

Despite the growth in popularity in aBCI, numerous fundamental yet critical research questions have not been fully studied resulting in the difficulty in generalizing the research conclusion across studies. The following session presents early attempts carried out by the author to clarify on certain unanswered issues including temporal continuity of emotion, music familiarity, appropriate features, effective classification techniques, annotation manipulation, and multimodality.

All of these studies employed the same dataset constructed from experiments of 15 subjects undergoing music-listening tasks while EEG signals were also recorded [101]. Fifteen males (mean age = 25.52, SD = 2.14) participated in the experiments. The music collection is a set of MIDI files comprised of 40 instrumental pop songs having different instrument and tempo. At the beginning, each subject was instructed to select 8 familiar songs and 8 unfamiliar songs from a collection of 40 musical excerpts and rate the level of familiarity from 1 to 6, while short samples of songs were also offered. Scalp EEG signals were recorded by a Waveguard EEG cap\(^1\), which is designed for medical usage and was placed on the subject’s head. The 12 electrodes, namely Fp1, Fp2, F3, F4, F7, F8, Fz, C3, C4, T3, T4, and Pz, out of 21 available electrodes were utilized as these electrodes are located nearby frontal lobe which is known to play a key role in emotion regulation [41, 87]. The sampling frequency was down-sampled to 250 Hz, and the impedance of each electrode was kept below 20 kΩ. Notch filter was also applied to attenuate power line noise. In each experiment, musical excerpts were presented to subject as MIDI sounds whose averaged duration was 2 minutes, while the subject was encouraged to close eyes and minimize body movement. Between each song, a 16-second resting period was included. After presenting all of the musical stimuli, EEG cap was removed and each subject proceeded to an annotation session. During the session, the subject was instructed to listen to the same set of songs again with additional instruction to annotate the emotion felt in the first listening session (with EEG recording) by continuously clicking on corresponding points in the arousal-valence space shown on a monitor. A brief guideline of the arousal-valence model was also given to acquaint the subject with this emotion model.

\(^1\)http://www.ant-neuro.com/products/waveguard
2.1.1 Temporal continuity of emotion recognition

Emotion while engaging with entertainment can change over time, for instance when listening to long-duration music, and this phenomenon has been reported in numerous studies using skin conductance measurement [25], fMRI [42], MEG [77], and EEG power spectra [84]. However, most of the previous works on EEG-based emotion recognition, unfortunately, neglected this temporal-variation characteristic of emotion [39] by the reason that the stimuli were shorter than 1 minute, which is far from the duration of general entertainment materials nowadays.

An early work of this research [101] serves as a very first attempt to take into account the characteristic of emotion that varies over the course of time. Unlike traditional method that each subject annotated emotion at song level, the subject in this experiment was allowed to annotate emotion at finer level by continuously clicking on arousal-valence space. This annotation data was used to label fractal dimension (FD) and power spectral density (PSD) features extracted from preprocessed EEG signals and then train support vector machine (SVM-Pearson VII function kernel (Puk) [104]), multi-layer perceptron (MLP) and C4.5 classifiers; the detail of feature extraction algorithms can be found in Section 3.3. A non-overlapping 4-second sliding window was
applied to allow tracking emotional change. For the sake of comparison, traditional methods that neglected emotional change over time was also simulated by expanding the size of the sliding window to the full length of the song and used majority voting to derive emotional label for that song. The performance on subject-dependent emotion recognition was evaluated by using 10-fold cross-validation approach, whereas the chance level was defined as a benchmark by the majority class of the testing data.

The results shown in Figure 2.1 demonstrated that considering emotion variation significantly improved the performance of arousal classification ($p < 0.01$) and valence classification ($p < 0.05$) compared with traditional approaches regardless of feature extraction or classification technique used. Furthermore, Figure 2.2 illustrates that the SVM model trained with all FD features can remarkably track the fluctuation of emotion of a subject. As can be seen, the model could handle the distinct oscillations of emotion appeared in the sessions of some songs, including the arousal shifts in song 2 and 12, and the valence alternation in song 1 and 5.

2.1.2 Familiarity effects

Experiencing music can be influenced by subjective issues including cultural background, age, gender, training, and familiarity level with the music [115]. In particular, listening to familiar music involves expectation and prediction based on prior knowledge to musical excerpts, memory of the listener might therefore play roles in musical percep-
tion and can influence the emotional responses. Recent neurophysiological studies have reported the correspondence between music familiarity and physiological signals. By using fMRI, a feeling of familiarity with music or odors was found to induce activation in the deep left hemisphere, while a feeling of unfamiliarity induced activation in the right hemisphere [75] and the researchers concluded that it is possible to trigger neural processes specific to the feeling of familiarity regardless of the type of triggering stimuli via processes that are likely related to the semantic memory system. Another fMRI study suggested that music familiarity is related to limbic, paralimbic, and reward circuitries reflected by the discovered correlation between brain activity and music appreciation [71]. A study on electrodermal activity also suggested that certain levels of expectation and predictability caused by familiarity play an important role in the experience of emotional arousal in response to music [105]. Evidence from event-related potentials (ERPs) observed along the frontocentral scalp demonstrated that musical melodies with a higher degree of familiarity producing more negative potentials potentially relevant to processing mechanism at the conceptual level [18]. Despite these evidences, previous works on EEG-based emotion recognition have overlooked the familiarity to musical stimuli as there has been no clear evidence direct from any full single-trial EEG study. On the other hand, if music familiarity indeed has an effect on EEG signals, ignoring familiarity would degrade the performance of EEG-based emotion recognition.

This reason motivated the origin of another early study of this research [100, 102] attempting to explore the EEG evidence of music familiarity and investigate the effect to emotion classification results. To acquire stronger evidences, two different datasets were employed; one was constructed from the conducted experiments as explained above [101], and one was drawn from the Database for Emotion Analysis using Physiological signals (DEAP) [43], an existing EEG database consisting of 32-channel EEG and peripheral physiological signals of 32 subjects performing 40 music-video watching tasks. Apart from emotional labels, each subject in DEAP experiment also assessed the level of familiarity to musical stimuli ranging from “never heard it before the experiment” (1 score) to “knew the song very well” (5 scores).
Investigation of EEG Correlates of Familiarity

To reveal EEG characteristics with regards to music familiarity, two types of analyses were performed: within-electrode analysis and cross-electrode analysis.

**Single-Electrode-Level Power Spectral Density Analysis** PSD features were extracted from EEG data in the conducted experiment to obtain delta-, theta-, alpha-, beta-, and gamma-band features. Similarly, theta-, alpha-, beta-, and gamma-band features were obtained from DEAP dataset (the raw signals were filtered above 4 Hz a priori resulting the impossibility to acquire delta-band PSD). A non-overlapping 4-second sliding window was applied to both dataset.

To investigate how the PSDs were affected by music familiarity and subject individuality, two-way analysis of variance (ANOVA) with replication was performed based on the test of null-hypotheses that the main effects of familiarity and subjectivity were not significant. For each frequency band and electrode, multiple PSDs were aggregated from all subjects and divided into two groups: low and high familiarity. Replication, i.e., multiple observations, involved obtaining multiple PSDs from each subject. The problem of instance-inequality across subjects was addressed by instance random selection technique. Tukey test was used to perform post-hoc comparisons. It should be noted that entire data or particular channel of some subjects in DEAP dataset were discarded because of excessive noises.

The results indicated the primary effects of inter-subject variability on the alternation of PSD. However, the familiarity was still found having statistically significant effects on PSDs in certain electrodes. To illustrate the familiarity effect, the difference of PSDs between both conditions (familiarity–unfamiliarity) was calculated within a subject and later grand-averaged across subjects. Figure 2.3 shows the topological plots of such differences on scalp map and indicates the positions where PSDs were found significantly distinctive owing to the extent of familiarity by the ANOVA technique ($p < 0.05$ for the conducted experiment and $p < 0.0001$ for DEAP dataset). In the DEAP dataset, the PSD alternation owing to familiarity was prominent in higher frequencies.

It was previously discovered that listening to unfamiliar songs relates to recollection, the cognitive ability to recall a former context associated with a musical excerpt by utilizing episodic memory [76]. It was hypothesized that subjects in the conducted experiment might recollect past experience from episodic memory to identify a novel song.
Previous research [12] that showed relatively higher gamma power over the parietal scalp during the act of recollection (as opposed to the act of experiencing familiarity) is consistent with the obtained results that showed a marginally higher gamma-PSD obtained from the Pz electrode while listening to an unfamiliar song. In addition, the implication of the frontal midline $\theta$ in working and episodic memory in which the associated memories could possibly be relevant to unfamiliar song listening was reported [31]. However, subjects in the DEAP experiment produced higher gamma and frontal midline $\theta$ power while watching familiar music videos; it might be reasonable to speculate that the subjects used memory to a greater extent to anticipate the next scene of a music video because they might have occasionally watched the music video versions of regularly listened-to songs. Unlike the conducted experiment, subjects in DEAP experiment who watched a particular music video for the first time or who had minimal experience with the video would engage so intensely enough in watching the video that they avoided using any recollection memory to associate the music with previous experiences. This evidence indicated that familiarity with video scenes had a higher influence on brain activities than familiarity with the music used as background sounds in the music video. Moreover, the increase in Fz $\theta$ power in the obtained results corresponds with previous reports of enhancement of frontal midline theta rhythm ($Fm\theta$) during focused attention [1]. A likely underlying reason for this is that song unfamiliarity induced the subjects to listen more attentively in order to successfully annotate emotions subsequently in the following phase.

**Functional Connectivity Analysis** A number of brain functions have been demonstrated to involve multiple brain regions. The analysis of the interrelation between EEG electrode pairs was found capable of offering more insights on the association between brain activities [82] including on the neural correlates of emotion [51].

To perform analysis in specific EEG frequency bands, a fifth order bandpass Butterworth filter was applied to extract EEG signals in multiple frequency bands. Then, the connectivity indexes were calculated from all pairs of electrodes separately in each frequency band using the three approaches that are sensitive to different characteristics of EEG signals; Correlation corresponds to the relationship between 2 signals from different brain sites. Given signals $x$ and $y$, the correlation at each frequency ($f$) is a
Figure 2.3: A topological plot of the variation of average PSD values across subjects produced by songs with high and low music familiarity (familiarity power – unfamiliarity power) from the conducted experiment (a) and DEAP dataset (b); positive areas represent regions in which high familiarity produces higher power than low familiarity, while negative areas depict where unfamiliarity produces higher power. White cross signs denote that PSDs taken while listening to music with high familiarity are significantly higher than those taken while listening to music with low familiarity and the gray cross signs have the same notation but in the opposite direction. Significant levels in the conducted experiment and DEAP dataset are $p < 0.05$ and $p < 0.0001$ respectively.

Function of cross-covariance $C_{xy}^f$ and auto-covariances, $C_{xx}^f$ and $C_{yy}^f$, of $x$ and $y$:

$$R_{xy}(f) = \frac{C_{xy}^f}{\sqrt{C_{xx}^f C_{yy}^f}}.$$  

(2.1.1)

Coherence is similar to correlation that also includes the covariation between 2 signals as a function of frequency. This index indicates how much two brain sites are working closely together at a specific frequency band. The coherence is a function of the respective power spectral densities, $P_{xx}(f)$ and $P_{yy}(f)$, of $x$ and $y$, and of the cross-power spectral density, $P_{xy}(f)$, of $x$ and $y$:

$$Coh_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}.$$  

(2.1.2)

Phase Synchronization Index (PSI) is a nonlinear measure of connectivity indicating the phase difference between signals from two different brain regions. PSI can be restricted to certain frequency bands reflecting specific brain rhythms. The PSI is defined as

$$PSI_{xy} = \left| \frac{1}{L} \sum_{l=0}^{L} e^{i[\phi_x(t) - \phi_y(t)]} \right|,$$  

(2.1.3)
where $\phi_x(t) = \arctan \tilde{x}(t) / x(t)$ is the Hilbert phase of signal $x$ and $\phi_y(t)$ is the phase of signal $y$, while $\tilde{x}(t)$ is the Hilbert transform of $x(t)$.

In each subject, a connectivity index was calculated from the filtered EEG signals in song level and then averaged within a group of familiarity (low-familiar and high-familiar groups) using arithmetic-mean method for coherence and PSI and quadratic-mean method for correlation. Afterwards, paired $t$-test was performed to discover any statistically significant difference in EEG functional connectivity associated with music familiarity on the aggregated data from all subjects.

Figure 2.4 illustrates pairs of electrodes whose functional connectivity significantly altered by the effect of familiarity. As an evidence from the conducted experiment,
the increase of connectivity was discovered mainly in the higher frequency bands when subjects listened to unfamiliar songs. A greater functional connectivity in the gamma band during an experience of recollection compared to that during an experience of familiarity was observed in a previous study [12], which is in line with the obtained results as higher connectivity owing to unfamiliar songs was found, especially in the gamma frequency range. In addition, another previous study reported higher delta and gamma band connectivity during performing autobiographical memory tasks [33]. In light of the hypothesis regarding episodic memory use during unfamiliar song listening, the obtained results are consistent with their findings. Additionally, functional connectivity in the DEAP experiment increased mainly in higher frequency bands when the subjects watched familiar music video excerpts. This incident may be related to cognitive recollection based on the speculation that the subjects perhaps made use of episodic memory to anticipate the next video scenes. Interestingly, the agreement between single-electrode-level analysis and functional connectivity analysis strongly confirmed that music familiarity elicited detectable changes in brain activities that might be relevant to memory recollection.

**Familiarity Effects in Emotion Recognition Systems**

As a consequence of the discovery that music familiarity affected EEG signals, further study on the merit of taking into account the music familiarity in emotion recognition was conducted using the feature extraction and classification techniques presented earlier. To demonstrate the benefit of controlling music familiarity, the data were separated into 2 groups in accordance with the familiarity level (low and high), and the third group that aggregated all data was also created representing the traditional approach that overlooked the familiarity effect.

The classification accuracies above the chance levels averaged over the subjects are shown in Figure 2.5. As can be observed, the performance of training using solely data from unfamiliar song sessions was superior to that using the overall dataset regardless of feature extraction techniques, classifying techniques, or dataset. In most of the cases, the data from familiar song sessions achieved the lowest performance. The empirical results of emotion recognition suggested that unfamiliar musical excerpts might be the most appropriate materials to induce emotion in music-emotion recognition experiment. In addition, experiencing unfamiliar musical stimuli would also eliminate the factors of
Figure 2.5: Arousal and valence classification accuracies above the chance levels for high familiarity (familiar songs), low familiarity (unfamiliar songs), and combined (all songs) data groups from the conducted experiment (a) and from the DEAP dataset (b).

expectation and predictability that have been reported to reduce emotional response to music [105]. In summary, this study has demonstrated that classifying emotion using typical algorithms can gain benefit from controlling the familiarity level of the subject to musical stimuli. In particular, using EEG responses elicited solely by unfamiliar stimuli might be appropriate and could help achieve higher classification accuracy.

