論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Cross-sensory EEG emotion recognition under multimodal stimulation
著者(和文)	GAO Chenguang
Author(English)	Chenguang Gao
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第12938号, 授与年月日:2024年9月20日, 学位の種別:課程博士, 審査員:三宅 美博,井上 中順,小野 功,瀧ノ上 正浩,吉村 奈津江
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第12938号, Conferred date:2024/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Cross-sensory EEG emotion recognition under multimodal stimulation

GAO Chenguang

Supervisor: Yoshihiro Miyake

School of Computing, Department of Computer Science, Major of Artificial Intelligence
Tokyo Institute of Technology

This dissertation is submitted for the degree of
Doctor of philosophy

August 2024

Acknowledgements

Firstly, I would like to say thanks to Professor Yoshihiro Miyake, my main academic supervisor gratefully and sincerely, for the invaluable support and advice for me academically and personally. Through careful guidance and advice during my doctoral studies, I have made continuous progress in research, and these formed me as a qualified researcher. My sincere thanks to Professor Natsue Yoshimura, my sub academic supervisor, for her all-time support for my research, especially the instruction for the EEG measurement. My sincere thanks to Professor Hirotaka Uchitomi and Professor Taiki Ogata, for your patience professional discussions, and advice during the research projects in my doctoral time, which is invaluable. Great thanks to Professor Isao Ono, Professor Masahiro Takinoue, and Professor Nakamasa Inoue, for your professional review of my doctoral thesis and presentation. Also, thanks to all the professors, faculty members, and students in Miyake Lab, the countless support and care during my doctoral time are irreplaceable.

Special thanks to the family members. My father and mother gave me the courage to go through the difficulties during my doctoral time. My grandfather, grandmother, and uncle take care of me both mentally and financially.

Abstract

Electroencephalography (EEG) based emotion recognition enables computers to understand human emotions, which is a prosperous topic in the affective Brain-Computer Interface (a-BCI). Previous studies in EEG emotion recognition were sensory-dependent, the cross-sensory EEG emotion recognition under multimodal stimulation could not be achieved. However, human can percept emotion using cross-sensory information. There was a lack of systematic study for cross-sensory EEG emotion recognition under multimodal stimulation. To address this challenge, in this study, the purpose was investigating and achieving cross-sensory EEG emotion recognition under multimodal stimulation as video. This study consisted of two parts: Part 1. Emotion EEG analysis under multimodal emotion stimulation; Part 2. Cross-sensory EEG emotion recognition method.

In Part 1, was an EEG experiment conducted with six conditions: two types of emotion (Pleasure /Unpleasure) × three stimuli modality (Audio /Visual /Audio-visual). Feature analysis revealed sensory and emotion patterns in spatial, spectral, and quantitative features of EEG. In detail, as spatial features, emotion-related changes exhibited greater diversity in regional changes of the brain. As spectral features, emotion-related changes exhibited greater diversity in sub-bands of EEG. As quantitative features, sensory-related changes exhibited greater spectral impact quantitatively than emotion.

In Part 2, based on the cross-sensory emotion EEG under multimodal stimulation and feature analytic results from Part 1, this study proposed Filter Bank Riemannian Feature with Adversarial Domain Adaptation (FBADR) method for cross-sensory EEG emotion recognition. Filter Bank, Riemannian feature extraction, and domain adaptation were coordinated with

sensory and emotion patterns in spectral, spatial, and quantitative aspects, respectively. Classification results showed that the proposed method achieved state-of-the-art performance with 89.01%±5.06% accuracy in cross-sensory EEG emotion recognition with high robustness. Generally, this research provided a comprehensive idea for inspecting and realizing the cross-sensory EEG emotion recognition under multimodal stimulation from Part 1 and 2, for understanding the neurological evidence of emotion and sensory patterns under multimodal stimulation and achievement of cross-sensory EEG emotion recognition, widening the potential application for EEG-based a-BCI.

Contents

List of Figures

List of Tables

1. General Introduction

1.1 EEG-based affective-brain computer interface

EEG-based affective brain-computer interface (a-BCI) is becoming more and more prosperous that combines knowledge from neuroscience, biomedical engineering, and computer science to identify and interpret an individual's emotional states by monitoring their brain's electrical activity through EEG [1]. EEG-based a-BCI holds promise for various applications, including the diagnosis of psychological disorders, emotional assistance therapy, human-computer interaction, and the development of emotion-aware technologies [2]. Since this research is specifically laid on the field of EEG-based a-BCI, this part will introduce the basic principles, research methods, application areas, and research trending in EEG-based a-BCIs.

The basic principle of EEG-based a-BCI it that EEG is a non-invasive method for monitoring the electrically neurological activity by scalp electrodes to measure changes in cortical electrical potentials. EEG signals are associated with the firing activity of neurons and can reflect various functional states of the brain, including cognition, emotion, and motor control [3]. Emotional states are often associated with specific EEG patterns, which can be used to identify and interpret an individual's emotional state. The core principle of EEG-based affective BCI is to reveal mapping relationship between EEG signals and emotional states using machine learning and pattern recognition techniques. Firstly, EEG data is collected from participants while simultaneously recording their emotional states, typically through questionnaires, physiological measures, or behavioral observations. Then, this data is used to train classifiers or regression models to associate EEG signals with emotional states. The trained model is t to predict an individual's emotional state with EEG.

EEG-based a-BCI has a wide range of potential applications like clinical trials, human-computer interaction and emotion-aware systems. EEG-based a-BCI can assist in diagnosing and treating psychological disorders such as depression, anxiety, and post-traumatic stress disorder [4]. By monitoring a patient's emotional state, clinicians can gain a better understanding of their condition and develop personalized treatment plans. Emotion recognition can enhance the user experience in human-computer interaction. For example, in virtual reality environments, systems can adapt content or interaction based on the user's emotional state, providing a more personalized and immersive experience [5].Emotion-aware technologies are emerging to develop computer systems with emotion understanding and response capabilities. By combining EEG-based a-BCI with natural language processing and computer vision, more intelligent human-computer interactions and emotion analysis can be achieved [6]. Currently, EEG-based a-BCI research is in a rapidly evolving phase, with researchers exploring new methods and application areas. Researchers are increasingly interested in combining EEG with other modalities for emotion recognition, such as facial expressions, voice, and physiological signals. Multimodal emotion recognition aims to capture a more comprehensive view of an individual's emotional state [7].Real-time emotion recognition is a challenging but promising area of research. Researchers are working on developing real-time EEG-based a-BCI systems [8], for applications in emotion monitoring, emotion-assisted therapy, and real-time feedback for emotion-aware technologies. Since everyone's brain activity patterns and emotional experiences are unique, personalized models are a significant direction in EEG-based a-BCI research [9]. Personalized models can be trained based on an individual's EEG data, improving the accuracy of emotion recognition. As the applications of EEG-based a-BCI continue to grow, data security and privacy have become important concerns. Researchers need to consider how to protect the security and

privacy of EEG data to prevent potential misuse and privacy violations [10]. Therefore, EEG-based a-BCI is essential for understanding and integrating human emotion towards external devices, enable computational methods and technologies to boost human's emotion intelligence.