2.1.3 Features and Window Size

A wide range of algorithms has been proposed to extract informative features from the EEG signals and can be broadly categorized into three domains: time domain, frequency domain, and time-frequency domain [34]. To gain more insights on the appropriate features to recognize emotion, FD, PSD, and discrete wavelet transform (DWT) based on Daubechies order 4 (db4) wavelet function were used to represent features from time, frequency and time-frequency domains respectively. Hereby, spectral power ($P_j = \frac{1}{N} \sum_{k=1}^{N} (d_j(k))^2$) was derived as DWT feature from each frequency band where $j$ is the level of wavelet decomposition, $k$ is the number of wavelet coefficients varying from 1 to $N$. The non-overlapping sliding window was identically applied to every
domain. However, it can be noticed that the size of sliding window could affect characteristics of the features extracted from non-stationary EEG signals. Accordingly, sliding window size was varied in the range from 1 to 8 seconds to investigate the effect of window size. SVM and C4.5 algorithms were used to build classifiers and the grand-averaged accuracies above the corresponding chance levels across subjects are presented in Figure 2.6 [97]

In comparison among three feature domains, performance of classification with FD features was superior to those with the others, while PSD features achieved the worst performance, and the results were partly in line with a previous work [110]. A plausible underlying reason was that time and time-frequency domain features could cope with the non-stationary characteristic of EEG signal better than frequency domain features. In addition, it is observable that the size of sliding window affected classification performance with time and time-frequency domain features; using smaller window size led to slightly better results in classifying FD and DWT features. In contrast, the size of sliding window did not have any notable influence on the classification performance with PSD features. The hypothesis was that time is an essential factor in time-related domain features, therefore varying window size could directly affect the characteristics of the obtained features. Empirical results also suggested that performance could be higher
when using a sliding window that was not larger than 4 seconds, which is corresponding with a previous report on optimal window size, 3–6 seconds, for DWT features [15].

2.1.4 Application of State-of-the-art machine learning techniques

Deep learning has gained greater attention in machine learning research [49]. The technique of deep belief networks (DBN), a deep learning approach, was adopted to detect emotional state from EEG signals [35, 116] but in static fashion where the variation of emotion was overlooked. This motivated another study of applying deep learning for EEG-based emotion recognition in dynamic manner [97].

DBN is an efficient unsupervised deep-learning algorithm that could overcome the complexity of training deep generative model [30]. DBN is a hierarchical structure of multiple layers of restricted Boltzmann machines (RBM). The process of training DBNs involves training each individual RBM one after another and then stacking them on top of each other. The outputs of the hidden nodes at layer $l-1$ is used as input data for training the next RBM at layer $l$. In this study, DBN with 2 hidden layers and a softmax classifier attached to the top was also applied. In training a DBN, firstly, RBMs were trained by greedy layer-wise unsupervised learning algorithm [6] and then unsupervised fine-tuning of all layers of DBN was performed with backpropagation on the unlabeled data. The resultant pre-trained DBN was then used to initialize the weights for supervised fine-tuning with backpropagation. In this study, linearly decaying learning rate and momentum were used. To prevent models from overfitting, 10 percents of instances in the dataset were selected randomly and used as a validation set and the remained instances were used to evaluate the performance using 10-fold cross-validation. DBN was implemented with DBNToolbox [114].

The results were integrated into Figure 2.6. In general, the results reflected the superior capability of DBNs to learn regardless of feature domain or classifier. The results confirmed that deep learning approach could improve EEG-based emotion recognition in dynamic manner as well. The great advantage of DBN is the learning ability to capture high-level feature representation from lower-level features in each consecutive DBN level. Therefore, DBN could be considered an interesting and promising approach for recognizing emotion using EEG.
2.1.5 Annotation Smoothing

The annotation of felt emotion plays an important role in the construction of emotion recognition system but also confronts with a variety of challenges \cite{61, 59} including tool unfamiliarity or limitation of tool leading to the contamination of noise in emotion annotation data \cite{61}. To address the problem, an early study of this research proposes a novel approach to manipulate annotation data by introducing the application of filtering techniques \cite{96, 98} to smooth annotation curve and alleviate the adverse effect of annotation noise. The resultant annotation data was expected to be more naturalistic and better matching with the intrinsic emotion than the original annotation data obtained by the mouse-clicking annotation that relies on the assumption that emotion of subject is temporarily stable if the subject does not click.

Three types of commonly-used filters in signal processing were applied: (1) Moving average filter takes the average of multiple points within a sliding frame; (2) Savitzky-Golay filter fits a low-degree polynomial to data points in a frame and then assesses the resulting polynomial at a single point within the approximation interval; it also does not greatly distort the signal \cite{86}; (3) Median filter takes the median within a sliding frame and has the advantage in preserving edges.

Emotion recognition was then performed by using classification and regression techniques using SVM and SVM regression (Gaussian radial basis function (RBF) kernel) respectively. For classification problem, the performance was measured by accuracies and MCC that will be explained in Section 3.5. For regression problem, the emotional labels were obtained by averaging values of the reported numerical arousal and valence within a sliding window, and the performance was measured by Pearson correlation and mean squared error (MSE) between the estimated curves and the smoothed annotation curves.

Figure 2.7 illustrates the exemplified resultant annotation curves from applying filtering technique with small, medium, and large filter frame size (501, 4001, and 8001 instances respectively). Applying larger filters resulted in smoother curves of annotation, where trends of curve were still preserved but high-frequency fluctuation was reduced. Nonetheless, larger filters diminished the intensity of annotation data, especially when applying moving average filter. The results of emotion recognition in subject-dependent approach using 10-fold cross-validation and subject-independent approach using leave-one-subject-out cross-validation are illustrated in Figure 2.8. In general, using large filter
could enhance the performance of emotion recognition by both using classification and regression techniques, and in both cross-validation approaches. Although the moving average filter achieved the highest performance and also has compelling less computational complexity, this filter should be used with adequate precaution concerning the intensity distortion of the resultant annotation curves.

2.1.6 Incorporating information from musical features

Although emotional responses to music stimuli can be subjective, there might exist similar patterns of responses for some extents. In particular, the emotion that is expressed in music might elicit common patterns of felt emotion among population. By this way, the information from musical stimulus itself could be utilized for predicting emotional states of listener, especially a new-coming user. This motivated another early study of this research attempting to fuse information from musicals contents with EEG signals with the aim to enhance the performance of emotion recognition system [99].

FD features served as representative features of EEG modality. High-level musical features, whose relevances to music-emotion were validated by a previous study [55], were extracted by using MIRtoolbox v.1.6.1 [48]; A dynamic feature of a song was derived from the frame-based root mean square (RMS) of the amplitude; In addition, the frame-based tempo estimation, the attack times and slopes of the onsets were extracted to represent rhythm of the song; Timbre reflects the spectro-temporal characteristics
of sound and can be characterized by spectral roughness, 13 Mel-frequency cepstral coefficients (MFCC) and their 1st derivatives, frame-decomposed zero-crossing rate, the low energy rate, and the frame-decomposed spectral flux; Furthermore, tonal features, including the key clarity, mode, and the harmonic change detection function (HCDF), were extracted to characterize tonal pattern of the song. Afterward, two modalities were fused at decision level. As two modalities were independent, linear fusion could be applied. For binary classification, let $p_{\text{EEG}}^x$ and $p_{\text{music}}^x \in [0, 1]$ denote the classifier outputs of EEG and music modality respectively for class $x \in \{1, 2\}$, the output class probability, namely $p_{\text{multimodal}}^x$, for class $x$ is then given by

$$p_{\text{multimodal}}^x = \alpha p_{\text{EEG}}^x + (1 - \alpha)p_{\text{music}}^x,$$

(2.1.4)
where $\alpha$ is the weighting factor that satisfies $0 \leq \alpha \leq 1$ and determines how EEG modality contributes to the final decision. Features were fed to SVM classifiers (RBF kernel) to estimate emotions in subject-dependent and -independents manners.

Figure 2.9 displays the results suggesting that the performance decreased when increasing the contribution of EEG modality (i.e., increasing $\alpha$ from 0 to 1), distinctively noticeably in subject-independent arousal classification. This suggested that musical features could be more stable information source to estimate music-emotions while EEG modality might suffer from inter-subject variability, which had dramatically adverse impact on arousal classification. Nevertheless, the emotion recognition system cannot always rely solely on information from musical stimuli given the hypothesis that emotional responses to music are subjective. Notwithstanding, the empirical results suggested that incorporating musical modality might address the cold start problem that system might encounter when starting to predict emotion of a new-coming subject with very limited data. The system can later turn to utilize EEG information after collecting sufficient training data.

### 2.2 Experimental Protocol

To design the experiment for data collection to construct emotion recognition models using wearable sensors, the insights learned from the above early works of this research and literature review on designing music-emotion experiment [24] were taken into account. In particular, the balance of the induced emotion was carefully controlled expecting more balanced dataset to ease model training and evaluation.
2.2.1 Subject Recruiting and Questionnaires

Experimental data were collected from 30 healthy subjects (25 males, 5 females). Averaged age was 28.44 years (SD 3.83, range 22.14-37.89 years). All subjects had minimal formal music education, and most of them were graduate students at Osaka University. Informed consent was obtained from all individual subject included in the experiment. Monetary rewards were given to the subjects requiring their participation. Prior to the experiment, a brief introduction was given and subjects were encouraged to ask questions when any clarification was needed and they were allowed to terminate the experiment whenever feeling uncomfortable to continue.

It was reported that emotional processing of left-handers is localized on the opposite cortical side to right-handers suggesting that handedness could affect the affective processing in the brain. In particular, the motivation-related alpha-power asymmetry pattern of right-handers was found reverse in left-handers [11]. To ensure the synchrony of EEG signals and investigate further on this phenomenon, handedness of each subject was assessed with the Edinburgh Handedness Inventory [69] (Table A.1).

As pre-experiment mood has been shown to affect emotional judgment of musical experience [109], subjects took the shortened version of the profile of mood states (POMS) [89] prior to the experiment to ensure that subjects were not in the severely turbulent mood. The questionnaire is shown in Table A.2. Furthermore, as self-referential emotional judgment could also be influenced by the level of self-emotion awareness, subjects were also instructed to complete the emotion awareness questionnaire (EAQ) [78] (Table A.3) to allow assessing the reliability of the annotation data. To facilitate questionnaire answering, all of the three questionnaires were carefully translated into Thai and Japanese with partly referencing to the previous questionnaire validation studies.

2.2.2 Musical Stimuli

Table 2.1 shows all musical stimuli that were used in the experiment, and their detailed information on song title, artist, and genre are shown in Table A.4. All songs were derived from music collection featured in previous studies in music-emotion research and therefore were already labeled by statistically verified emotional ratings. In this research, the ratings were simply classified into 4 emotion classes in total: Q1, Q2, Q3, and Q4 corresponding to 4 quadrants of the arousal-valence emotion model. Each song was selected based on its potential to successfully elicit emotion of listeners and
Table 2.1: Musical stimuli used in the experiment