Precise emotion recognition is important for a-BCI system. The main points are experimental paradigms, feature extraction and pattern recognition. In EEG-based a-BCI research, specific experimental paradigms are often used to induce different emotional states. For example, emotional movies, music, or images can be used to elicit emotions while recording participants' EEG signals. These experimental paradigms help researchers collect EEG data related to emotions for model training and evaluation [11]. Feature extraction involves transforming raw EEG signals into numerical features for machine learning. Commonly used features include time-domain features, such as mean power [12], frequency-domain features, such as spectral energy distribution [13], spatial feature, such as channel covariance matrix [14] and time-frequency domain features, such as wavelet transform coefficients [15]. The goal of feature extraction is to extract informative features that help classifiers better discriminate between different emotional states. Pattern recognition in a-BCI is the process of modeling the relationship between features and emotional states. Common pattern recognition methods include support vector machines, artificial neural networks, decision trees, and random forests [6]. These methods can classify, or regress emotional states based on combinations of features, enabling emotion recognition. When take attention into sensory modality on previous proposed methods, the wide-existing methods are sensory dependent. Specifically, one method only applied for one specific type of stimulation, if types of stimulation changed, the method must be changed correspondingly. For instance, Audio-induced EEG emotion recognition: Lin et al [16] utilized power spectrum density (PSD) feature and multilayer perceptron (MLP)/ SVM classifier

to decode music-induced EEG for joy, anger, sadness, pleasure, which achieved mean acc = 82.29%. Visual-induced EEG emotion recognition: Mehmood et al [17] utilized Hjorth parameters feature into multiple classifiers with separate or ensemble to decode affective picture induced EEG for sad, scared, happy, calm, which achieved mean acc = 76.6%. Audiovisual-induced EEG emotion recognition: Liu et al [18] utilized short-time Fourier transform (STFT) feature and LDA feature selection with SVM classifier to decode audiovisual-induced EEG for non-neutrality/neutrality, achieved real-time EEG emotion recognition with mean acc= 92.26%. Sensory-dependency decoding methods fade the further application of a-BCI systems.

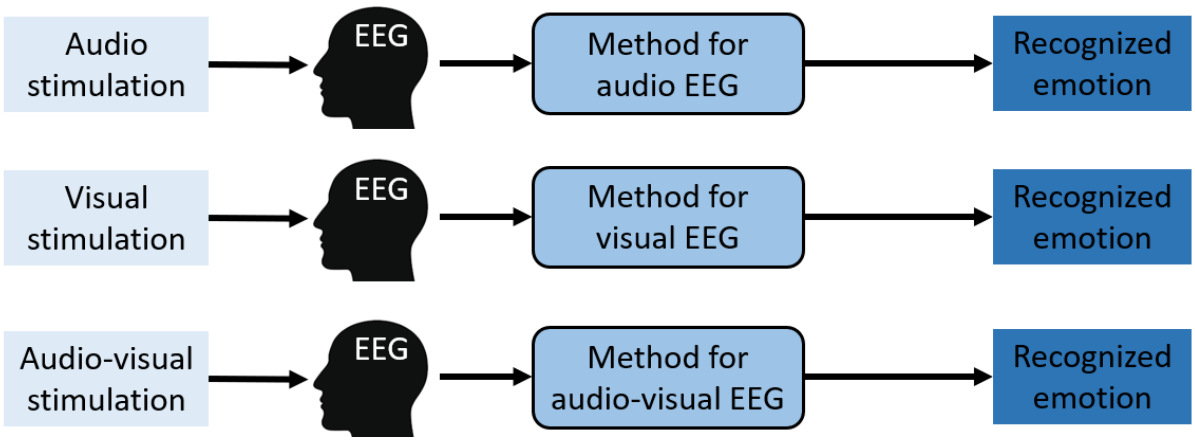


Figure 1.1. Current sensory dependency in EEG emotion recognition.

1.2 Emotion recognition and perception

Emotions is the basic element of modal society and appears and influence human's daily experiences, participating in people's interactions with meaning, shaping people's perceptions of the world. In this expansive experience of human emotions, people encounter a rich and diverse spectrum of emotional states, including joy, sadness, excitement, fear, anger, surprise, and countless others [19]. Emotions are not only a short time influence, but they also constitute a long-term impact in everyone's lifespan. Emotion composes the experience of our inner lives and serves as vital building factor s to construct of our humanity. Emotions are wide-existing, and lead to various domains of our lives [20]. In the process of social interactions, they are the silent but crucial characters, guiding us in our interactions with others. A smile exchanged between friends, the shared laughter among loved ones, and the comforting embrace of a family member, all showed emotions play in formatting of our social connections [21].Moreover, emotions are not only confined solely to subjective experiences [22] but also possess the power to shape human physical well-being [23]. Stress, as a physiological response to challenging situations, can impact on our heart rate, blood pressure, and immune system [24]. Conversely, positive emotions, such as contentment and joy, can bestow upon us a sense of relaxation and serenity, bolstering our immune function and contributing to our overall health and resilience [25].

Both emotion recognition and emotion perception are important concepts in studying different aspects of human emotions. Emotion recognition refers to judging human emotional states by analyzing external clues, such as facial expressions, voice, body posture, EEG, etc. [26]. Technical tools, such as computer vision, deep learning, and machine learning, are usually used to identify emotional features and patterns from data in an automatic or semi-automatic way [27]. In contrast, emotion perception studies and explores the process by which individuals, especially

humans, subjectively perceive, understand, and experience emotions [28]. This includes individuals' cognition and response to emotions, which is usually studied through psychological experiments, questionnaires, and other methods [29]. Both aim at understanding nature and expression of human emotions. However, emotion recognition focuses more on the analysis and identification of external cues [27], while emotion perception focuses more on an individual's subjective perception and experience of emotions [30]. From a certain level, emotion recognition is a machine-based and patterned process of human emotion perception. Through emotion recognition, researchers enable machines to understand emotions like humans and provide corresponding feedback.

Emotion recognition, driven by technological advancements and its diverse applications across various domains, has undergone significant evolution in recent years. This transformative field is marked by several key trends and processes. The multimodal approaches emerged as a milestone, where researchers integrate diverse data sources, including facial expressions, vocal cues, physiological signals, and textual data. This comprehensive approach enhances the accuracy and robustness of emotion recognition systems by harnessing complementary information from multiple modalities [7]. Deep learning, notably convolutional and recurrent neural networks, excels at automatically extracting features from raw data, enabling more precise emotion recognition [31]. Transfer learning has gained attention as a strategy to improve emotion recognition performance [32]. Real-time emotion recognition systems are being actively developed, these systems enable dynamic and adaptive responses based on users' emotional states, enhancing human-computer interaction and user experiences [33]. Privacy and ethical considerations have become increasingly significant in the field of emotion recognition technology. There is a growing awareness of the need to address data privacy, obtain informed

consent, and ensure responsible use of this technology [34]. Efforts to enhance cross-cultural sensitivity are underway, recognizing that emotions can manifest differently across cultures [35]. Emotion-aware AI systems are being developed to create more personalized and empathetic human-computer interactions. These systems can adapt their responses and behaviors based on users' emotional states, fostering more meaningful and effective communication [36]. Emotion recognition has found valuable applications in healthcare, particularly in mental health monitoring. Detecting and tracking emotional states in patients, aids in early diagnosis and treatment, potentially improving overall well-being [27]. In the area of market research and advertising, emotion recognition plays an important role in understanding consumer reactions to products, advertisements, and user interfaces. This data-driven approach informs marketing strategies and decision-making processes [37]. Emotion recognition technology can accurately identify an individual's emotional state by analyzing various data such as voice, facial expressions, and physiological indicators, providing more accurate human-computer interaction for intelligent systems [38]. At the same time, multimodal stimulation technology integrates multiple senses such as audio, visual, olfactory, etc., and can deeply influence and regulate the emotional experience of users through multi-faceted stimulation [25]. The development of these technologies had broad application prospects in virtual reality, games, medical, and other fields.

The current state of research on emotion perception is constantly evolving. Researchers are deeply exploring the cognitive neural mechanisms of emotion perception, individual differences, and the relationship between emotional experience and behavior [39]. Through a variety of methods such as neuroimaging [40], psychological experiments [41], and computational modeling [42], they tried to parse the brain's role in emotion perception and changes in emotion processing . As neurological mechanism of cross-sensory emotion perception,

brain mechanisms of cross-sensory were investigated by fMRI, different sensory may be mapped onto amodal emotion processing [43][44], the brain regions include pSTS (posterior superior temporal sulcus), PFC (prefrontal cortex) and pCC (posterior cingulate cortex) would be activated for processing amodal emotion, it enables modality-nonspecific for emotional meaning, which means Human can percept the emotions regardless the types of stimulations [45]. Related to the emotion recognition mentioned above, an important research question is how to make machines capable of cross-sensory recognition of emotions like humans.

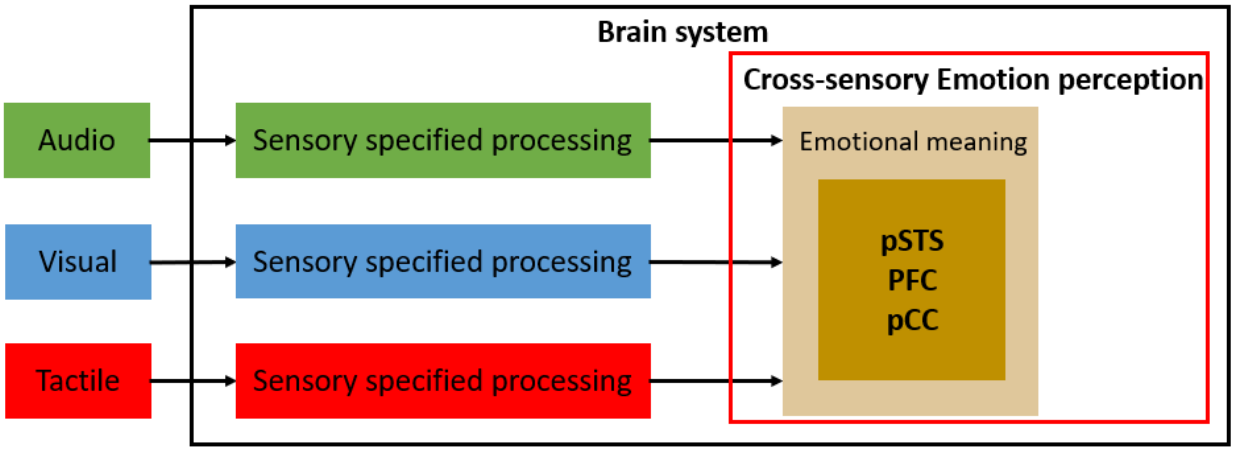


Figure 1.2 Brain mechanism of cross-sensory emotion perception

1.3 Remaining problems

Emotion recognition is a key point for effective EEG-based a-BCI. The precise emotion recognition for EEG enables the successful operation of a-BCI. As for the experimental environment and diagram in a-BCI, external stimulations always serve as the origin for elite user emotions. The stimulations human experienced, most of them are multimodal stimulations. When humans are encountering multimodal stimulation, they can perceive the same types of emotion from partial or entire sensory inputs [45]. Previous studies in EEG emotion recognition were sensory-dependent, the a-BCI can't recognize cross-sensory emotion like humans. Since EEG is subjected to changes from both sensory and emotion, sensory patterns and emotion patterns are not clear from EEG under multimodal stimulation by far. There is a lack of systematic study for cross-sensory EEG emotion recognition under multimodal stimulation. The main problems for solving cross-sensory EEG emotion recognition are two aspects: emotion EEG analysis under multimodal stimulation and cross-sensory EEG emotion recognition methods. These two problems mainly indicate the impact of multimodal emotion stimulation on EEG responses and the sensory dependency in EEG-based emotion recognition.

1.3.1 The impact of multimodal emotion stimulation on EEG responses

The EEG cues from EEG are fundamental basement for understanding users neurological feedback for a-BCI system. The very first issues for making cross-sensory EEG emotion recognition unsolved is the EEG patterns, especially emotion and sensory patterns under multimodal stimulation are unclear. Previous study indicate that emotion pattern and sensory pattern with emotion and sensory-related changes respectively [46] [47], but lack of the comparatively study for EEG behave with respect to the sensory-related changes and emotion-related changes under multimodal stimulation. The impacts of multimodal emotion stimulation are major points. It not only would bring the psychological understanding of cross-sensory emotion perception, but most importantly, it enable the further development of and achievement of cross-sensory EEG emotion recognition under multimodal stimulation.

1.3.2 Sensory dependency in EEG-Based emotion recognition

The difficulty for applying EEG-based a-BCI as sensory-independent as human emotion perception is from the current sensory-dependent EEG emotion recognition methods. One framework only applied for one specific type of stimulation, if types of stimulation changed, the framework must be changed correspondingly. With retrospective to the previous research in EEG emotion recognition, audio-stimulated emotion EEG [16], visual-stimulated EEG [17], audio-visual-stimulated EEG [18], their emotions could be decoded with proposed methods. However, there is no certain method focusing on cross-sensory EEG emotion recognition, like one method could be appliable for emotion recognition regardless of sensory inputs. In reality, the multimodal stimulation and cross-sensory situations are widely existing. Hence, the solution for current sensory dependency in EEG-based emotion recognition is crucial for the further development of EEG-based a-BCI.

1.4 Research purpose and approach

With respect to existing problems in EEG-based a-BCI research and its major difficulties which mentioned before, the general research purpose is achieving cross-sensory EEG emotion recognition under multimodal stimulation. The multimodal stimulation in this research, which focused on audio-visual stimulation (video). The cross-sensory situations for multimodal stimulation are audio, visual (partial sensory inputs) and audio-visual (complete sensory input) with two types of emotion component respectively. The research will focus on emotion EEG analysis under multimodal stimulation and cross-sensory EEG emotion recognition methods. These two parts will be presented in chapter 2 and chapter 3, which respect to problem 1.3.1 and 1.3.2.

As emotion EEG analysis under multimodal stimulation, this research would like to investigate the brain's response which from collected EEG, to clear out how emotion and sensory-related changes will influence on EEG as emotion and sensory patterns. 20 participants would participate in the experiments and their EEG would be collected under the experimental conditions with multimodal emotion stimulation. The collected EEG after preprocessing, would go through feature analysis for revealing emotion and sensory pattern. The feature analytic results would provide a comparative idea for emotion and sensory patterns characters under multimodal stimulation and bring the EEG inspection for cross-sensory EEG emotion perception.

For cross-sensory EEG emotion recognition, the feature interpolation of emotion pattern would be enabled by analytic results from emotion EEG analysis under multimodal stimulation. This study would build a method for cross-sensory EEG emotion recognition. The realization and validation of proposed methods would be taken for supporting proposed cross-sensory EEG emotion recognition's effectiveness.

1.5 Structure of thesis

The thesis is organized as follow structure in Figure 1.3. Chapter 1 is for envision of EEG-based a-BCI , emotion recognition and their sensory-dependent problem, which brought motivation and purpose for this research. Chapter 2 is for experimental design and feature analysis of cross-sensory emotion EEG under multimodal stimulation. Chapter 3 is for cross-sensory EEG emotion recognition method, the method design, realization and validation. Chapter 4 is general discussion of results from chapter 2 and chapter 3 as well as future works. Chapter 5 is for concluding the works in this study.

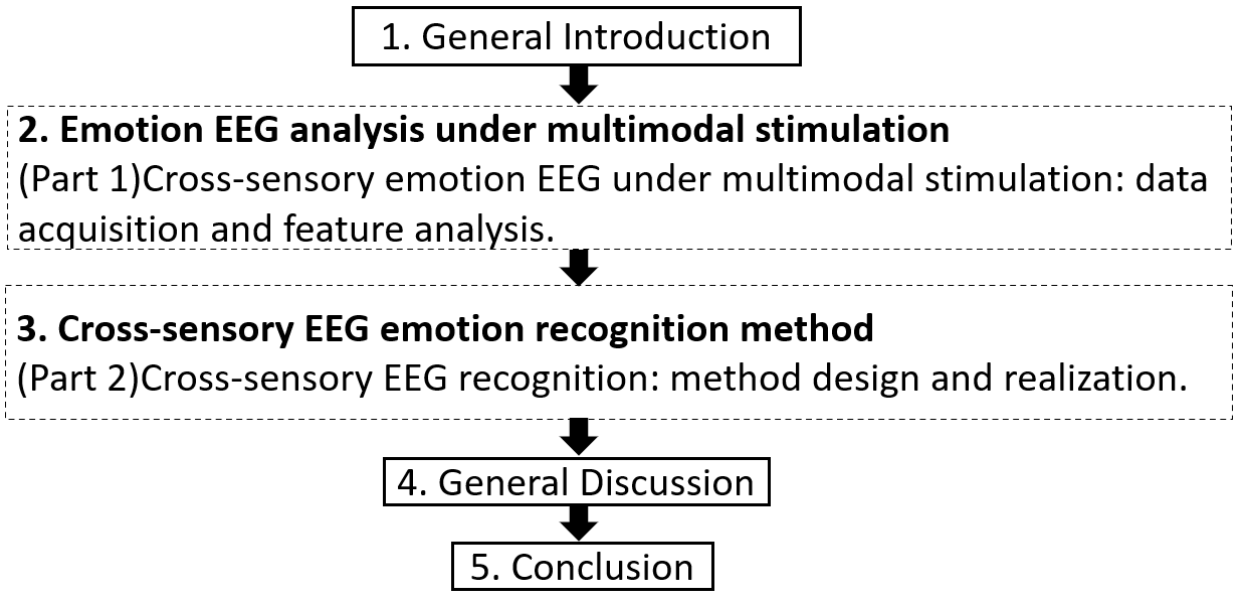


Figure 1.3. Workflow of paper structure.