<table>
<thead>
<tr>
<th>Song ID</th>
<th>Block</th>
<th>Type</th>
<th>Expected Emotion</th>
<th>ID in DB</th>
<th>Cut Start (m:ss)</th>
<th>Cut Stop (m:ss)</th>
<th>Length (m:ss)</th>
<th>Sample Start (m:ss)</th>
<th>Bit Rate (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>P</td>
<td>MIDI</td>
<td>A+V+</td>
<td>[107] G04</td>
<td>0.00</td>
<td>0.10</td>
<td>0.10</td>
<td>-</td>
<td>128</td>
</tr>
<tr>
<td>P2</td>
<td>P</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[107] A02</td>
<td>0.00</td>
<td>0.10</td>
<td>0.10</td>
<td>-</td>
<td>128</td>
</tr>
<tr>
<td>P3</td>
<td>P</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[107] T03</td>
<td>0.00</td>
<td>0.10</td>
<td>0.10</td>
<td>-</td>
<td>128</td>
</tr>
<tr>
<td>P4</td>
<td>P</td>
<td>MIDI</td>
<td>A+V-</td>
<td>[107] P02</td>
<td>0.00</td>
<td>0.10</td>
<td>0.10</td>
<td>-</td>
<td>128</td>
</tr>
<tr>
<td>001</td>
<td>1</td>
<td>WAV</td>
<td>A+V+</td>
<td>[91]</td>
<td>343</td>
<td>0.25</td>
<td>1.10</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>002</td>
<td>1</td>
<td>WAV</td>
<td>A-V+</td>
<td>[91]</td>
<td>360</td>
<td>2.09</td>
<td>2.54</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>003</td>
<td>1</td>
<td>WAV</td>
<td>A-V-</td>
<td>[91]</td>
<td>488</td>
<td>4.12</td>
<td>4.57</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>004</td>
<td>1</td>
<td>WAV</td>
<td>A+V-</td>
<td>[91]</td>
<td>631</td>
<td>0.06</td>
<td>0.51</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>005</td>
<td>2</td>
<td>WAV</td>
<td>A-V+</td>
<td>[91]</td>
<td>960</td>
<td>0.52</td>
<td>1.37</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>006</td>
<td>2</td>
<td>WAV</td>
<td>A-V+</td>
<td>[91]</td>
<td>891</td>
<td>0.27</td>
<td>1.12</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>007</td>
<td>2</td>
<td>WAV</td>
<td>A-V-</td>
<td>[91]</td>
<td>739</td>
<td>0.42</td>
<td>1.27</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>008</td>
<td>2</td>
<td>WAV</td>
<td>A-V-</td>
<td>[91]</td>
<td>702</td>
<td>0.28</td>
<td>1.13</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>009</td>
<td>3</td>
<td>MIDI</td>
<td>A+V+</td>
<td>[7]</td>
<td>10</td>
<td>7.56</td>
<td>8.41</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>010</td>
<td>3</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[7]</td>
<td>6</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>011</td>
<td>3</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[7]</td>
<td>9</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>012</td>
<td>3</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[7]</td>
<td>17</td>
<td>0.38</td>
<td>1.23</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>013</td>
<td>4</td>
<td>MIDI</td>
<td>A+V+</td>
<td>[88]</td>
<td>-</td>
<td>0.00</td>
<td>1.30</td>
<td>1.30</td>
<td>-</td>
</tr>
<tr>
<td>014</td>
<td>4</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[88]</td>
<td>-</td>
<td>0.00</td>
<td>1.30</td>
<td>1.30</td>
<td>-</td>
</tr>
<tr>
<td>015</td>
<td>4</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[88]</td>
<td>-</td>
<td>0.12</td>
<td>1.42</td>
<td>1.30</td>
<td>-</td>
</tr>
<tr>
<td>016</td>
<td>4</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[25]</td>
<td>-</td>
<td>0.00</td>
<td>1.30</td>
<td>1.30</td>
<td>-</td>
</tr>
<tr>
<td>017</td>
<td>5/6</td>
<td>WAV</td>
<td>A-V+</td>
<td>[91]</td>
<td>635</td>
<td>1.05</td>
<td>1.50</td>
<td>0.45</td>
<td>0:11.5</td>
</tr>
<tr>
<td>018</td>
<td>5/6</td>
<td>WAV</td>
<td>A-V+</td>
<td>[108]</td>
<td>H4</td>
<td>0.10</td>
<td>0.55</td>
<td>0.45</td>
<td>0:01.0</td>
</tr>
<tr>
<td>019</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[94]</td>
<td>87</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>0:16.6</td>
</tr>
<tr>
<td>020</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[87]</td>
<td>-</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>0:00.0</td>
</tr>
<tr>
<td>021</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[91]</td>
<td>647</td>
<td>0.23</td>
<td>1.08</td>
<td>0.45</td>
<td>0:00.6</td>
</tr>
<tr>
<td>022</td>
<td>5/6</td>
<td>WAV</td>
<td>A-V+</td>
<td>[108]</td>
<td>T2</td>
<td>1.01</td>
<td>1.46</td>
<td>0.45</td>
<td>0:00.4</td>
</tr>
<tr>
<td>023</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[94]</td>
<td>21</td>
<td>0.02</td>
<td>0.47</td>
<td>0.45</td>
<td>0:00.4</td>
</tr>
<tr>
<td>024</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[87]</td>
<td>-</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>0:04.5</td>
</tr>
<tr>
<td>025</td>
<td>5/6</td>
<td>WAV</td>
<td>A-V-</td>
<td>[91]</td>
<td>197</td>
<td>0.11</td>
<td>0.56</td>
<td>0.45</td>
<td>0:11.1</td>
</tr>
<tr>
<td>026</td>
<td>5/6</td>
<td>WAV</td>
<td>A-V-</td>
<td>[108]</td>
<td>S2</td>
<td>0.00</td>
<td>0.23</td>
<td>0.45</td>
<td>0:04.1</td>
</tr>
<tr>
<td>027</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[94]</td>
<td>160</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>0:25.8</td>
</tr>
<tr>
<td>028</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[87]</td>
<td>-</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>0:11.7</td>
</tr>
<tr>
<td>029</td>
<td>5/6</td>
<td>WAV</td>
<td>A-V-</td>
<td>[91]</td>
<td>456</td>
<td>0.02</td>
<td>0.47</td>
<td>0.45</td>
<td>0:08.8</td>
</tr>
<tr>
<td>030</td>
<td>5/6</td>
<td>WAV</td>
<td>A-V+</td>
<td>[108]</td>
<td>F2</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>0:17.8</td>
</tr>
<tr>
<td>031</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[94]</td>
<td>190</td>
<td>0.27</td>
<td>1.12</td>
<td>0.45</td>
<td>0:11.5</td>
</tr>
<tr>
<td>032</td>
<td>5/6</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[87]</td>
<td>-</td>
<td>9:46</td>
<td>10:31</td>
<td>0.45</td>
<td>0:05.6</td>
</tr>
<tr>
<td>033</td>
<td>7</td>
<td>WAV</td>
<td>A-V+</td>
<td>[91]</td>
<td>477</td>
<td>0.00</td>
<td>1.46</td>
<td>1.46</td>
<td>-</td>
</tr>
<tr>
<td>034</td>
<td>7</td>
<td>MIDI</td>
<td>A-V-</td>
<td>[83]</td>
<td>-</td>
<td>0.56</td>
<td>2:11</td>
<td>1:15</td>
<td>-</td>
</tr>
<tr>
<td>035</td>
<td>7</td>
<td>WAV</td>
<td>A-V+</td>
<td>[84]</td>
<td>-</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>-</td>
</tr>
<tr>
<td>036</td>
<td>7</td>
<td>WAV</td>
<td>A-V-</td>
<td>[91]</td>
<td>158</td>
<td>0.00</td>
<td>1.46</td>
<td>1.46</td>
<td>-</td>
</tr>
<tr>
<td>037</td>
<td>7</td>
<td>MIDI</td>
<td>A-V+</td>
<td>[83]</td>
<td>-</td>
<td>0.28</td>
<td>1:43</td>
<td>1:15</td>
<td>-</td>
</tr>
<tr>
<td>038</td>
<td>7</td>
<td>WAV</td>
<td>A-V-</td>
<td>[84]</td>
<td>-</td>
<td>0.00</td>
<td>0.45</td>
<td>0.45</td>
<td>-</td>
</tr>
</tbody>
</table>
its possibility to be less familiar for the subjects since it has been shown that it benefits
the model if subjects have a limited familiarity to the stimuli [100, 102], whereas the
familiarity level was judged by the opinion of the experimenter.

The 1000 Songs dataset [91] provides 45-second 1000 excerpts each of which was
annotated by 10 subjects at minimum using arousal and valence ratings. The dataset
has been actively used in music-emotion research. The chosen songs in this research
were selected by their potential to elicit extreme emotion which was indicated by having
the relatively high summation of squared mean arousal, squared mean valence, and low
standard deviation across annotators in each quadrant of the arousal-valence space. The
work [107] provided a dataset containing 56 songs in happy (equivalent to Q1), scary
(equivalent to Q2), sad (equivalent to Q3), and peaceful (equivalent to Q4) emotions.
The responses in arousal and valence from 39 listeners substantially matched the in-
tended emotion of music composer. The songs to be used in this research were selected
using the same strategy as in selecting songs from the 1000 Songs dataset. In [7], 27
classical musical excerpts were studied to investigate the influence of song duration on
emotional response in musician and non-musician groups, and the results suggested that
merely 1 second of music contained sufficient cue for inferring emotional reaction. The
songs in this research were selected from the ones that could elicit emotion successfully
both in musician and non-musician groups who annotated emotions in arousal and va-
lence space in the experiment. In [88], annotation software for continuously tracking
emotional response to music in arousal-valence space was developed and 67 subjects
were recruited to annotate 4 long classical music excerpts that could expectedly elicit
emotions in 4 quadrants of the arousal-valence model. As the main purpose of [88] was
to track emotional variation over the course of time, the songs in this research tended
to have emotion fluctuation. Therefore, the musical excerpts derived from this dataset
were sampled with longer duration of 90 seconds, to allow the investigation of emotion
variation. The songs corresponding to Q1, Q3, and Q4 were selected from this dataset.
In addition, the song that was rated high on the dimension of arousal and low on the
dimension of valence in [25], which was also another study of emotion variation in
music, was selected as another representative long song of Q2. The work [108] studied
emotional responses to a set of songs selected from of their own musical dataset [23] to
further explore the role of individual difference variables (such as personality and mood)
in music-induced emotions. The songs could successfully induce happy (corresponding to
Q1), scary (corresponding to Q2), sad (corresponding to Q3), and tender (corresponding to Q4) emotions of 148 participating subjects. The 4 songs to be used in this research were selected by seemingly having lowest popularity. The MoodSwings Turk Dataset [94] contains 240 songs with the emotional labels in arousal and valence which were collected by the developed MoodSwings, a collaborative online game that enables crowdsourcing emotion ratings for music. Four songs with extreme emotions were selected to be used as stimuli in this research by adopting the same strategy as in selecting songs from the 1000 Songs dataset. One of the early reports of frontal alpha asymmetries includes [87] that employed 4 classical music to induce intense-pleasant (corresponding to Q1), intense-unpleasant (corresponding to Q2), calm-unpleasant (corresponding to Q3), and calm-pleasant (corresponding to Q4) emotions of 59 right-handed subjects. The pre-experiment ratings in the study were reported to be in line with the intended emotions. In this research, the same set of musical stimuli was used to allow the study of the validation of the putative frontal alpha asymmetry. In addition, musical stimuli that were used in the previous discovery of chilling effect [83] and unpleasantness effect [84] were also utilized in an additional block of experiment (block 7 in Table 2.1) to allow the investigation of these effects.

After completing the questionnaires, subjects proceeded into the music-stimuli selection phase of the experiment. The musical selection database was comprised of a balanced proportion of excerpts in MIDI and WAV format. Almost all of the musical excerpts had the duration of 45 seconds that fell into the range of commonly-used music length in music-emotion research [24], except for the 4 practice songs (10 seconds) and 4 aforementioned long songs (90 seconds). The length and the number of songs were designed by following the guideline in a literature review on designing music-emotion experiment [24].

In the process of assembling the set of music stimuli to be analyzed by each subject, the findings from [24] which suggest that in order to maximize the efficacy of music elicitation, each set of musical stimuli must be comprised of a mix of subject-selected songs and researcher-selected songs. Moreover, each set should maintain a good balance of samples from each emotional quadrant. Adopting a stratified music-selection approach, the 8 chosen songs would comprise a set of songs that subject selected from block 5/6 in Table 2.1. In song selection, subjects were presented with a set of 4 songs from each emotion class. From each set, they selected the top 2 songs that they felt
Figure 2.10: Experimental Protocol; each song in one block has different expressing emotion varying from high (Hi) to low (Lo) arousal (A) and from positive (Pos) to negative (Neg) valence (V).

the songs would be most likely to elicit the assigned emotion. Together with the 16 researcher-selected songs (block 1-4 in Table 2.1), the resultant set of musical stimuli featured 24 songs to be used for the music-listening phase. The musical excerpts in each block would be of similar duration and would include a representative sample from each emotion class. Another block comprising of 6 songs was also included for possible future work (block 7 in Table 2.1).

2.2.3 Music-listening Task

After assembling the set of stimuli, subjects then listened to and annotated each song. In particular, after 4 short songs that were used at the beginning of each session as part of the practice session, each subject proceeded into task session which was to listen to 6 main blocks and 1 additional block of music. To aid discussion, it can be referred to the act of listening and annotating a single song as one trial (see Figure 2.10 bottom row). Each trial begins with a resting period (R0) lasting 10 seconds meant to minimize the lingering effects of the previous song. This is followed by an emotionally-neutral white-noise listening period (WN) lasting 5 seconds, followed by a relaxation period (R1) lasting another 5 seconds. Next, a song is played back (PL) to the subject for either 45 or 90 seconds depending on song length. This is followed by another 5-second relaxation period (R2). To minimize EEG artifacts, subjects were requested to close their eyes and minimize body movement while still remaining comfortable for the entire interval between WN and R2. Afterwards, subjects would proceed into the continuous annotation period (A1).

During A1, subjects were presented with the same song for the second time, this time
to annotate the emotions they felt during the preceding PL. Continuous annotations were provided through a user interface by marking milestones of emotion variations by clicking on the arousal-valence space presented on the screen as the music unfolded over time. Finally, subjects provided another set of annotations (A2) which provided an overall assessment of each song. The first was the overall arousal and valence (on a Likert scale of 1–9) using the Self-Assessment Manikin [10], followed by a rating of their familiarity to the musical stimuli (on a Likert scale of 1–5 ranging from “never heard before” to “listen every day”). Subjects also rated their confidence in the annotations provided in A1 (on a scale of 1–3), and whether or not they liked the song. Data collection was performed on the own in-house software coded in Java; Figure 2.11 shows a screenshot during performing emotion annotation. As emotional label smoothing was reported to successfully overcome the limitation of mouse-clicking emotion annotation [96, 98], moving average filtering technique was thus applied to the acquired numerical arousal and valence ratings, where the filter size was set as ≈ 4 seconds (1025 sample points).

### 2.2.4 Brain and Physiological Recording

Brainwave and physiological signals were recorded (Figure 2.12) throughout the sessions by wearable sensors. A wearable EEG headset with active sensor nodes and miniaturized electronics was used to capture brainwaves, an ECG patch was used for recording heart activity, and a wrist-worn GSR band was used to track skin responses to musical stimuli. All wearable sensors were developed by imec [26], streamed data to PC via Bluetooth...
communication, and were used to record signals at a sampling frequency of 256 Hz using a synchronization software also provided by imec. The specification of wearable devices is summarized in Table 2.2. Original EEG signals from 8 electrodes without applying filter available in the software were used in this research. Only one channel of ECG signals selected by the experimenter was utilized. Similarly, one proper channel of GSR signals was used. It is noteworthy that EEG electrodes were located nearby the frontal lobe which has been demonstrated to play a key role in emotion regulation [41]. Although the impedance of each electrode was not controlled throughout the experiment, the quality of signals was closely monitored by the experimenter and the headset was carefully re-adjusted whenever necessary. In most of the cases, setting up each sensor could be finished within 1 minute demonstrating the practicability of wearable sensors, although some subjects spent slightly longer time in fitting up the EEG headset owing to the unique shapes of their heads.

2.3 Summary

In summary, the early exploratory studies of this research have paved the way toward the better design of experiment for data collection. First, emotion recognition should take into account the emotion variation over the course of time, and this can be done by sliding windowing technique, while smaller window is preferable. Second, empirical results suggested that familiarity to musical stimuli affected the recorded EEG signals, and using solely unfamiliar music is highly recommended. Next, time-related domain
features would be more stable compared to frequency domain features and this evidence suggested the necessary for frequency-domain features to use more robust power-spectral estimation, especially when having short-period data extracted within a sliding window. Deep learning was found achieving slightly better performance than SVM but introduced a new challenge in the increase of computational cost. Furthermore, manipulating annotation data would be a necessary step prior to using them as the labels for training classifiers or estimators, and smoothing annotation by moving-average filter would be a compelling candidate of solutions. Finally, emotion-related musical features can be partly used to estimate emotion of a newly-coming subject yet with precaution owing to the fact that emotion response to music is subjective.