2. Emotion EEG Analysis under Multimodal Stimulation

2.1 Introduction

Multimodal emotion stimulation is widely used in modern society for evoking various emotional states in humans via external sensory inputs, including auditory, visual, olfactory, and tactile cues. These stimulations are prevalent in different media, such as movies, animation, and virtual reality, and they can be used in education, marketing, advertising, psychological counseling, and entertainment [48–50]. A study on the response of the human brain to these stimuli is critical for understanding the process of emotional multimodal stimulation. EEG is commonly used to analyze and understand how emotion and sensory stimulation affects human brain activity. EEG measures the electrical activity of the brain by using electrodes placed on the scalp and recording the signals generated by neurons. By analyzing EEG data, researchers can better understand how these stimulations affect specific areas of the brain and elicit emotional responses in humans [51].

BCI are promising in the future of multimodal emotion recognition [52][53]. EEG signals can be used to recognize emotions from visual and audio-visual stimuli [54], and the fusion of EEG signals and audio-visual features can improve emotion recognition performance [55]. However, there is a lack of research on the process of multimodal emotional stimulation and how it affects human brain activity. Therefore, research in this area could be significant in improving the effectiveness of these stimulations. This research can assist in the development of new methods used for designing more effective emotional stimuli and for attaining a better understanding of how various multimodal stimulations impact different individuals.

Recent studies using EEG have investigated the neural correlates of emotional processing. Jiang et al. [56] found that alpha and beta oscillations in the prefrontal cortex are modulated by

the valence and arousal of emotional stimuli during an affective priming task. In another study, Mu et al. [57] found that beta and gamma oscillations in the temporal and parietal regions were associated with the processing of social feedback during interpersonal trust games. Schelenz et al. [58] investigated the multisensory integration of dynamic emotional faces and voices and proposed a method for simultaneous EEG and fMRI measurements. Additionally, Li et al. [59] demonstrated that alpha oscillations in the prefrontal cortex are associated with cognitive reappraisal during an emotion regulation task, whereas Yu et al. [61] found that delta and theta oscillations in the anterior cingulate cortex and insula are modulated by the perceived intensity of emotional pain during a social exclusion task. The medial temporal pole was observed to be responsible for driving amygdala activity in response to emotional stimuli by Sonkusare et al. [61]. Ahirwal and Kose investigated correlated EEG channel agendas, and the results were examined with respect to audio-visual emotion classification [62]. In the spectral domain, Balconi and Vanutelli investigated how vocal and visual stimulation, congruence, and lateralization affect brain oscillations in emotional interspecies interactions [63]. Temporal analyses were conducted by Balconi and Carrera [64], and they showed that the cross-modal integration of an emotional face and voice leads to the P2 ERP effect. These studies highlight the continued use of EEG in revealing the neural mechanisms underlying emotional processing.

EEG signals have been used to investigate the neural mechanisms unrevealing sensory processing, including visual, auditory, and somatosensory processing. For instance, Gallotto et al. [65] investigated the neural correlates of visual perception using EEG and found that alpha and beta oscillations were modulated by attentional demands during visual processing. Craddock et al. [66] found that alpha oscillations in the somatosensory cortex were associated with the detection of tactile stimuli. Klasen et al. [67] found that emotional information processing in

multimodal environments is highly integrated, with interactions and regulatory effects between different sensory channels. Additionally, Bugos et al. [68] used EEG to investigate the neural mechanisms underlying auditory processing and found that theta oscillations in the auditory cortex are modulated by musical features. Zimmer et al. [69] found that emotional cues presented in multiple sensory modalities can improve visual attention. Ross et al. [70] investigated the neural correlates of somatosensory processing using EEG and found that gamma oscillations in the somatosensory cortex are associated with the processing of vibrotactile stimulation. Brain functional area studies were utilized by Magosso et al. [71], revealing that EEG alpha power is as-sociated with attentional changes during cognitive tasks and virtual reality immersion sensory stimulations. The effects of prolonged waking-auditory stimulation on EEG synchronization and cortical coherence were investigated by Cantero et al. [72]. Baumgartner et al. [73] found that different types of sensory stimuli evoke emotions in different ways, but the essential characteristics of emotional experience are similar. These studies demonstrate the utility of EEG in uncovering the role of oscillatory activity in sensory processing and its relationship with perception and attention.

However, previous EEG-related research has mutually and exclusively focused on the influence of stimulation modality and emotional stimuli. The effects of EEG with changes in multimodal emotional stimulations have not been thoroughly studied, especially the spatiotemporal analysis under its uni-stimulus-modality and multi-stimulus modality. Overall, much remains to be understood regarding the neural mechanisms underlying emotional and sensory processing.

In this chapter, which is aimed to investigate the changes in human brain activity in response to multimodal emotional stimulation and to assess the effects of the emotional mode

and sensory modality individually and together with respect to EEG. These investigations were conducted because movies are recognized as useful naturalistic stimuli for neuroimaging studies, and they were utilized in previous studies [74]. To achieve this, this study selected short film clips as the source of multimodal (visual + auditory) emotional stimulation and recruited 20 participants. This study assessed two emotions (pleasure/unpleasure) and three modes (audio/visual/audio-visual); thus, a total of six conditions were assessed by EEG and self-assessment data analyses. The subjective evaluation of the self-assessment metric (SAM) and objective PSD analysis of EEG with brain functional area allocation and sub-band analysis. Emotion patterns and sensory patterns of EEG under multimodal stimulation could be obtained with respective and comparative analysis. The results of this study can be used to gain an in-depth understanding of the relevant changes in human brain activity in response to multimodal emotional stimuli, thus contributing to research and application in the areas of emotion regulation, psychotherapy, and EEG-based a-BCI.

2.2 Materials and Methods

2.2.1 Experimental setup

2.2.1.1 Participant information

Referring to the previous EEG experiment using the emotion mode and sensory modality [75–77], twenty individuals were invited to participate in this study. Of the participants, 11 were men and 9 were women, aged 24.7 ± 1.9 years. All participants were inter-national students at the Tokyo Institute of Technology and native Chinese speakers, right-handed, and had no history of mental illness or recent mental or physical trauma. All participants provided written informed consent to participate in this study. The study was conducted in accordance with the principles of the Declaration of Helsinki. The study was approved by the Ethics Committee of the Tokyo Institute of Technology.

2.2.1.2 Emotional stimulus

Next, 20 emotional video clips were selected to elicit 2 emotions: pleasure or unpleasure (sad). Each emotion was shown in 10 video clips, with each lasting 30 s. Stimulus video clips were sourced from the New Standardized Emotional Film Database for Asian Culture [78]. The detailed information for selected video clips were referred to Appendix Table A1. This study trimmed the original videos from the database to 30 s to meet the stimulus duration requirements for this study.

2.2.1.3 Experimental protocol

During the experiment, subjects were asked to sit in a comfortable chair with their hands placed flat on a table. A 24-inch monitor with 1024 × 768 resolution was connected to a computer that sent control commands to play the experimental instructions and visual stimuli on

the screen. The subjects were asked to wear AirPods Pro2 in the noise-canceling mode while playing audio stimuli. The distance between the participants and the monitor was approximately 90 cm, measured with their chins resting on a chinrest to prevent the shaking of the head. The experimental setup is shown in Figure 2.1, the photo used in Figure 2.1 had been approved by the participant.

Before the experiment, a questionnaire survey was conducted to assure the physical and mental condition of the subjects. The contents of questionnaire were referred Appendix: A3. Questionnaire before emotion EEG experiment. After the questionnaire, according to the type of stimulus, there were three experimental conditions/modalities: audio stimulus, visual stimulus, and audio-visual stimulus. Resting conditions were used as control.

Audio stimulus modality: The film clip was played with a black screen, and only audio was provided. Visual stimulus modality: The film clip was played while muted, and only visual stimulus was provided. Audio-visual stimulus modality: The film clip was played with original audio and visual information. Resting state: No audio or visual information was provided

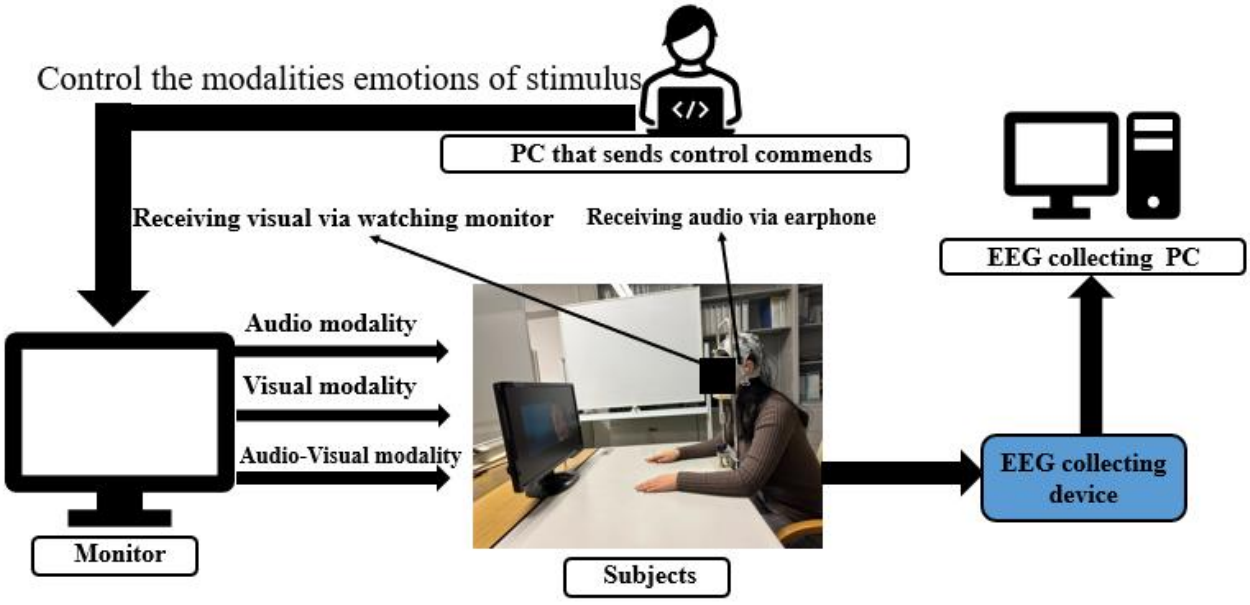


Figure 2.1. Illustration of the experimental setup

Each subject first performed the experiment in the resting state to enable the measurement of EEG in the resting state. In the resting state, the subjects maintained a sitting posture that was the same as that in stimulus-corresponding tasks. The duration for the resting state EEG recording was 300 s.

The experimental workflow for the three stimulus modalities of multimodal emotional stimulation is shown in Figure 2.2. The experimental execution of the multimodal emotional conditioning task consisted of 60 trials, each lasting approximately 2.5 min. At the beginning of each trial, a 5 s cue indicated the starting time, followed by an 8 s cross-fixation to direct the participant's attention. The stimulus lasted 30 s and was pseudo-randomly presented with one of two emotional stimuli (pleasure or unpleasure) in one of the three stimulus modalities. Pseudo-random protocol was referred to Appendix A2. Pseudo-random protocol for the experiment. After the presentation of the stimulus, participants completed a self-assessment questionnaire to rate the type and degree of emotion experienced, spending approximately 20 s. Then, there was a 60 s rest without any visual or auditory stimulation. For each subject, the entire experiment, which included the requisite rest during the emotional conditioning task and resting state EEG measurements, lasted approximately 2.5 h; a 5 min break was provided after every 20 trials to avoid fatigue effects.

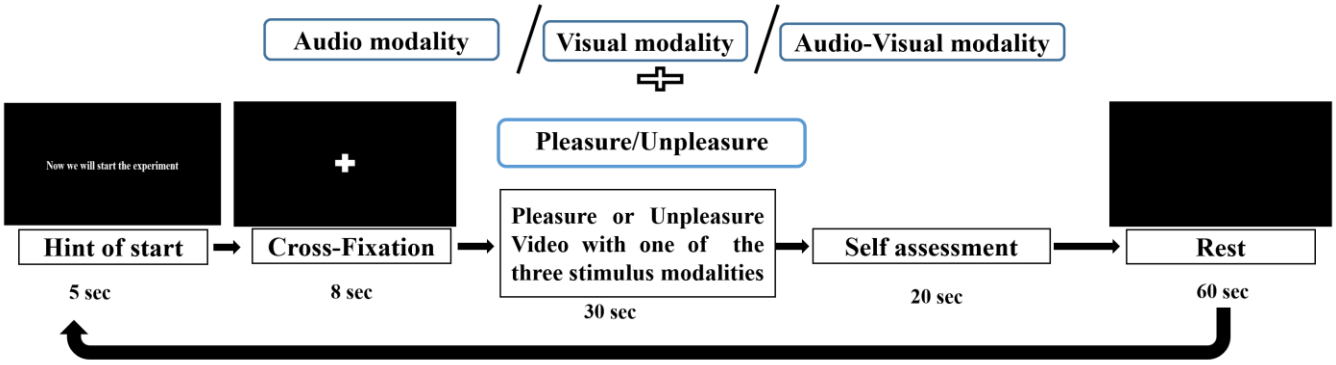


Figure 2.2. Workflow for stimulation representation

2.2.2 Data acquisition

2.2.2.1 EEG data acquisition

In this study, a 32-channel EEG system (Brain Products GmbH, Gilching, Germany) was used to collect EEG data, and the electrodes were as follows: Fp1, Fp2, F3, F4, F7, F8, FT9, FT10, FC5, FC6, T7, T8, CP5, CP6, TP9, TP10, P7, P8, FC1, FCz, FC2, C3, Cz, C4, CP1, CP2, P3, Pz, P4, O1, O2, and Oz. The positions of the channels on the scalp are shown in Figure 2.3.

Before the experiment, each subject wore a headset for EEG collection. By the application of a special gel to the scalp, the impedance between each electrode and the scalp was reduced to below 10 kΩ. Using Brain Vision Recorder software, an amplifier was used to amplify weak electrical signals from the brain, and the data acquisition system was used to digitize and store the data. This study used FCz as the reference electrode, with a sampling frequency of 500 Hz. The data from these experiments are detailed in Table 2.1.