Based on the insights learned from these studies together with literature reviews, the new experimental protocol was designed. Musical stimuli were drawn from a number of previous studies, whereas each song already had emotional label either from the population of annotators or from experimenters in those studies (the success of emotion induction was reported). The number, duration, and balance of musical stimuli were suitably controlled. In one trial of the experiment, each subject listened to white noise, took a brief rest, listened to music, took another rest, annotated the feeling by continuously clicking on arousal-valence space while music unfolded, and rated overall music. While performing the music-listening task, wearable brain and physiological sensors were used to record bodily responses of the subject; EEG headset recorded cortical activities nearby scalp; ECG patch recorded heart activity; wrist-worn GSR sensor monitored the change in skin conductance. The setting-up time for these devices was impressing short (in general, less than 5 minutes) indicating the potential of wearable sensor usage in daily-life basis.
Table 2.2: Wearable device specification

<table>
<thead>
<tr>
<th>Devices</th>
<th>Properties</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEG</td>
<td>Components</td>
<td>Single headset</td>
</tr>
<tr>
<td></td>
<td>Sensor position</td>
<td>Fp1, Fp2, F3, F4, F7, F8, T3, T4 in 10-10 system</td>
</tr>
<tr>
<td></td>
<td>Ground position</td>
<td>AFz in 10-10 system</td>
</tr>
<tr>
<td></td>
<td>Reference position</td>
<td>FCz in 10-10 system</td>
</tr>
</tbody>
</table>
|         | Channels | 8 raw EEG signals  
9 channels  
8 filtered EEG signals  
8 I-impedance signals  
8 Q-impedance signals |
|         | Sampling frequency | 256 Hz |
|         | Data transmission | Bluetooth |
|         | User interface | Nyx-legacy 1.54 x86 software running on MS Windows 10 |
|         | Output format | HDF5, CSV |
|         | Battery life | 5 hours |
|         | User fitting | By adjusting velcro straps |
|         | Device controlling method | On/off button |
|         | ECG Components | Sensor and disposable patch |
|         | Channels | 2 ECG channels  
3D accelerometer  
5 channels  
256 Hz for ECG channels, 32 Hz for other channels |
|         | Sampling frequency | 256 Hz for ECG channels, 32 Hz for other channels |
|         | Data transmission | Bluetooth |
|         | Data storage | Streaming to PC, SD card |
|         | User interface | Nyx-legacy 1.54 x86 software running on MS Windows 10 |
|         | Output format | HDF5, CSV |
|         | Battery life | 7 days |
|         | User fitting | Strapping patch on user’s chest |
|         | Device controlling method | Tapping |
|         | GSR Components | Single wristband |
|         | Channels | 1 GSR low-amplitude channel (0.01 µS to 1 µS) for dry skins  
1 GSR high-amplitude channel (1 µS to 20 µS) for hydrated skins  
3D accelerometer  
Body temperature (Celsius) |
|         | Sampling frequency | 256 Hz for GSR channels, 32 Hz for other channels |
|         | Data transmission | Bluetooth |
|         | Data storage | Streaming to PC, SD card |
|         | User interface | Nyx-legacy 1.54 x86 software running on MS Windows 10 |
|         | Output format | HDF5, CSV |
|         | Battery life | 7 days |
|         | User fitting | Wearing on user’s left hand |
|         | Device controlling method | Tapping |
Chapter 3

Emotion Recognition based on Electroencephalography

It is widely accepted that CNS responses are closer to emotional process in the brain than ANS responses and could thus provide better emotional cues [27, 66]. Correspondingly, the research is in general an extension of emotion recognition study based on EEG signals, with the paradigm shift toward wearable sensors and the fusion with another physiological modality. To aid discussion, this chapter is devoted to the presentation of emotion recognition study using merely data from wearable EEG sensor.

3.1 Questionnaires Results

The results of the questionnaire of handedness using the Edinburgh Handedness Inventory [69] are shown in the Figure 3.1. It is apparent that most of the subjects are right-handers; 24 of them reported right-hand orientation, while 5 reported left-hand orientation and one uses left hand as often as the right.

POMS standardized test reveals the state of mood of the subject before experiment as the extents of tension, depression, anger, fatigue, confusion, and vigor. Following the calculation of scores based on the standard protocol [89], the scores were then normalized within subject as the percents of total moods. The statistics of the normalized values (mean, median, minimum, and maximum) are shown in Figure 3.2. The scores indicated that subjects were generally in mildly positive mood prior to the experiment and had low levels of negative moods resulting in the low total mood disturbance (TMD) scores. It has been demonstrated that by having high vigor mood, happiness could be strongly included [109], which might benefit the experiment when considering the success of happiness elicitation using music.
Figure 3.1: Results of Edinburgh Handedness Inventory

Figure 3.2: Results of Profile of Mood States (POMS); Median, 25th and 75th percentiles, minimum and maximum of the score are shown. The diamond symbol denotes mean.

With regards to EAQ results shown in Figure 3.3, all of the medians and means were higher than 0.5 suggesting that each subject possessed excellent skill to distinguish different emotions. As a result, the emotions that were reported from the subjects were adequately reliable. In addition, the score of music-familiarity, which was preferred to be low, was found extremely low; the averaged familiarity score across subjects and songs was 1.49 (SD 0.88, Median 1) on the scale of 1–5. Furthermore, subjects also reported remarkably high score of confidence in the annotation data (i.e., the extent how well the annotated emotion corresponding to the emotion that the subject actually felt in
the preceding phase); on the scale of 1–3; the averaged score across subjects and songs was 2.79 (SD 0.43, Median 3). By this evidence, the reliability of emotions reported by subjects was sufficiently guaranteed.

### 3.2 Signal Preprocessing

Acquiring high-quality EEG signals from soft dry electrodes in a real-world scenario is challenging owing to various sources of electrical, mechanical, and physiological artifacts [62, 67]. Accordingly, preprocessing the acquired EEG signals is indispensable, and the protocol used for cleaning EEG signals in this research is illustrated in Figure 3.4.

#### 3.2.1 Band-pass Filters

At the beginning, by following a standardized EEG signal preprocessing guideline that has been verified by a study on large-scale collection of dataset [8], high-pass filter at 1 Hz was applied and followed by applying notch-filter to clean power-line noise at 60 Hz (with the notch width of 2 Hz). Signals were filtered by using the function `pop_eegfiltnew` of EEGLAB [19, 21], a freely-available MATLAB toolbox for the analysis of EEG dynamics, since the function has been proven to avoid any phase distortion or delay distortion.
3.2.2 Artifact Rejection using ASR method

More recently, artifact subspace reconstruction (ASR) method [67] was proposed to detect and remove artifactual data including eye blinks and muscular activities. The ASR filter is designed to detect and remove high-amplitude data components of high amplitude relative to some artifact-free reference data while recovering EEG background activity that lies in the subspace spanned by the artifact components. The method relies on a sliding-window principal component analysis (PCA), which statistically interpolates any high-variance signal components exceeding a threshold relative to the covariance of the calibration dataset. Each affected time point of EEG is then linearly reconstructed from the retained signal subspace based on the correlation structure observed in the calibration data. In this research, the cut-off parameter for removal of bursts was set at 3 SDs which might be the aggressive but acceptable for avoiding the loss of massive EEG data. Optional techniques of channel rejection by flat-line duration, channel correlation, and line noise criteria and time window rejection were not applied.
3.2.3 Rejecting Epochs in Resting-State Data

In the experiment, despite the instruction of eye closing and body-movement minimizing during WN, R1, PL, and R2 periods, it turned out that some subjects did not return to ready posture in time after a freely-relaxing period of R0. This resulted in the contamination of artifacts especially in WN and R1 that immediately followed R0. Unlike PL, WN and R1 were supposed to be emotional-neural resting-state data, therefore the time synchronization with emotional labels can be discarded. Accordingly, removing some severely-artifactual EEG portions would be reasonable and helpful for the subsequent analysis. Therefore, the technique for rejecting artifacts in continuous data of EEGLAB (pop_rejcont function [20]) was applied to the resting-state data. First, EEG signals in WN, R1, and R2 periods were concatenated and then epoched into 0.25-second windows. The number of contiguous epochs necessary to label a region as artifactual was set at 4. By considering signals in 1-40 Hz frequency range (Hamming window was used prior to fast Fourier transform (FFT)), regions of contiguous epochs were marked as artifacts and removed whenever the spectrum deviated from baseline for 10 dB.

3.2.4 Decomposition by ICA and IC Evaluation

Afterwards, independent component analysis (ICA) was performed on EEG signals using EEGLAB and info-max algorithm was used for decomposition [37]. The technique is capable of separating independent sources whose activities are linearly mixed to generate scalp potentials. The obtained independent components (ICs) were consequently evaluated to be categorized as brain activity or non-brain artifact (eye movement artifact, muscle artifact, or unrelated noise) using the following methods.

Basic Methods

Each IC was assessed to be labeled as brain, muscle, eye, and other noise by first visual inspection on the salient characteristic PSD, locations of the equivalent current dipoles, scalp topography, and time-course activation. For instance, eye-movement ICs tend to have dipoles whose locations are near the edge of ventral frontal brain regions and their PSDs tend to be high at low frequencies with no spectral peak. Muscle artifacts can be identified by having the locations of dipoles outside the brain or near lower head region and having characteristic mean spectral plateau above 25 Hz. Brain activity tends to have PSD feature of 1/f curve occasionally with peaks in the alpha band.
Figure 3.5: Visualization of the properties of brain IC (a), eye-movement IC (b), and muscular-activity IC (c); the used EEG dataset was drawn from a previously published study on image familiarity judgment [64]. Top images show scalp topography and time-course activation (epoched into 4 seconds); Middle images display PSDs and estimated locations of the equivalent dipoles; The bottom images show the comodulograms illustrating cross-frequency power-power couplings.

MARA Toolbox

Classifying and labeling ICs have been formidable challenges in EEG research for decades. Not all ICs would have clear and typical properties in the combinations of EEG features. Recently, MARA Toolbox [112, 113], a freely-available toolbox released as a plug-in for EEGLAB, has been developed aiming to assist automatic labeling ICs. The features derived from IC’s spectrum include FitError and lambda features, which describe the deviation of component’s spectrum from a typical 1/f curve ($\lambda$ parameter and mean squared error of the approximation to the real spectrum are used as features), and the average log band power of alpha ($8-13$ Hz feature). In addition, the technique also employs the Mean Local Skewness feature derived from IC’s time series, which is defined as the mean absolute local skewness of time intervals of 1 second and 15 seconds duration. Additional features from IC’s pattern include Range Within Pattern feature, which is defined as the logarithm of the difference between the minimal and the maximal activa-
tion in a pattern, and \textit{Current Density Norm} feature, which is obtained from the density in the strongest source’s position \(x, y,\) and \(z\). MARA core model was trained by a supervised machine learning algorithm to learn on binary classification problem for classifying 1290 ICs labeled by expert raters. In practice, MARA Toolbox automatically provides the probability that the inputted ICs would be artifacts. In this research, MARA was employed as an additional tool to aid IC evaluation. However, merely MARA might not be sufficient for evaluating an ambiguous IC which needs to be assessed by more robust method.

\textbf{Cross-frequency Power-Power Coupling}

To gain more insights for evaluating IC, this research proposes a novel method for IC evaluation based on a newly-discovered characteristic pattern. Recently, researchers have found that the brain signals show various forms of cross-frequency couplings \cite{36}. In a part of this research, the cross-frequency power-power coupling (cross-frequency PPC) was found to benefit the analysis of ICA-decomposed EEG data. In particular, each class of brain, eye, and muscle has its own patterns in the comodulogram, which is a correlation plot of PPC of the combination of all pairs of frequencies. The classification was supported by physiological interpretations from their estimated source distributions. In a brain-activity IC, the cross-frequency PPC was found between 8-30 Hz (alpha to beta), and that in the off-diagonal area was well-suppressed. The comodulogram of an eye-movement IC tended to have characteristic block-diagonalization on the lower end of the frequencies in 1-8 Hz (delta to theta band) and the higher end of the frequencies in 23-50 Hz (beta to gamma band). Lastly, in the muscle-activity IC, characteristic \textit{parallel lines} were observed between 30-50 Hz (gamma band). To facilitate understanding, the exemplified comodulograms were shown in the bottom images in Figure 3.5.

To compute cross-frequency PPC, EEG data are first decomposed by ICA. On continuous IC time-series data, short-term Fourier transform (STFT) with 50%-overlapping 1-s sliding window is employed to compute spectrogram and implemented by Matlab’s \texttt{spectrogram()} function to generate semi-logarithmically spaced 1-50 Hz divided into 100 frequency bins. To clean the data, 20% outliers detectable by distances from median values were excluded. The correlation between the power spectrum \(S(f)\) at a frequency \(f_i\) and at another frequency \(f_j\) can be calculated using the expression

\[
corr_{ij} = \frac{\sum_k (S_k(f_i) - \bar{S}(f_i))(S_k(f_j) - \bar{S}(f_j))}{\sigma_i \sigma_j}
\]  

\[(3.2.1)\]
where $S_k(f_i)$ is the power spectral density at frequency $f_i$ in time-window $k$, $\overline{S(f_i)}$ the averaged power spectral density at frequency $f_i$ over all sliding windows, $\sigma_i$ the standard deviation of the power spectral density at frequency $f_i$, and $k$ ranges over all sliding windows.

Currently, a toolbox for detecting cross-frequency power-power coupling has been being collaboratively developed as a free, open-source software that can be plugged into the EEGLAB environment\(^1\). After trimming top 10% and bottom 10% from the data distribution for data cleaning, cross-frequency PPC was computed and IC was evaluated by the corresponding comodulogram.

### 3.2.5 EEG signals Reconstruction

Each individual IC was evaluated by all of the above-mentioned methods and labeled as brain non-brain ICs. Afterwards, ICs accounting for non-brain activities (eye movement artifact, muscle artifact, or unrelated noise) were removed from the IC set. Finally, the remained ICs were used to back-project to the original space to reconstruct the artifact-free EEG signals.

### 3.3 Feature Extraction

To enable capturing emotion recognition in temporal continuous fashion [90, 101], a non-sliding window technique was applied. The size of window was set as 2 seconds as this length provided satisfactory results in the previous work [97] across feature types and classification methods and was also close the optimal size (3–6 seconds) recommended by a study in the literature [15].

In the past decade, there are a plenty of feature extraction algorithms proposed for emotion recognition studies. Recently, the work [34] impressively summarized the algorithms with the aim to discuss the advantages and disadvantages of each feature. By this motivation, all of the feature extraction algorithms summarized in the work [34] were applied in this research, except for the High Order Spectra (HOS) and Hilbert-Huang Spectrum (HHS) owing to their expensive computational costs yet poor preliminary results. Generally, the features can be categorized into time domain, frequency domain, time-frequency domain, and feature calculated from the combination of electrodes, namely cross-channel features. The brief introduction to the feature extraction

\(^1\)https://sccn.ucsd.edu/wiki/Power-Power_Coupling_Analysis_Tool
algorithms is given here using the following notation: $X(i)$ denotes the vector of the time series of a single electrode where $i = 1, ..., N$ and $N$ is the number of time-samples in $X$; $X'(i)$ denotes the time derivative of $X(i)$.

### 3.3.1 Time Domain Features

**Statistics of Signals** In this research, the following statistical features were extracted.

- **Power**\(^2\): $P_X = \frac{1}{N} \sum_{i=1}^{N} |X(i)|^2$
- **Mean**: $\mu_X = \frac{1}{N} \sum_{i=1}^{N} X(i)$
- **Standard deviation**: $\sigma_X = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X(i) - \mu_X)^2}$
- **First difference**: $\delta_X = \frac{1}{N-1} \sum_{i=1}^{N-1} |X(i+1) - X(i)|$
- **Normalized first difference**: $\overline{\delta_X} = \frac{\delta_X}{\sigma_X}$
- **Second difference**: $\delta_X^2 = \frac{1}{N-2} \sum_{i=1}^{N-2} |X(i+2) - X(i)|$
- **Normalized second difference**: $\overline{\delta_X^2} = \frac{\delta_X^2}{\sigma_X}$

**Hjorth Features** In this research, the following Hjorth features were extracted.

- **Mobility**: $M_X = \sqrt{\frac{\text{var}(X'(i))}{\text{var}(X(i))}}$
- **Complexity**: $C_X = \sqrt{\frac{\text{var}(M'(i))}{\text{var}(M(i))}}$

**Non-Stationary Index** Non-Stationary Index (NSI) is a simple measure of the consistency of the local average values, independent of the fluctuation magnitude of the original time series. Higher NSI values indicate more inconsistent local averages. After normalization with mean and variance, the normalized time series is then divided into $K$ segments, and the local average is computed in each segment. The NSI is then obtained by calculating the SD of these $K$ means \([28]\). In this research, $K$ was set to 32.