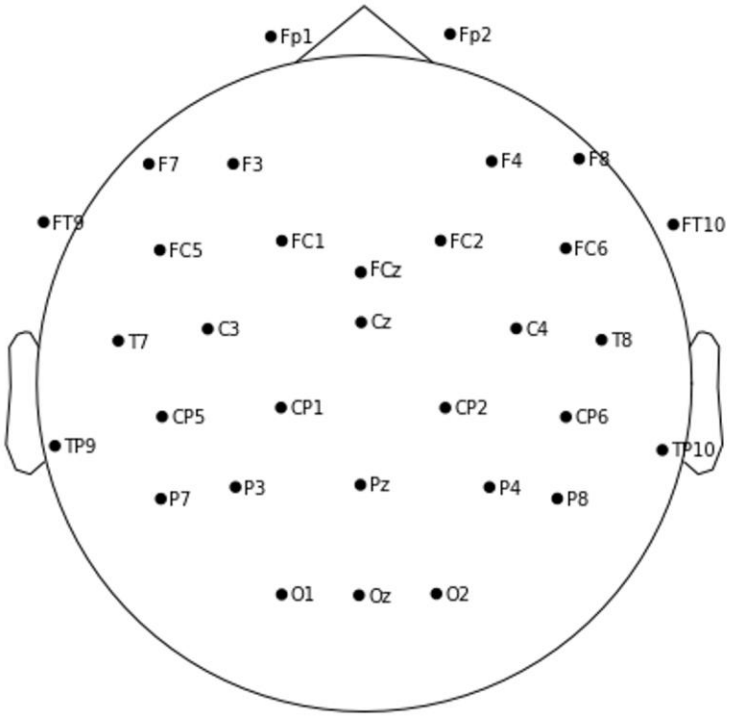


Figure 2.3. EEG electrode distribution.

Table 2.1. EEG data collected during the experiment.

Audio-Unpleasure	Visual-Unpleasure	Audio-visual-Unpleasure	Audio-Pleasure	Visual-Pleasure	Audio-visual-Pleasure	Resting State
200 trails	200 trails	200 trails	200 trails	200 trails	200 trails	20 trails

EEG data were divided into seven categories. Audio unpleasure EEG data were collected with the audio stimulus modality, and film clips were collected with unpleasure emotions. For visual unpleasure data, EEG data were collected with the visual stimulus modality, and film clips were collected with unpleasure emotions. For audio-visual unpleasure data, EEG data were collected with audio-visual stimulus modalities, and film clips were collected with unpleasure emotions. For audio pleasure data, EEG data were collected with the audio stimulus modality, and film clips were collected with pleasure emotions. For visual pleasure, EEG data were collected with the visual stimulus modality, and film clips were collected with pleasure emotions. For audio-visual pleasure, EEG data were collected with the audio-visual stimulus modality, and film clips were collected with pleasure emotions. Resting state EEG data were collected in the resting state.

Collected data comprised the following: 20 (subjects) × 2 (emotional classes) × 3 (stimulus modalities) × 10 (trials) = 1200 trials, with 200 trials for each emotion with 1 stimulus modality. The duration of each trial was 30 s. Resting state = 1 × 20 (subjects) trials were collected before the main experiment with a duration of 300 s for each trial.

2.2.2.2 Self-Assessment data

In each trial of the multimodal emotional conditioning tasks, self-assessment data were collected post-stimulus. According to the self-assessment metric (SAM) [79], the participants' self-reported scores during each trial, as shown in Figure 2.4.

During the self-assessment in each trial, subjects named the level of emotional arousal they experienced and assigned an integer score from 0 to 10. If the participant observed that the emotion was pleasure, it was recorded as the original score, and if the emotion was unpleasure, it was recorded as a negative absolute value of the score. All readings were recorded by the researcher.

Self-assessment data were used as a subjective index to validate multimodal emotional stimulation, and the data were explored using EEG analysis.

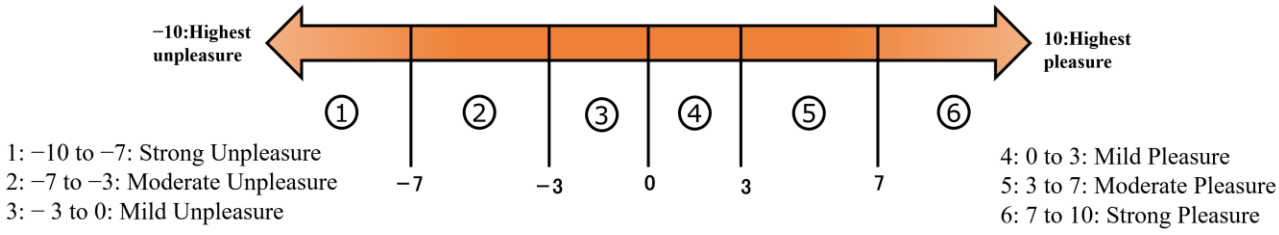


Figure 2.4. Self-assessment metric [79].

2.2.3 Data analysis

2.2.3.1 EEG data preprocessing

When preprocessing EEG data to enhance data quality, it utilized several methods to improve the accuracy and reliability of the data. Specifically, the following steps were taken:

Filtering: Appling a notch filter with a frequency of 50 Hz to remove electromagnetic interference from the EEG data. Additionally, it used the average potential as the reference electrode and re-referenced the EEG electrodes to further enhance data quality. These steps helped reduce noise and artifacts in the EEG signals.

Artifact removal: It identified and removed EEG signals that were greater than 100 μv, which are commonly considered noise in EEG data. This step helped further reduce the impact of artifacts on the data.

ICA decomposition: It used independent component analysis (ICA) to identify and separate the different independent sources of the EEG signals. This step helps identify components that are not related to the neural activity of interest, such as eye movements or muscle artifacts. The ICA result would be used in ICLabel.

ICLabel: It used the ICLabel tool in EEGLAB [80] to classify the independent components obtained by ICA. ICLabel is a commonly used automatic classification tool that classifies independent components into different categories, including eye movement, muscle artifact, and neural activity. It used ICLabel to remove components corresponding to eye movements and muscle artifacts.

By utilizing these methods, it were able to preprocess the EEG data and prepare it for further analysis. The combination of filtering, artifact removal, ICA decomposition, and ICLabel classification helped enhance data quality and reduce the impact of noise and artifacts on the data.

2.2.3.2 Power spectral density (PSD) analysis of EEG

For the analysis of EEG data after preprocessing, we first conducted frequency domain analysis. In frequency domain analysis, the main index used was PSD, which is the result of the spectral analysis of the power of EEG signals. Different frequency bands of EEG can be obtained using the above experimental conditions and the distribution of PSD in different regions; thus, the significance of brain activity under different experimental conditions can be specifically explained [81]. Meanwhile, the inner condition comparison can be easily conducted using PSD. Before explaining the PSD calculation method adopted in this study, this part specify the strategy for the PSD analysis. As this experiment pertained to emotion and stimulation modalities, it used frequency band and brain functional analyses together. It aimed to reveal the conditions under different frequency ranges and the activity of each functional brain area corresponding to the EEG obtained from different experimental conditions.

EEG sub-band Analysis, which EEG spectrum band research is a well-recognized analysis method in frequency domain studies and has significance in the study of emotion and sensory. In this EEG, signals are divided into various frequency bands [82]. Commonly used sub-spectral bands include delta: 1-4 Hz; theta: 4-8 Hz; alpha: 8-14 Hz; beta: 14–30 Hz; and gamma: 30–100 Hz [83]. In this EEG study with emotional and multimodal (visual, auditory, and audio-visual) stimulations, this study selected three bands—delta, theta, and alpha—and used a bandpass filter to obtain sub-spectrum bands.

Increases in delta and theta waves are associated with negative emotions, such as unpleasure, while increases in alpha waves are associated with positive emotions, such as pleasure [84]. In addition, different stimulation modalities can elicit different EEG patterns that can be used to differentiate stimulus types, such as visual, auditory, or multimodal stimuli. These

frequency bands are particularly relevant in the analysis of emotion and stimulus patterns as they provide valuable insights into the physiological changes that occur in response to different emotional and stimulus conditions [85]. By analyzing these specific frequency bands, it can better understand the neural mechanisms underlying emotional processing, such as unpleasure and pleasure. For example, it enables compare pleasure and unpleasure emotional stimuli by comparing the intensity changes in different frequency bands. In addition, this study compared the effects of visual, auditory, and audio-visual stimuli on different frequency bands, thereby assessing the effects of different stimulation modalities on emotional and sensory responses.

EEG functional brain mapping, which studies of divided functional brain areas is important in the field of brain science [86]. Functional brain division refers to the subdivision of the entire brain into smaller regions as per their contribution to different functions. In this study, five functional areas were defined as corresponding to the EEG channels utilized in the experiment: prefrontal, temporal, central, parietal, and occipital [87]. Except for the original reference channel FCz, the remaining 31 channels were allocated as per Figure 2.5.

The prefrontal cortex (Fp1, Fp2, F3, and F4) is mainly responsible for thought, cognition, and behavioral control. The temporal cortex (F7, F8, FT9, FT10, FC5, FC6, T7, T8, CP5, CP6, TP9, TP10, P7, and P8) is primarily responsible for language processing, auditory cognition, and memory processing. The central cortex (FC1, FC2, C3, Cz, and C4) is primarily responsible for functions, such as perception, sensation, and sensory fusion. The parietal cortex (CP1, CP2, P3, Pz, and P4) is primarily responsible for body perception, spatial orientation, and hand–eye coordination. The occipital cortex (O1, O2, and Oz) is primarily responsible for visual processing and cognition.

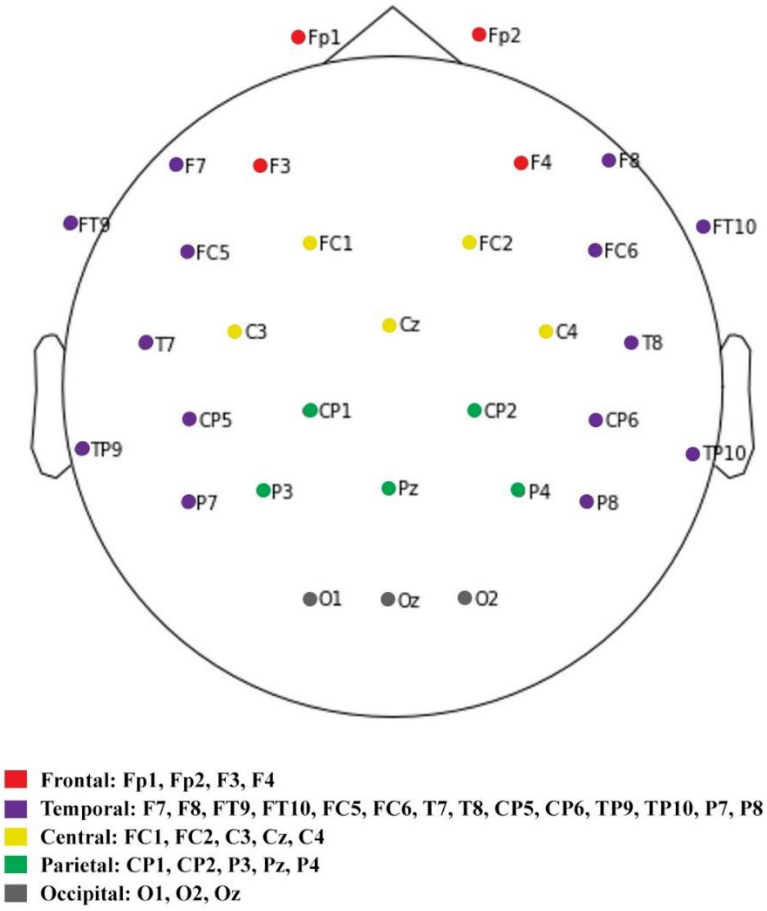


Figure 2.5. Brain functional area allocation [87] with EEG electrodes.

For each trial, the processes of a one-channel EEG with bandpass filtering and its PSD were calculated using the Welch method. The following process was performed using the MNE library in Python [88][89].

Segmentation: It divided the pre-processed EEG signals using the default setting for calculating the PSD using Welch's method in a library, which usually includes overlapping windows. Adjacent windows overlapped by a certain percentage of their length, typically 50%. This allowed for a more accurate estimation of the PSD by reducing variance.

PSD estimation: It used a function for estimating PSD using Welch's method to estimate the PSD of each window of each EEG channel. The default settings for this function included overlapping windows, a Hanning window, and a Welch window size of 250.

Denoting a one-channel EEG from one trial as vector X, the PSD can be determined as follows:

$$p_{psd(\omega)=\frac{1}{n}|\sum_{m=0}^{n-1}X(m)\cdot e^{-jwm}|^2} \quad (2\text{-}1)$$

where X is a vector with n elements, as indicated by the recordings from one channel. Average PSD: The function for estimating the PSD using Welch's method returns an average PSD estimate for each channel in each window. The average PSD was calculated from the PSD estimates across all windows for each channel. Using the process shown in Figure 2.6, it obtained the PSD for each channel in each frequency band under different experimental conditions.

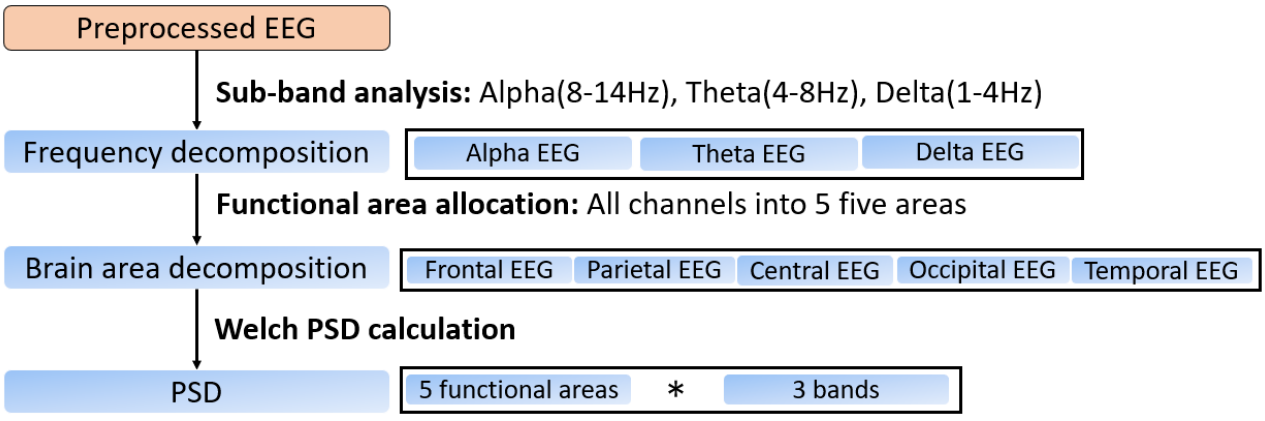


Figure 2.6. Block diagram of spectral EEG analysis.

2.3 Results

2.3.1 Self-Assessment

First, considering the subjective self-assessment of emotion induced by multimodal emotional simulation, in 20 subjects, each person experienced 60 trials of stimulation; the total number of stimulation conditions was 6, 2 (pleasure/unpleasure) × 3 (audio/visual/audio-visual). The self-assessment results for each subject in terms of the experimental conditions are shown in Table 2.2. By calculating the self-assessments of all subjects for the six experimental conditions, it obtained the average self-assessment score [90][91] for each experimental condition shown in Table 2.2.

Table 2.2. Self-assessment results of each subject in different experimental conditions.