**Fractal Dimension** FD values characterize the complexity of time-varying signals and the method is commonly used in affective computing owing to its simplicity yet excellent results \([56, 93]\). Higher FD values of EEG signals reflect the higher activity of the brain. In this research, the Higuchi algorithm \([29]\) was applied to calculate FD values.
Table 3.1: Frequency Band Ranges and Decomposition Levels of EEG signals Recorded at a sample frequency of 256 Hz

<table>
<thead>
<tr>
<th>Bandwidth (Hz)</th>
<th>Frequency Band</th>
<th>Decomposition Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–4 Hz</td>
<td>Delta (Unused)</td>
<td>A5</td>
</tr>
<tr>
<td>4–8 Hz</td>
<td>Theta</td>
<td>D5 (4–8 Hz)</td>
</tr>
<tr>
<td>8–12 Hz</td>
<td>Alpha</td>
<td>D4 (8–16 Hz)</td>
</tr>
<tr>
<td>12–30 Hz</td>
<td>Beta</td>
<td>D3 (16–32 Hz)</td>
</tr>
<tr>
<td>30–40 Hz</td>
<td>Gamma</td>
<td>D2 (32–64 Hz)</td>
</tr>
</tbody>
</table>

Higher Order Crossings  Recently, higher order crossings (HOC) has been adopted to estimate emotional states using EEG data [72]. The technique indicates the extent of the fluctuation of filtered EEG signals; The number of filters was set as 10 in this work.

3.3.2 Frequency Domain Features

Power Spectral Density  Over the last few decades, the PSD method has been the most prominent approach to reveal affective states from brainwaves [39]. PSD indicates signal power in specific frequency ranges. This method is basically based on FFT converting data from time domain into frequency domain and vice versa. Recently, multi-taper PSD was proposed to estimate PSD with the advantages of having minimal bias (which is a critical issue for the state-of-the-art PSD estimation techniques including Periodogram and Welch’s method) and robustness under stochasticity. Although it is not widely used, it has been proven to have many important optimality properties and is appropriate to estimate short-period PSD for EEG signals [4] as it does not assume the stationarity of the signal for the duration of a trial. Accordingly, in this research, PSD was estimated by using multi-taper PSD implemented by an open-source Chronux toolbox [63]. The taper bandwidth was set to 5 Hz in each 1-second sliding window, and the number of tapers was set to 9. Afterwards, PSD features, namely PSD-band features, were extracted from the obtained PSD by averaging power spectra within specific frequency bands: theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–40 Hz) bands. Alternative PSD features were extracted by computing the averaged power within small equal-sized bins of 2 Hz ranging from 4 Hz to 40 Hz to extract 18 of so-called PSD-bin features.

3.3.3 Time-Frequency Domain Features
**Discrete Wavelet Transform**  In time-frequency domain, discrete wavelet transform (DWT) is a recently-proposed feature extraction technique that decomposes signals into different approximation and detail levels corresponding to particular frequency ranges while conserving the time information of the signal [68]. As the sampling frequency of EEG recording in this research was 256 Hz, time-frequency features were extracted via db4-based DWT by first extracting the components A5, D5, D4, D3 and D2, where the correspondence of frequency bands and wavelet decomposition levels is shown in Table 3.1. Then, spectral power, root mean square (RMS), recursive energy efficiently (REE), Log-REE, and absolute Log-REE features were calculated to construct a feature set.

### 3.3.4 Cross-Channel Features

**Magnitude Squared Coherence Estimate**  The magnitude squared coherence estimate (MSCE) indicates how well two signals $X(i)$ and $X(j)$ correspond to each other at each frequency and is calculated from cross PSD and PSDs of the two signals. The MSCE features were computed from all combination of electrodes mounted on EEG headset. To reduce a large amount of features, $C_{ij}$ was averaged over the frequency bands of theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–40 Hz).

**Lateral Asymmetry**  Correspondence between the asymmetries of brain activity between the left and right hemispheres and the emotional responses were occasionally reported in the literature popularizing the use of bilateral asymmetric indexes to classify emotions [43, 54, 55, 93]. Likewise, this research employed this technique by constructing features from the differences in power bands of corresponding pairs of electrodes, namely differential asymmetry features ($\text{Diff}$):

$$
\text{Diff}_{\text{band}} = PSD_{\text{band}}^l - PSD_{\text{band}}^r
$$

(3.3.1)

where $l$ and $r$ are symmetric pairs of electrodes on the left/right hemisphere of the scalp. In this research, 4 $\text{Diff}$ features were drawn from PSD-band features of Fp1-Fp2, F3-F4, F7-F8, and T3-T4 pairs separately in each frequency band. Similar to $\text{Diff}$ features, the rational asymmetry feature ($\text{Ratio}$) can be obtained by:

$$
\text{Ratio}_{\text{band}} = \frac{PSD_{\text{band}}^l}{PSD_{\text{band}}^r}
$$

(3.3.2)
3.4 Feature Adjustment by Subtracting the Baseline

The brain signals recorded from EEG are the summation of a number of brain processes. In music listening, the brain can involve in sound processing, emotion processing, memory usage, and even motor movement as a result of enjoyment. Although some of the unrelated signals, e.g. muscular activities, can be removed by ICA, the remaining signals might still be contaminated by the activities from certain mechanisms. Especially, sound processing activities can inevitably mix with signals from the emotional process to generate the resultant EEG signals. Therefore, this overlapping necessitates the removal or suppression of sound processing activity from the acquired EEG data. To suppress sound processing activities, the extracted features were subtracted by the baseline, which was defined as the feature extracted from the period of emotionally-neutral white-noise listening (WN). In special case that WN data was missing due to epoch removal during signal preprocessing, either data from R1 or WN in preceding song session was used instead.

To demonstrate the usefulness of feature adjustment by baseline subtraction, the correlations between features and the annotation data (numerical values in the range [-1,1]) in arousal and valence space were calculated. Figure 3.6 and 3.7 display the comparison of the correlations before and after adjusting the PSD features (as a representative of frequency domain) and time-domain features respectively. By statistical \( t \)-test, the correlation in frequency-domain features extracted from certain electrodes was found

![Figure 3.6: Correlation of PSD-band features and arousal/valence ratings comparing between features before (a) and after adjustment by subtracting the baseline (b); White cross signs and plus signs in (b) denote the significant differences at \( p < 0.05 \) and \( p < 0.01 \) respectively by paired \( t \)-test](image)
Figure 3.7: Correlation of time domain features and arousal(A)/valence(V) ratings comparing between features before (a) and after adjustment by subtracting the baseline (b); The notation is the same as in Figure 3.6

significantly increased as a result of adjustment. Similarly, the features in time domain were found more correlated with the annotation data in general after subtracting the baseline. Interestingly, some features turned to have correlation in the opposite direction after adjustment. Higher correlation infers the potential of being good features for
emotion classification. This evidence, therefore, suggests the importance of subtracting
the extracted features with their baselines, which was a usually-overlooked technique
in most of the previous studies. To clarify, the difference and ratio in the PSD Diff
and Ratio features were adjusted after feature extraction. In other words, The adjusted
PSD Diff feature can be obtained by:

\[ \text{Diff}_{\text{adjusted}} = (P_{\text{Diff}}^{l} - P_{\text{Diff}}^{r}) - (P_{\text{Diff}}^{l}_{\text{calibration}} - P_{\text{Diff}}^{r}_{\text{calibration}}), \]  

(3.4.1)

where \( P_{\text{calibration}} \) is the feature of EEG during WN period. Similarly, the adjusted
PSD Ratio feature can be obtained by:

\[ \text{Ratio}_{\text{adjusted}} = \frac{P_{\text{Diff}}^{l}}{P_{\text{Diff}}^{r}} - \frac{P_{\text{Diff}}^{l}_{\text{calibration}}}{P_{\text{Diff}}^{r}_{\text{calibration}}}. \]  

(3.4.2)

### 3.5 Emotion Classification

Emotion recognition was converted into binary classification of arousal and valence sep-
arately for the sake of simplicity. Arousal ratings were quantized into 2 classes of high
and low, and valence ratings were quantized into 2 classes of positive and negative.
When assigning emotion labels to features, a majority method was applied to a particu-
lar window containing the variation in emotional class. Afterwards, all features for each
individual subject were normalized into the range of 0 and 1 using min-max approach.

For emotion classification, SVM based on RBF kernel was adopted and implemented
using the open-source LibSVM [16]. To evaluate the performance of emotion classifi-
cation for each subject, a leave-one-block-out (LOBO) cross-validation was adopted
given the expectation that each block contained balanced samples in each emotional
class. Henceforth, the results reflect the performance of subject-dependent emotion-
recognition. In addition to classification accuracy, the Matthews Correlation Coefficient
(MCC) was also used as it is considered as an accurate metric of performance of a
classifier with unbalanced classes [60]. The maximal coefficient +1 represents a perfect
classification (100% accuracy) and the minimal coefficient -1 represents total disagree-
ment (0% accuracy). Interestingly, the coefficient 0 indicates that the classification is
one-class random guessing. Given a confusion matrix of binary classification, MCC can
be calculated by

\[ MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \]  

(3.5.1)
Figure 3.8: Valence (a) and arousal (b) classification accuracies comparing between using features before and after adjustment by subtracting the baseline; The corresponding MCCs for features before and after adjustment are shown as the bottom and top values respectively over the SD bars. Two-star mark denotes the accuracy is significantly higher than the chance level at $p < 0.01$ by paired t-test and one-star mark represents the significance at $p < 0.05$.

where $TP$, $TN$, $FP$, and $FN$ are the numbers of true positives, true negatives, false positives, and false negatives respectively. It is noteworthy that the optimal SVM hyper-parameters ($\gamma$ and $C$) were obtained by grid searching in the range $2^{-5}, 2^{-4}, ..., 2^{5}$ based on the classification accuracy, and merely the best results are presented.

### 3.6 Emotion Classification Result

The valence and arousal classification accuracies, together with MCCs, are shown in Figure 3.8 with the main comparison between the performance of features before and after adjustment by subtracting the baseline. As can be seen, all of the EEG features before the adjustment could achieve significantly higher performance than the chance level in valence classification but merely some of them can repeat the same significant success in arousal classification. However, after the adjustment, all of the features can be significantly superior to the chance level ($p < 0.01$) suggesting the importance of feature adjustment. Especially, frequency-domain features gained the maximal benefits from feature adjustment; this might be from the fact that the PSD values are very sensitive and could dramatically fluctuate over time. This principle was preliminarily confirmed by inspecting the plot of resting-state PSD.

The cross-electrode PSD $Ratio$ and $Diff$ features after adjustment achieved the highest accuracies in valence recognition of 68.04% (SD 6.08%) and 68.01% (6.08%)
respectively. The frequency domain features were prominent in recognizing valence, while the FD features performed best in time domain. In arousal classification, the best result (67.95%, SD 5.46%) was achieved by classifying PSD features, while the Hjorth features from time domain also achieved decent performance (67.44%, SD 4.62%).

3.7 Enhancing Performance by Feature Selection

Feature selection plays a crucial role in machine learning classification problems. In EEG domain, only some of the all features are capable of capturing task-specific information while the other features might be redundant or irrelevant to the task. Feature selection technique principally allows the model to select only informative features that could be useful for classification and remove unimportant ones leading to the computational efficacy.

In this research, ReliefF algorithm [80] was exploited to extract the most informative features for each subject’s model. The usefulness of this feature selection method has been demonstrated in previous affective computing works [34, 53]. To overcome high sensitivity to noise, ReliefF uses \( n = 5 \) nearest hits/misses. Multiple classes are accounted for by searching for the \( n \) nearest instances from each class weighted by their prior probabilities.

The assigned number of remained features after selection was varied in the range \( F \times k/8 \) where \( k = 1, ..., 8 \) and \( F \) is the total number of features, to elucidate the feature selection technique. At a fixed \( k \), features were chosen for each subject and emotion classification was performed on this remained features. The results are illustrated in Figure 3.9. As can be seen, feature selection cannot substantially improve the results when applying it with the feature type whose total number of features was relatively low (typically less than 32). In particular, the accuracies and MCCS of PSD-Band, Hjorth, NSI, FD, PSD-Band-Diff and PSD-Band-Ratio became lower when reducing the size of feature set. However, the performance of using other feature types did not drop dramatically when the number of selected features decreased. In some cases, feature selection could increase the accuracies of emotion classification.

However, EEG signals can vary dramatically by subjects or even the different day of recording [52]. It might be reasonable to speculate that the optimal number of informative features might not be identical across all subjects. Therefore, assigning the same \( k \) for remaining features for all subjects might not necessarily improve the accuracies but
Figure 3.9: Emotion classification accuracies when varying the number of remaining features (selected by feature selection technique) separated by the used features; The corresponding MCCs for valence and arousal classification are shown as top (blue) and bottom (red) texts respectively over the graph.

instead degrade the performance. To allow flexibility, the feature selection method was thus modified by allowing the model of each subject set its own $k$. Figure 3.10 shows the results of permitting this flexibility. When compared to the results of fixing $k$ as shown in Figure 3.9, this flexibility remarkably improved the emotion classification results in general. The enhancement of performance was found significant by the paired $t$-test.

On the other hand, this results offered another evidence suggesting the inter-subject variability in EEG signals.
Figure 3.10: Valence (a) and arousal (b) classification accuracies comparing between using all feature set and applying feature selection; The results were obtained by finding the optimal number of remaining features in each subject. The corresponding MCCs for method using all feature set and applying feature selection are shown as the bottom and top values respectively over the SD bars. The star denote that feature selection method significantly outperforms the method of using all feature set at $p < 0.01$ by paired $t$-test.

### 3.8 Summary

Signal preprocessing is indeed a vital process. In this research, typical band-pass and notch filters were applied to the acquired EEG signals. Then, the recently-proposed ASR technique was applied to reconstruct the artifact-disrupted signals. However, the most widely-accepted technique, which has been verified by a huge amount for works, to clean EEG signals is to decompose the original scalp EEG signals into source domain using ICA technique and subsequently remove the components contributing noises or artifactual signals to the summed EEG from the source space before back-projecting into scalp EEG domain. However, one of the main challenges for this technique is to identify the class of each component among brain signal, eye artifact, and muscle artifact classes. This research proposes an alternative approach to classify the component by visualizing the cross-frequency power-power coupling as each class of components was found to feature different patterns of such coupling. With this method, artifactual IC can be accurately identified and removed from source space, and the quality of the back-projected EEG signals can be significantly improved.
After acquiring cleaner EEG signals, informative features were extracted using verified algorithms appearing in the literature. Normalization with the white noise listening session, which was designated to be the baseline, was found to increase the correlation between extracted features and annotation data suggesting that the adjustment might cancel the signature of brain’s sound processing which might conceal the emotion-related information. The empirical results on subject-dependent emotion recognition suggested that the performance of using the features after adjustment was magnificently better than those before adjustment. The best accuracies above corresponding chance levels were 7.62% for valence classification and 6.18% for arousal classification. Exploiting merely the most informative features and ignoring less-relevant or redundant features offered benefits in potentially increasing classification accuracies and reducing computational cost. Accordingly, the most commonly-used feature selection technique (ReliefF) was applied in this research. The results indicated that fixing the number of remaining features for all subjects might prevent the benefit of the feature selection technique owing to the issue of inter-subject variability. Accordingly, by searching the optimal number of remaining features for each subject, the performance of emotion recognition was remarkably improved. The accuracy above chance level for valence classification was increased to 9.19% and 8.15% for arousal classification.

Despite challenges in wearable EEG sensors, the better performance over one-class random guessing proves the potentials of using wearable sensors to recognize emotion effectively in more realistic scenarios. It is noteworthy that the accuracies could be further improved by applying alternative state-of-the-art classification techniques in machine learning which is extremely active research field in the recent years.
Chapter 4

Multimodal Adaptive Emotion Recognition

Recent efforts to increase accuracies of emotion recognition include multimodality approach, a technique of integrating information from different sources to improve the robustness of the system. However, utilizing higher number of information source is facing with challenge in exhaustively consuming longer time for device setting up. Thanks to the enormous growth in the technology of wearable sensing, recording signals at minimal set-up time has become possible and could enable the creation of highly practical emotion recognition system which could be integrated into daily life. However, combining information from multiple modalities also introduces another challenge. It is not necessary that each of modality should contribute to the final decision equally as some might be disrupted by noise. Directly dropping portions of bad signals in one modality might have adverse effects on the synchronization with another modality. On the other hand, blindly combining information sources from every modality might degrade the performance of the system rather than enhance it. To address this problematic issue, this research proposes a novel framework for multimodal fusion. Instead of discarding bad signals, the context-aware framework regulates the relative contribution of a particular modality based on its stability in decision fusion. The details of this method are presented in this chapter.

4.1 Data Acquisition and Signal Preprocessing

The EEG features were the same as those presented in Chapter 3. To obtain cleaner ECG signals, a band-pass filter of the frequency range 0.5–40 Hz, implemented by infinite impulse response Butterworth filter, was applied to mitigate the adverse effects of
low-frequency trend and power-line noise. Consequently, ECG and GSR features were extracted.

4.2 Feature Extraction

**ECG Features** Again, 2-second non-overlapping sliding window was applied to extract ECG and GSR features. The set of 22 ECG features was selected by partly following previous works [79, 74] and comprised of:

- Averaged heart rate (HR); The R-R interval, where the R-peak was detected by MATLAB’s *findpeaks* function, was calculated to estimate HR,
- Difference between the averaged HR in that window and the averaged HR in R1 period,
- Zero-crossing rate,
- First 4 statistical moments,
- Normalized length density (NLD),
- Non-stationary index (NSI); The same NSI algorithm for EEG time-domain feature extraction was employed,
- Spectral entropy,
- 12 averaged power spectra within frequency bin of 2 Hz in the range 3–27 Hz.

**GSR Features** Partly following a previous work [43], the set of 13 GSR features was comprised of:

- 10 averaged PSD within the frequency bin of 0.25 Hz in the range of 0–2.5 Hz,
- averaged GSR$^N$,
- average of first derivative$^N$,
- proportion of negative samples in the first derivatives and total sample$^N$,

where $^N$ denotes that the features were normalized by subtracting with the features extracted from R1 period in order to remove the effect of global trend of GSR signal (very low frequency).
4.3 Signal Quality Quantification

In this research, the contribution of each modality is dynamically varied based on the instability of each modality. First, time-series of instability metrics were calculated for each modality, and then a sliding window technique was applied to acquire the instability indexes that were synchronized with EEG, ECG, and GSR features.

**EEG**

To quantify the degree of instability of EEG signals, the artifact detection algorithm of EEGLAB [37] was adopted. The data points were analyzed at the granularity of 0.5 seconds (128 points) referred as one epoch. All data points inside an epoch can be marked as containing artifacts whenever the signals of at least one electrodes were classified as artifacts by an instability metric. In this research, 6 different metrics were used.

- Abnormal trend: this first metric was calculated by EEGLAB’s *pop_rejtrend* function, which is used to detect abnormal trend. The method fits signal to a straight line and marks data points as artifacts if the slope exceeds a linear regression *R*-square value threshold (set as 0.3).

- Abnormal joint probability: this metric was calculated by *pop_jointprob* function that computes probability of occurrence of each epoch and marks data points as artifacts if probability of the epoch exceeds a threshold (set as 3 SD) of deviation from the mean probability distribution calculated from all trials.

- Abnormal distribution by kurtosis: this additional metric was calculated by *pop_rejkurt* function based on the assumption that artifactual data epochs (e.g. discontinuity in signals) can have very peaky activity value distributions that can be detected by measuring the kurtosis of probability distribution.

- Abnormal spectra: 3 additional metrics were calculated by *pop_rejcont* function, which detects abnormal spectra, using 3 different thresholds. The method marks data points as artifacts if the power spectra exceed a threshold of deviation from the global mean over all trials; The thresholds were set as 6, 7, and 8 dB.

Figure 4.1 illustrates the data epochs that were marked as containing artifacts by each metric, where the exemplified signals were drawn from a 150-second data of Subject 53.
11. As can be seen, the noisy signals occurred intermittently and can be noticed by the increase of amplitudes of those noises (marked by red lines). Some metrics were found capable of detecting this noisy period.

**ECG and GSR**

Both wearable ECG patch and GSR band recorded sensors’ accelerometer data. Since subjects were instructed to minimize body movement between the WN and R2 periods, movements could be detected by calculating deviations from the median value of each accelerometer channel. By assuming that signal quality might drop dramatically during movement, it might be reasonable to reduce the contributions of movement-laden modalities, whereas the movement can be detected in accelerometer data. For each accelerometer channel (corresponding to x, y and z motion), signals from each modality can be marked as containing artifacts when the accelerometer data exceeds a threshold (set as 0.05 SD) of deviation from the median of each trial. As a result, 3 instability metrics for the ECG and GSR modalities were obtained.
Stability in a sliding window

The same sliding window used in EEG, ECG, and GSR feature extraction were once again applied to quantify the instability in each window. A preliminary observation on data suggested that it was unlikely that the periods containing artifacts would last longer than the window size (2 seconds). In other words, subjects would not commit huge movement continuously for longer than 2 seconds. To allow capturing intermittent movements, the threshold used to mark a window as unstable $\sigma$ should be shorter than the window size. Given an instability metric of a modality, let $mD$ be the number of data points marked as unstable, the instability index is defined ($\text{instab}_m^k$) as

$$\text{instab}_m^k = \begin{cases} 1 & mD \geq \sigma \\ \frac{mD}{\sigma} & mD < \sigma \end{cases}$$

(4.3.1)

where $m \in \{1, ..., M\}$ is the set of modalities (i.e., EEG, ECG, or GSR) and $k \in \{1, ..., K\}$ is the set of metrics for each modality. The $\sigma$ was deliberately set as 256 data points (equivalent to 1 second); if $mD$ exceeds this threshold, the $\text{instab}_m^k$ is set to 1, otherwise it is calculated as the ratio between $mD$ and $\sigma$.

4.4 Decision-level Multimodal Fusion

Since classification performance was previously assessed and optimized for each modality, multimodal fusion can be performed using the optimal unimodal models. The final classification can then be decided via decision-level fusion. Majority voting (MAJ), a popular decision-level multimodal fusion technique for binary classification [3], was used as the baseline method. In this research, an alternative method incorporating the instability index was proposed to dynamically regulate the contribution of each modality based on their assessed reliability. Given $M$ modalities, each of which has $K$ stability metrics, the fusion weighting parameter $w_m$ for each modality $m$ is defined as

$$w_m = \frac{\left(\frac{1}{K} \sum_{k=1}^{K} (1 - \text{instab}_m^k)\right)\theta_m}{\sum_{m=1}^{M} \left(\frac{1}{K} \sum_{i=1}^{K} (1 - \text{instab}_m^i)\right)\theta_m}$$

(4.4.1)

where $\theta_m$ acts as a penalty parameter for each modality. A larger $\theta_m$ more heavily penalizes lower values of stability $0 \leq (1 - \text{instab}_m^k) < 1$, resulting in a much lower weighting parameter $w_m$. The sum of $w_m$ over all modalities is then normalized to 1. In the rare case that all modalities are totally unstable (i.e., the denominator in equation 4.4.1 is 0), $w_m$ of a modality is defined as $\frac{1}{M}$. 

55
For binary classification, given $p_m \in \{0, 1\}$ is an output from a modality $m$, the multimodal fusion output is generated as

$$p_{\text{fusion}} = \sum_{m=1}^{M} w_m \cdot p_m. \quad (4.4.2)$$

The final output is set to 1 when $p_{\text{fusion}}$ is at least 0.5, and 0 otherwise. It can be noticed that if every modality is stable (i.e. $\text{instab}_m^k$ is 0 for any $m$ and $k$), each modality will contribute to the final decision equally and the output would be the same as from majority voting.

To assess the consequence of introducing the penalty parameter, models were tested using different values of $\theta_m$. Specifically, $\theta_m$ for each modality $m$ was varied from 1 to 10 and only the best result was presented. This method is referred, in the succeeding sections, as adaptive method with varied penalty (AVP). Another variation of this approach, wherein $\theta_m$ was fixed to 1 for every modality, is also presented and aptly referred as adaptive method with fixed penalty (AFP).

Recently, a technique of incorporating performance of inner cross-validation on training set into decision-level fusion was proposed and found to be capable of improving the accuracy of bimodal emotion recognition [44]. This research includes an extension of this method to 3 modalities by performing inner 5-fold cross-validation on training set separately for each modality $m$ to obtain accuracy $acc_m$, which is later used to define the weighting parameter $t_m = \frac{acc_m}{\sum_{m=1}^{M} acc_m}$, and then applying it to the decision-level fusion as

$$p_{\text{fusion}} = \sum_{m=1}^{M} w_m \cdot p_m \cdot t_m. \quad (4.4.3)$$

where the same criteria (threshold as 0.5) was used to generate the final output of classification. This technique is referred as adaptive method with varied penalty and inner cross-validation (AVPC). In order to demonstrate that the improvement of the classification performance is not mainly from incorporating the weighting parameter $t_m$ but from including both $t_m$ and $w_m$, another method of solely integrating the $t_m$ into the decision-level fusion, namely adaptive method with fixed penalty and inner cross-validation (AFC), is defined to allow comparison. Table 4.1 summarizes all of the above-mentioned settings.
Table 4.1: Summary of unimodal and multimodal methods for emotion classification

<table>
<thead>
<tr>
<th>Modality</th>
<th>Method</th>
<th>Abbrev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unimodal</td>
<td>EEG</td>
<td>EEG</td>
</tr>
<tr>
<td>Unimodal</td>
<td>ECG</td>
<td>ECG</td>
</tr>
<tr>
<td>Unimodal</td>
<td>GSR</td>
<td>GSR</td>
</tr>
<tr>
<td>Multimodal</td>
<td>Majority Voting</td>
<td>MAJ</td>
</tr>
<tr>
<td>Multimodal</td>
<td>Adaptive, Fixing Modality Degree ($\theta_m = 1$)</td>
<td>AFP</td>
</tr>
<tr>
<td>Multimodal</td>
<td>Adaptive, Varying Modality Degree ($\theta_m = [1, \ldots, 10]$)</td>
<td>AVP</td>
</tr>
<tr>
<td>Multimodal</td>
<td>Adaptive, Varying Modality Degree ($\theta_m = [1, \ldots, 10]$), Incorporating inner cross-validation ($t_m$)</td>
<td>AVPC</td>
</tr>
<tr>
<td>Multimodal</td>
<td>Adaptive, Incorporating inner cross-validation ($t_m$)</td>
<td>AFC</td>
</tr>
</tbody>
</table>

4.5 Results

The emotion classification accuracies and MCCs are shown in Figure 4.2. In all of the cases in valence classification, the proposed adaptive methods (AFP, AVP, and AVPC) could outperform unimodal classification or conventional multimodal fusion. Likewise, the performance of arousal classification can be enhanced when using the most of the proposed adaptive stability-sensitive approaches. In particular, the AVPC achieved the best results measured by accuracies and MCCs on both valence and arousal classification. Notably, EEG was found achieving the best results in valence classification compared with the other unimodality. In addition, GSR unimodality generally achieved better arousal classification performance compared with the other unimodality, and this is corresponding with the fact that GSR has been popularly employed to recognize arousal since the first era of affective computing [74].

To gain more insights on the effect of feature extraction techniques, the performance of EEG unimodality with applying the flexible feature selection technique and that of AVPC technique were illustrated in Figure 4.3 for direct comparison. According to the results, every feature type of emotion recognition gained benefit from multimodal fusion using the context-aware adaptive technique. The PSD-Ratio features achieved highest valence classification accuracies (71.24%, SD 5.14%) and the time-domain Hjorth features accomplished the best performance in the arousal classification (70.84%, SD 5.41).
Figure 4.2: Emotion recognition accuracies of each unimodality and each method of multimodal fusion; The corresponding MCCs are shown over the SD bars. A starred pair represents significantly difference using multiple comparison with Bonferroni method at $p < 0.01$.

Figure 4.3: Valence (a) and arousal (b) classification accuracies comparing between EEG unimodality (after adjustment without feature selection) and multimodality (AVPC); The corresponding MCCs for EEG unimodality are multimodality are shown as the bottom and top values respectively over the SD bars. Two-star mark denotes the accuracy is significantly higher than the chance level at $p < 0.01$ by paired $t$-test.
4.6 Summary

Multimodal fusion technique that integrates information from different sources does not always guarantee the improvement of classification. In particular, if a single modality introduces noise or unreliable data, blindly fusing information from that modality might degrade the system rather than improve it. Henceforth, the reliability of each modality at particular time should be evaluated and used to determine the reliability before integrating its information into decision making. To answer the question which modality to rely on, this research proposes making use of the instability index calculated from input signal as a metric to indicate the reliability of each modality and use for regulating decision fusion. EEG instability was derived by adopting artifact detection techniques, and ECG and GSR instability indexes were drawn from the mounted accelerometers. By considering the context of stability of each modality, information from each source can be integrated effectively.

Empirical results suggested that the proposed method could enhance the performance of emotion recognition in a subject-dependent manner. The best accuracy over chance level for valence classification was 10.82% which was significantly higher than the conventional majority-voting method and any of unimodality method. Similarity, arousal classification could achieve 9.23% accuracy above chance level and significantly outperformed unimodality and majority-voting approaches. The excellent results suggested the promise of the adaptive approach and call for the attention to consider modality stability during multimodal fusion.
Chapter 5

Toward Subject-independent Emotion Recognition using Transfer Learning

The individual difference in EEG signals across subjects poses a grand challenge for building an effective generalized brain-computer interface (BCI) for decades including the emotion recognition system. The inter-subject variability makes the model that is successfully-trained to estimate emotion for a specific subject fails to recognize emotion of another subject. In other words, emotion recognition model tends to be subject-dependent and has limited performance in testing in subject-independent fashion. Accordingly, to create a new model for a specific subject, there exists the requirement to collect a large number of training data with emotional labels, which is a time-consuming process and reducing the practicability of the system.

Recently, transfer learning has drawn increasing attention as a successful approach to alleviate this issue. Transfer learning technique allows transferring knowledge from source domain to target domain with the scarcity of training sample available [70]. Current BCI research employs transfer learning to transfer data, classifiers, hyper-parameters, or another information acquired from source subjects to a target subject who is usually a new-coming user. Using the various techniques of domain adaptation, an effective model can be constructed and adjusted to match the new subject at the computationally effective cost with minimal prerequisite to collect the training data from the subject. Hence, the success of this technique infers higher practicability of the system in overall. This chapter presents a study on the application of transfer learning in EEG-based emotion recognition, and a new approach that allows the system rely on fewer amount of source subjects is proposed.
5.1 Methodologies

To successfully apply transfer learning method to enhance the subject-specific emotion-classification model by reducing the amount of labeled data imperative for each individual, there are several issues [70] that the framework needs to deal with including (1) what to transfer; and (2) how to transfer. This section describes the methodology for the proposed transfer learning framework with the answers to these questions.

5.1.1 Acquiring Classifiers

In this research, a leave-one-subject-out cross-validation was adopted to evaluate the performance in subject-independent fashion. In each trial, data from 29 subjects were used to build an emotion classifier for a left-out subject and the performance across trials was averaged to obtain the overall result.

SVM

An intuitive and straightforward approach is to train the classifier on the aggregated data from the population of subjects and then make a prediction of emotion on the unseen data acquired from a new subject. This generic classifier is apparently computational ineffective as it relies on the massive amount of training data. In addition, a plenty of works reported its poor performance owing to the issue of inter-subject variability in EEG data [43, 117]. To allow comparison, this method was also adopted as a benchmarking approach in this research. Hereby, SVMs based on linear and RBF kernel were implemented; the cost parameter (C) was fixed as 1 for both types of kernels.

Transductive Parameter Transfer

Recently, transfer learning has been successfully demonstrated in solving problems in BCI studies including EEG-based emotion recognition [117, 53]. Among the attempts, transductive parameter transfer (TPT) approach was reported to achieve the highest accuracy [117] and hence it was used as a representative transfer learning technique in this research. TPT [85] is a technique that learns a regression function that maps the relationship between feature distribution and the parameters of multiple individual classifiers that are trained for individual source subjects and then uses the regression model to construct a new classifier to a target subject. The technique consists of three main steps that are explained here to answer the question, how to transfer.
First, multiple individual classifiers are learned on each source dataset $D_i$. In this research, SVM with linear kernel was used as a classifier, where the hyperplane in the feature space is defined as $\theta = [w_i, b_i]$; it can be estimated by solving the optimization problem:

$$\min_{w, b} \| w \|^2 + \lambda L \sum_{j=1}^{n_i} l(w' x_j^s + b, y_j^s)$$  \hspace{1cm} (5.1.1)$$

where $l(\cdot)$ is the hinge loss. In the implementation in this research, the best model, which was obtained by searching the optimal cost parameter ($C$) for linear kernel in the range $[2^{-5}, 2^{-4}, ..., 2^{15}]$, was kept for each individual subject.

Second, a regression function $f$ is trained to learn the relation between the source data distributions $P_i$ and the hyperplane parameters of the optimal corresponding SVM: $f : \varphi \rightarrow \Theta$. Since the data distribution and the hyperplanes are relevant, the hyperplane of the target classifier can be obtained via the learned mapping function without label information of the target subject. As it is reasonable to assume that the elements in $\theta$ are correlated (i.e., a feature in EEG signals might correlate with another feature), the vector-valued regression approach can be applied; the Multioutput Support Vector Regression (MSVR) framework [103] is adopted in this research to create distribution-to-classifier mapping. In computation, a kernel $\kappa(X_i, X_j)$ that represents the similarity between pairs of datasets $X_i$ and $X_j$ is necessary and the density estimation (DE) kernel [9] was used in this research, where the C cost parameter to calculate the kernel was the averaged value of the optimal C in training each individual classifier in the first step.

Finally, the parameter vector of the target classifier can be predicted by $\theta_t = f(X^t)$ without any label information from target subjects. Given the target features $x$ and the classifier parameters $\theta$, the label can be predicted by the decision function: $y = sign(w^t x + b_t)$. In this research, the open-source TPT library contributed by [85] was adopted. Hereby, merely the best results, which were obtained by grid-searching of cost parameter (C) of regression in the range $[2^{-5}, 2^{-4}, ..., 2^{15}]$ and $\epsilon$ parameter of the loss function of MSVR in the ranges $[2^{-15}, 2^{-4}, ..., 2^5]$, are shown in order to demonstrate the potentials of the approach.

5.1.2 Assessing the Extent of Similarity

Although transfer learning has been proposed, the previous works mainly overlooked the subject transferability in data. Specifically, data from every single subject might
not necessarily be appropriate for inclusion into transfer learning to build a classifier for a new subject, and this kind of data would even have adverse impact on the performance. It might be reasonable to hypothesize that there might exist certain common neurophysiological responses in people having the same or similar emotions, accordingly the improvement of classification performance obtained by transfer learning should be positively correlated with the extent of similarity of the subjects being included into transfer learning. Based on this principle, a conditional transfer learning framework was recently proposed [53]; The framework measures the similarity between subjects by calculating the Pearson’s correlation coefficient between the target’s EEG feature space and that of each source subject and then combines the data from the most similar source subjects with the data of the target subject to train a Gaussian Naïve Bayes classifier with careful consideration of class-imbalance problem by discarding some data. Although the concept of subject transferability was impressively taken into account and the results were found promising, it is noticeable that the labels of training data from the target user are still necessary for the process of alleviating class-imbalance issue and training a new classifier, and this requirement of the labeled data collected from a new subject can be a great barrier for the practical system as the subject still needs to perform emotion-elicitation task to provide emotional labels for the framework.

To address this issue, this research proposes a new transfer learning framework focusing on an alternative approach for finding similar subjects and building classifier for a new subject without the necessity to access label information of the new-coming target subject. To find similar source subjects whose data could be transferable to the target subject, the extent of similarity can be calculated by the following methods based on two emerging important questions: what to be used and how to measure?

**The Data Used for Measuring Similarity**

To demonstrate and answer the question what to be used, the PSD features after baseline adjustment from source subjects and target subject were used. To further specify which (data) to be used, there were two options with the different advantages and disadvantages.

**Resting-state Data** Resting-state EEG data can be used to find similar subjects based on the assumption that there might be certain EEG signatures that are partially common among subjects during the resting state. Making use of resting-state EEG
data for compensating the individual difference has been proposed in BCI [65]. One merit of using merely the resting-state features is to yield practical emotion recognition system that requires minimal recording of calibration data; By recording short resting-state EEG signals, a new classifier can be constructed and emotion can be subsequently predicted within reasonably short time. In this research, PSD extracted from WN period were used as resting state data.

**Emotional-task Features** It is also reasonable to hypothesize that neurophysiological responses to emotion can differ by subjects owing to a number of factors including handedness, experiment engagement, musical expertise, or gender. In particular, the extent of the increase or decrease of EEG power of one subject might be different from that of another subject. Based on this assumption, estimating the extent of similarity between subjects should be done directly on the features obtained from the session in which subjects were performing the emotion-elicitation tasks (note that the labels were not accessed). However, using emotional-task features can be considered as less practical technique, compared to using resting-state data, as the system needs to wait until the end of data collection to start measuring the similarity posing the difficulty in enabling a real-time emotion recognition system.

**Distance Matrix Calculation**

To respond to the question *how to measure*, the difference of PSDs between two subjects can be quantified using either of the following methods.

**Root-mean-squared distance (RMSD)** After normalizing each features across subjects using *min-max* strategy, the root-mean-squared distance (RMSD) can be calculated by:

\[
RMSD = \sqrt{ \frac{1}{k} \sum_{k=1}^{F} (\xi_k^i - \xi_k^j)^2 },
\]

(5.1.2)

where \( k = 1, ..., F \) is the number of feature, \( \xi_k^i \) is the mean of feature \( k \) of subject \( i \). It is noticeable that this RMSD is equivalent to \( L2 \)-distance. The feature distances between all combination of pairs of subjects are calculated to generate a distance matrix, and the most similar subjects to a particular target subject can be identified by using this matrix.
**Clustering**  In this research, feature clustering was also used to measure the subject similarity based on the hypothesis that similar subjects might generate alike EEG patterns and the resultant PSD features might be positioned closely to each other in the feature space. After feature normalization, *k*-mean clustering technique was applied. Throughout the experiment, the number of clusters (*k*) was arbitrarily defined as 5 by referring to a preliminary study of fine-tuning the total distance to centroids and the consequent capability of finding similar subjects. From the obtained clustering result, the subject similarity index (*SIM*(i, j)) between subject *i* and *j* is defined as:

\[
SIM(i, j) = \sum_{c=1}^{k} N^i_c \times N^j_c
\]  

(5.1.3)

where *c* = 1, ..., *k* is the number of clusters and *N^i_c* is the total number of instances of subject *i* that are grouped into the *c*-th cluster. The intuition is that if a source subject and the target subject have high number of instances that fall into a certain same cluster, two subjects would be more similar and the similarity index should be increased. By calculating the similarity of all combination of subjects, the matrix of similarity is obtained. Then, similarity matrix is normalized into the range [0 1] by the minimum and maximum of the matrix. In addition, the self-similarity is enforced to be 1 afterward. Finally, the inverse of similarity is calculated by subtracting the latest *SIM*(i, j) with one in order to obtain the distance matrix for the sake of comparison with RMSD method.

The distance matrices calculated by RMSD or clustering method using resting-state features or emotional-task features are illustrated in Figure 5.1. The most similar subject was one who had the smallest distance from the target subject in the distance matrix.

**5.1.3 Experimental settings**

To demonstrate the efficacy of the proposed method, emotion recognition was evaluated using the leave-one-subject-out validation with applying all of the combination of the methods described above. In the first setting, the subject transferability was not taken into consideration, and SVM (linear kernel and RBF kernel) and transfer learning (TL) were used to build a classifier. In the next setting, with subject transferability consideration, data from 5 most similar subjects were included to construct a new classifier for the target subject using SVM and transfer learning method. The distance was measured by either RMSD or clustering method. The data for measuring the distance were either
resting-state data or emotional-task features. Table 5.1 summarizes all of the experimental settings for subject-independent emotion classification experiment. It should be noted that feature selection technique was not applied.

5.2 Results

The results of emotion classification are displayed in Figure 5.2 suggesting that making use of emotional-task features to find similar subjects prior to constructing emotion classifier achieved the highest performance of both arousal and valence classification. Especially, the transfer learning method performed best in arousal classification and would provide the second best result in valence classification. Making use of resting-state data to find similar subjects achieved poor performance. Compared to SVM classifiers, the SVM constructed by transfer learning method is computationally less expensive as the classifier is trained on each individual source subject hence utilizing much fewer training data; this evidence suggested the potentials of transfer learning in the real-world usage. In general, the results suggested that aggregating all available subjects
Table 5.1: Summary of methods used for subject-independent emotion classification.

<table>
<thead>
<tr>
<th>Methods</th>
<th>No. of used subjects for training</th>
<th>Classifiers Used</th>
<th>Data Features</th>
<th>Distance Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM-Linear-All</td>
<td>29</td>
<td>SVM-Linear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-RBF-All</td>
<td>29</td>
<td>SVM-RBF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL-All</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-Linear-SMLR-REST-RMSD</td>
<td>5</td>
<td>SVM-Linear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-RBF-SMLR-REST-RMSD</td>
<td>5</td>
<td>SVM-RBF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL-SMLR-REST-RMSD</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-Linear-SMLR-REST-CLST</td>
<td>5</td>
<td>SVM-Linear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-RBF-SMLR-REST-CLST</td>
<td>5</td>
<td>SVM-RBF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL-SMLR-REST-CLST</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-Linear-SMLR-FEAT-RMSD</td>
<td>5</td>
<td>SVM-Linear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-RBF-SMLR-FEAT-RMSD</td>
<td>5</td>
<td>SVM-RBF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL-SMLR-FEAT-RMSD</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-Linear-SMLR-FEAT-CLST</td>
<td>5</td>
<td>SVM-Linear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-RBF-SMLR-FEAT-CLST</td>
<td>5</td>
<td>SVM-RBF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL-SMLR-FEAT-CLST</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.2: Emotion classification accuracies of subject-independent classification using each method; The corresponding MCCs are shown over the SD bars.
together did not necessarily yield good performance due to the inter-subject variability, and selectively adopting data from similar subjects who had alike emotional-task features could enhance the performance of emotion classification.

To gain more insights on the applicability of transfer learning, the enumerated performance on each individual target subject using transfer learning to learn from the 5 most similar subjects (measured by clustering method on emotional-task features) is illustrated in Figure 5.3. Although the majority of subjects could gain benefit from...
Transfer learning, the performance for valence and arousal classification were lower than the chance level in 9 and 10 subjects respectively. In some unsuccessful cases, the corresponding classes were found imbalanced rendering the difficulty for performance to surpass the chance level. Another plausible reason suggested by a previous work [53] is that the subjects might have difficulty in spontaneously engaging the emotion-elicitation experiment, hence the recorded EEG dynamics of those subjects were less likely to be representative and informative about their emotional responses resulting in negative transfer. Further investigation on the underlying cause for unsuccessful transfer learning is awaited. In addition, future studies on alternative methods for identifying similar subjects and on the investigation of the optimal number of transferable subjects are also encouraged.

5.3 Summary

Transfer learning, a machine learning technique allowing the transfer of information from another subject, has been proposed to alleviate the inter-subject variability issue that poses the requirement of long recording of calibration data to make a classifying model for a new-coming subject. Recently, the principle of subject-transferability that determines the appropriateness of the information from a source subject for inclusion into transfer learning for a target subject was introduced to the BCI research community. However, the existing work relied on the correlation between features and labels to compute the transferability while accessing labels information necessitates full labeling which is the most exhaustive process in machine learning and limits practicability of the system. To solve this issue, this research proposes a new methodology to measure the subject-transferability without accessing the label information by calculating the similarity of the input which could be either resting-state data or emotional-task features to be classified. The similarity could be computed either by $L_2$-distance or clustering technique. The empirical results suggested that the transfer learning could outperform straight-forward method that directly trains SVM classifier on the aggregated data from source subjects. In addition, the proposed method of measuring subject-transferability could improve the classification accuracies that remarkably exceeded the chance levels. The proposed approach could expectedly shed light on the development toward subject-independent emotion recognition.
Chapter 6

Discussion and Conclusion

This research aims to construct more practical emotion recognition system using brain and physiological signals. It has been demonstrated that the proposed methods are promising but there are still rooms remaining for discussion.

6.1 Discussion

6.1.1 Access to testing data

In a real-world using situation, the ideal scenario for emotion recognition system is to have a new-coming user wear the wearable sensors and the system immediately starts detecting emotion of the user. On the other hand, a system in the most annoying case needs to record the resting-state signals of a user and then he/she also performs emotional tasks and annotates emotional labels, and the collected data is then used to train an emotion classification model. Although the proposed algorithms can avoid the worst scenario, the system could never become the ideal scenario as well. Due to the limitation of the proposed algorithms, there is still the necessity to record resting-state data and sometimes access the data that will be used for predicting emotion. This might slightly reduce the practicability of the system but not severely.

To specify, the measurement of stability needs to access raw signals including the resting-state and emotional-task periods to compute the probability of the occurrence of signals to detect the signals deviating significantly from the mean point. This can be avoided by using solely resting-state data to compute the mean values, but further studies are awaited in order to validate this technique. Alternative techniques with the minimal access to the testing data for detecting the artifact or outliers are also worthy to be studied.

In addition, the obtained results suggested that the performance of subject-independent
emotion classification can be substantially enhanced by finding transferable subjects using clustering technique on the emotional-task features to be later used for classification. However, this would disable the possibility of real-time emotion detection but it would still be possible for batch-type emotion recognition system that predicts emotion in every certain period of time. Another option is to find similar subjects by using only the resting-state data but the empirical results did not support this idea.

Therefore, designing the experiment for constructing emotion recognition model should not consider only on the classification results but also the usage in real-world scenarios. Threading off between the performance and the practicability should be carefully taken into account.

6.1.2 Comparison to previous works using conventional devices

One of the primary contributions of this research is to make use of multiple wearable devices to estimate emotional states at satisfactory accurate level. These devices may be keys to enable real-time measurement of brain states from small samples of noisy physiological signals recorded in real-world scenarios. As shown earlier, the highest accuracies obtained from EEG unimodality were 69.67% for valence classification (chance level 60.42%) and 69.91% for arousal classification (chance level 61.76%). In addition, the best accuracies for multimodal approach were obtained from context-aware adaptive methods achieving 71.24% and 70.99% accuracies for valence and arousal classification respectively.

Nevertheless, the accuracies are still not comparable to ones reported in previous works that used conventional devices. In those works, valence and arousal classification performance could reach 80-90% of accuracies [54, 43, 111]. However, it is noteworthy that each work had its own experimental protocol, tested on a variety of cross-validation techniques, and utilized different datasets. Unlike another machine learning works, the research on emotion recognition based on EEG is still in its infancy, hence there is no common benchmarking dataset or standardized experimental design that would allow direct comparison between each work. Conducting experiments by repeating the same experimental protocol in this research yet with conventional wet devices might be an option for allowing direct comparison between the performance of using wearable devices and conventional devices.
6.1.3 Limitation

Despite the increase of the innovation in wearable sensing technology, the development of practical wearable sensors is still an ongoing research. A number of wearable devices – including the one used in this research – have yet been officially approved on the usability in clinical usage. Despite comfortability, the signal quality of wearable devices is still not sufficiently comparable to a research-grade wet system and cannot be guaranteed throughout the recordings because the device might easily dislocate during the long-recording session. Future study should investigate the correlation of signal quality and real-world using factors in cluttered scenario such as the relationship with the degree of user movement. Such connection might also provide more insights for emotion detection research in real-world scenarios.

In addition, it is apparent that the resulted presented in this research were the best results with the best hyper-parameters. However, refining hyper-parameter should be done in the validation data rather than the testing data. Therefore, interpretation of the results of this research should be done with precaution and realization on their origins. It is also noteworthy that annotation played an important role in this research, especially in supervised training the classifiers. Meanwhile, the annotation itself was a difficult task for any non-expert annotator who was not familiar with self-emotion evaluation. Although annotation smoothing and emotion-awareness questionnaire were introduced, the trustworthy of this ground-truth information is always arguable from different aspects. Unlike another machine learning research, the performance of self-annotated emotion recognition always has particular ceiling owing to this limitation with regard to unclean ground-truth.

6.2 Conclusion

In summary, this research addressed two main concerns in emotion recognition based on wearable brain and physiological sensors to increase practicability. First, this research aimed to increase the robustness of the system by using multimodal technique with appropriately consider the stability of each modality. The contribution of each modality in decision-level fusion was regulated by that stability. The results demonstrated the promise of the proposed technique. Second, the challenge of inter-subject variability, which prevented practical usage and led to exhaustive training data recording, was alleviated by transfer learning technique with the consideration of subject transferability.
The data drawn from the most similar subjects were used to fine-tuning a model that would be used for classifying the emotions of a new-coming user. With the fabulous features of wearable sensors, the system applying the proposed algorithms is expected to be applicable in real-world situation smoothly.

6.3 Recommendation for future studies

As mentioned earlier, the performance on practical emotion recognition system can be further improved. This research mainly made use of SVM that was a popular machine-learning technique back to 2000s. However, the emergence of deep learning has opened a new era of machine learning and the preliminary results of an early work presented in Chapter 2 also demonstrated its potential. Alternative powerful architectures of deep learnings are expected to improve the emotion classification in this work and worthy to be studied in the future works.

Another possible future work is to take the psycho-physiological characteristics of emotion into account properly. In particular, emotion is momentary stable and does not rapidly vary in a short period of time (i.e., every 2 seconds). Rather emotional state changes when external or internal stimuli has sufficient influence. However, the empirical visualization of the predicted emotion showed that the estimation of emotion fluctuated over the course of time which was opposed to the natural characteristic of emotion. One solution would be incorporating the emotion predicted in preceding time frame and using emotional momentum to slow down the alternation. The momentum would prevent the fluctuation of emotion prediction and it would allow the change only when detecting substantial external or internal change. Alternative methods in time-series data mining research are applicable to this research as well.

Emotions in this research were simply quantized into 2 classes of arousal and 2 classes of valence, resulting the capability to predict merely 4 basic emotions. Due to the advantage of arousal-valence space, this research can be extended to recognize emotions directly as a point on the continuous space resulting in the increased granularity of the predicted emotions enabling the capacity to distinguish any similar emotions such as happy-joy and scare-angry.
Bibliography


Table A.1: Edinburgh Handedness Inventory; Which hand you prefer for that activity?

<table>
<thead>
<tr>
<th>Activities</th>
<th>Left always</th>
<th>Left mostly</th>
<th>No Preference</th>
<th>Right mostly</th>
<th>Right always</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Writing</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>2. Drawing</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>3. Throwing</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>4. Using scissors</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>5. Brushing teeth</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>6. Using a knife</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>(without a fork)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Using a spoon</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>(dominant hand)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Using a broom</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>(dominant hand)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Striking a match</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>10. Opening a jar</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>
Table A.2: Profile of Mood States (POMS) – Short Form; How you have been feeling in the past week including today?

<table>
<thead>
<tr>
<th>Adjective</th>
<th>Not at All (0)</th>
<th>A Little (1)</th>
<th>Moderately (2)</th>
<th>Quite a Lot (3)</th>
<th>Extremely (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tense</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Angry</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Worn Out</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Unhappy</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Lively</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Confused</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Peeved</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Sad</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Active</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>On Edge</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Grouchy</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Blue</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Energetic</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Hopeless</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Uneasy</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Restless</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Unable to Concentrate</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Fatigued</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Annoyed</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Discouraged</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Resentful</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Nervous</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Miserable</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Cheerful</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Bitter</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Exhausted</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Anxious</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Helpless</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Weary</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Bewildered</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Furious</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Full of Pep</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Worthless</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Forgetful</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Vigorous</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Uncertain about things</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Bushed</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>
Table A.3: Emotion Awareness Questionnaire (EAQ); Note that the items marked with stars are reversed-scored. The scores and stars do not appear in the questionnaires given to subjects.

<table>
<thead>
<tr>
<th></th>
<th>Not True</th>
<th>Sometimes True</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am often confused or puzzled about what I am feeling*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I find it difficult to explain to a friend how I feel*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Other people don’t need to know how I am feeling*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I am scared or nervous, I feel something in my tummy*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>It is important to know how my friends are feeling</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I am angry or upset, I try to understand why</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>It is difficult to know whether I feel sad or angry or something else*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I find it hard to talk to anyone about how I feel*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I am upset, I try not to show it*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I feel upset, I can also feel it in my body*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I don’t want to know how my friends are feeling*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>My feelings help me to understand what has happened</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I never know exactly what kind of feeling I am having*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I can easily explain to a friend how I feel inside</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I am angry or upset, I try to hide this*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I don’t feel anything in my body when I am scared or nervous</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>If a friend is upset, I try to understand why</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I have a problem, it helps me when I know how I feel about it</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I am upset, I don’t know if I am sad, scared or angry*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I am feeling bad, it is no one else’s business*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>My body feels different when I am upset about something*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I don’t care about how my friends are feeling inside*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>It is important to understand how I am feeling</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Sometimes, I feel upset and I have no idea why*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I am upset about something, I often keep it to myself*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>When I am sad, my body feels weak*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I usually know how my friends are feeling</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I always want to know why I feel bad about something</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I often don’t know why I am angry*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>I don’t know when something will upset me or not*</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Song ID</td>
<td>Music Title</td>
<td>Artist</td>
<td>Genre</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>001</td>
<td>Clear Blue Sky</td>
<td>Chatham County Line</td>
<td>Country</td>
</tr>
<tr>
<td>002</td>
<td>Nice And Easy</td>
<td>Jason Shaw</td>
<td>Country</td>
</tr>
<tr>
<td>003</td>
<td>Cold Summer Landscape</td>
<td>Bear Moon</td>
<td>Electronic</td>
</tr>
<tr>
<td>004</td>
<td>Lansdowne</td>
<td>WWIII</td>
<td>Jazz</td>
</tr>
<tr>
<td>005</td>
<td>Helix Nebula</td>
<td>Anamanaguchi</td>
<td>Rock</td>
</tr>
<tr>
<td>006</td>
<td>Jupiter The Blue</td>
<td>Gillicuddy</td>
<td>Rock</td>
</tr>
<tr>
<td>007</td>
<td>Chantiers Navals 412</td>
<td>L.J. Kruzer</td>
<td>Electronic</td>
</tr>
<tr>
<td>008</td>
<td>Shipyard</td>
<td>WWIII</td>
<td>Jazz</td>
</tr>
<tr>
<td>009</td>
<td>Trio for piano, violon, and horn in E-flat major, Op. 40, mvt. 2</td>
<td>Johannes Brahms</td>
<td>Classical</td>
</tr>
<tr>
<td>010</td>
<td>Traumerei Op. 15 mvt. 7</td>
<td>Robert Schumann</td>
<td>Classical</td>
</tr>
<tr>
<td>011</td>
<td>Tristan, Act 3, Tr 47</td>
<td>Richard Wagner</td>
<td>Classical</td>
</tr>
<tr>
<td>012</td>
<td>Tasso: Lamento e Trionfo, S.96</td>
<td>Franz Liszt</td>
<td>Classical</td>
</tr>
<tr>
<td>013</td>
<td>Pizzicato Polka</td>
<td>Johann and Joseph Strauss</td>
<td>Classical</td>
</tr>
<tr>
<td>014</td>
<td>Peer Gynt No. 1 (Morning mood)</td>
<td>Edvard Grieg</td>
<td>Classical</td>
</tr>
<tr>
<td>015</td>
<td>Concerto De Aranjuez, Adagio</td>
<td>Joaquin Rodrigo</td>
<td>Classical</td>
</tr>
<tr>
<td>016</td>
<td>A Skull Full of Maggots</td>
<td>Cannibal Corpse</td>
<td>Death Metal Rock</td>
</tr>
<tr>
<td>017</td>
<td>The Third Man Theme (Live @ KEXP)</td>
<td>Rat City Brass</td>
<td>Jazz</td>
</tr>
<tr>
<td>018</td>
<td>Pride &amp; Prejudice Track 4</td>
<td>Soundtrack</td>
<td>Film score</td>
</tr>
<tr>
<td>019</td>
<td>Calling You</td>
<td>Aqua</td>
<td>Pop Dance (90s)</td>
</tr>
<tr>
<td>020</td>
<td>Brandenburg Concerto No. 5 mvt. 1</td>
<td>Johann Bach</td>
<td>Classical</td>
</tr>
<tr>
<td>021</td>
<td>Scattered Knowledge</td>
<td>Revolution Void</td>
<td>Jazz</td>
</tr>
<tr>
<td>022</td>
<td>Shine Track 10</td>
<td>Soundtrack</td>
<td>Film score</td>
</tr>
<tr>
<td>023</td>
<td>All I Have to Do Is Dream</td>
<td>The Everly Brothers</td>
<td>Pop (60s)</td>
</tr>
<tr>
<td>024</td>
<td>Spring mvt. 2</td>
<td>Antonio Vivaldi</td>
<td>Classical</td>
</tr>
<tr>
<td>025</td>
<td>Eight</td>
<td>Marcel Pequel</td>
<td>Classical</td>
</tr>
<tr>
<td>026</td>
<td>The Portrait of a Lady Track 9</td>
<td>Soundtrack</td>
<td>Film score</td>
</tr>
<tr>
<td>027</td>
<td>No Surprises</td>
<td>Radiohead</td>
<td>Alternative Rock</td>
</tr>
<tr>
<td>028</td>
<td>Adagio for Strings, Op. 11</td>
<td>Samuel Barber</td>
<td>Classical</td>
</tr>
<tr>
<td>029</td>
<td>Crack in Time</td>
<td>Stickfigure</td>
<td>Electronic</td>
</tr>
<tr>
<td>030</td>
<td>Batman Returns Track 5</td>
<td>Soundtrack</td>
<td>Film score</td>
</tr>
<tr>
<td>031</td>
<td>Down with the Sickness</td>
<td>Disturbed</td>
<td>Metal Rock</td>
</tr>
<tr>
<td>032</td>
<td>Peter and the Wolf, wolf excerpt</td>
<td>Sergei Prokofiev</td>
<td>Classical</td>
</tr>
<tr>
<td>033</td>
<td>Night Drive</td>
<td>Decktonic</td>
<td>Electronic</td>
</tr>
<tr>
<td>034</td>
<td>Morceaux de Fantaisie, No. 2, Prelude in C# Minor</td>
<td>Rachmaninoff</td>
<td>Classical</td>
</tr>
<tr>
<td>035</td>
<td>Catkantei</td>
<td>Valentin Haussmann</td>
<td>Baroque Dance</td>
</tr>
<tr>
<td>036</td>
<td>5 00 AM</td>
<td>Peter Rudenko</td>
<td>Classical</td>
</tr>
<tr>
<td>037</td>
<td>New World Symphony - mvt. 2</td>
<td>Dvorak</td>
<td>Classical</td>
</tr>
<tr>
<td>038</td>
<td>Soul Intro - The Chicken</td>
<td>Jaco Pastorius</td>
<td>Jazz</td>
</tr>
</tbody>
</table>