	Audio Pleasure	**Visual Pleasure**	**Audio-visual Pleasure**	**Audio Unpleasure**	**Visual Unpleasure**	**Audio-Visual Unpleasure**
Sub01	7.8	8.2	7.4	−5.3	−4.6	−8.3
Sub02	5.2	7.4	6.5	−6.7	−8.3	−6.1
Sub03	6.4	5.5	7.9	−5.0	−6.2	−7.9
Sub04	7.5	4.2	7.8	−8.0	−4.3	−6.8
Sub05	6.0	5.2	7.1	−5.4	−5.2	−7.5
Sub06	5.6	8.0	8.5	−7.0	−3.8	−6.4
Sub07	6.5	7.1	3.8	−6.6	−4.1	−6.7
Sub08	5.2	5.1	6.1	−7.9	−6.7	−8.5
Sub09	7.9	7.7	7.7	−8.5	−5.6	−6.9
Sub10	3.4	5.3	8.2	−5.9	−6.3	−5.1
Sub11	7.2	8.3	7.8	−4.4	−3.6	−7.9
Sub12	7.2	7.4	7.5	−2.7	−7.3	−6.3
Sub13	4.8	6.0	6.0	−8.3	−7.2	−7.0
Sub14	7.6	3.9	8.2	−6.2	−3.3	−6.1
Sub15	6.9	4.7	6.6	−4.8	−8.6	−7.1
Sub16	6.9	7.7	7.0	−4.6	−8.1	−6.3
Sub17	4.7	4.7	8.0	−6.6	−7.0	−7.8
Sub18	6.6	5.2	7.4	−6.2	−5.2	−7.9
Sub19	8.3	7.4	8.0	−8.4	−6.3	−7.5
Sub20	8.2	3.4	7.0	−4.8	−6.4	−5.8
Average	**6.50 ± 2.40**	**6.17 ± 2.51**	**7.23 ± 1.91**	**−6.17 ± 2.35**	**−5.90 ± 2.30**	**−7.00 ± 2.02**

It is referred to pleasure as PL and unpleasure as UNPL. The results of the self-assessment of all subjects in the six conditions are audio pleasure = 6.50 ± 2.40; visual pleasure = 6.17 ± 2.51; audio-visual pleasure = 7.23 ± 1.91; audio unpleasure = −6.17 ± 2.35; visual unpleasure = −5.91 ± 2.30; and audio-visual unpleasure = −7.00 ± 2.02, respectively. The result of two-way repeated ANOVA [92] analysis indicated that there is a significant difference in Emotion (there is a significant difference between pleasure and pleasure $p = 8.12 * 10^{-14} < 0.01$), which indicated as *** in Figure 2.7. No significant (NS) difference in Modality ($p = 0.9868 > 0.1$). This means that subjected SAM could be recognized as pleasure or unpleasure to the same extent regardless of Modality, the results had been indicated in Figure 2.7

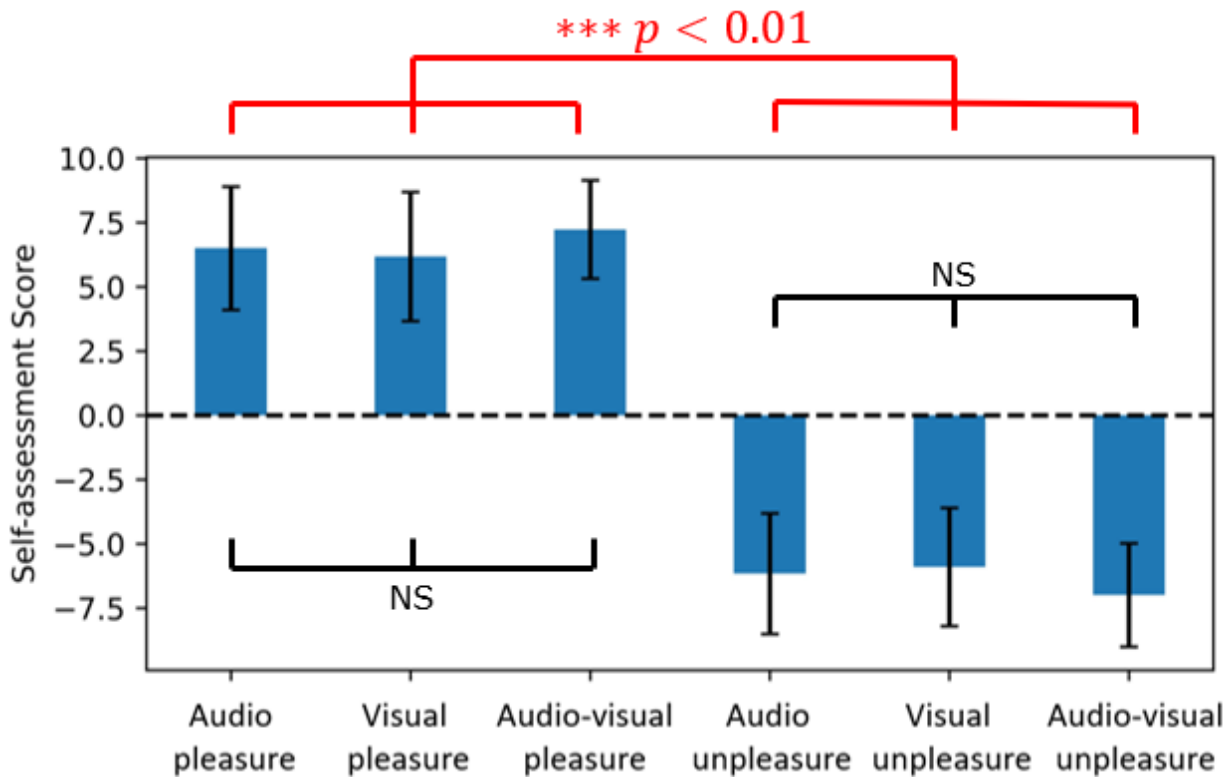


Figure 2.7. ANOVA analytic results of self-assessment data

Next, it calculated the relative difference between the two pairs of conditions within the six conditions. The p-values from the one-way ANOVA [92][93], were as follows: for audio pleasure and visual pleasure, $p = 0.48 > 0.1$, indicating no significant difference as NS in Figure 2.7; for audio unpleasure and visual unpleasure, $p = 0.61 > 0.1$, indicating no significant difference; for audio pleasure and audio-visual pleasure, $p = 0.06 < 0.1$, indicating a significant difference as NS in Figure 2.7; for visual pleasure and audio-visual pleasure, $p = 0.02 < 0.1$, indicating a significant difference; for audio unpleasure and audio-visual unpleasure, $p = 0.01 < 0.1$, indicating a significant difference; and for visual pleasure and audio-visual pleasure, $p = 0.02 < 0.1$, indicating a significant difference.

According to the results of ANOVA, analysis were concluded that the audio and visual conditions have no significant difference ($p > 0.1$) for the two emotional states of pleasure and unpleasure. However, the audio-visual condition was significantly different ($p < 0.1$) from audio- and visual-only conditions.

The results also suggest that a combination of multiple sensory modalities (audio-visual modality) can have a greater impact on mental states than stimulation from a single sensory modality (audio only or visual only) in view of subjective assessments.

2.3.2 Comparison of PSD between resting EEG and multimodal stimulation EEG

To explore the neurological effects of multimodal emotional stimulation in the six conditions, it first studied the PSD during stimulated and pre-stimulated phases. This study chose the start of the stimulation as 0 s, an EEG of 0~30 s for 3000 ms EEG in the stimulated phase, and the 500 ms EEG of −5~0 s for the pre-stimulated phase. The stimulus EEG and pre-stimulus EEG were transformed by epoching original EEG data. The changes before and after stimulation were compared. The extracted stimulus and pre-stimulus EEG datasets were preprocessed, and the PSD of the band-divided brain functional area was calculated to obtain the pre-stimulated and stimulated PSD in decibels. It calculated the percentage change in PSD. For each frequency band interval corresponding to each experimental stimulus condition and functional area, it used the following formula:

$$Change\ of\ percentage_{PSD} = \frac{stimulated_{PSD} - prestimulated_{PSD}}{|prestimulated_{PSD}|} \times 100\% \quad (2\text{-}2)$$

where the $prestimulated_{PSD}$ had a negative value, and it used its absolute value for correction. The topography PSD calculation for 20 subjects were described in Figure 2.8. For statically analysis for the data of 20 subjects, it calculated their respective results and used a paired one-tailed t-test to assess statistical significance. It considered results with a p-value < 0.1 as significantly different and denoted 0.05 < p < 0.1 as *, 0.01 < p < 0.05 as **, and 0.01 > p as ***. The results are shown in Table 2.3.

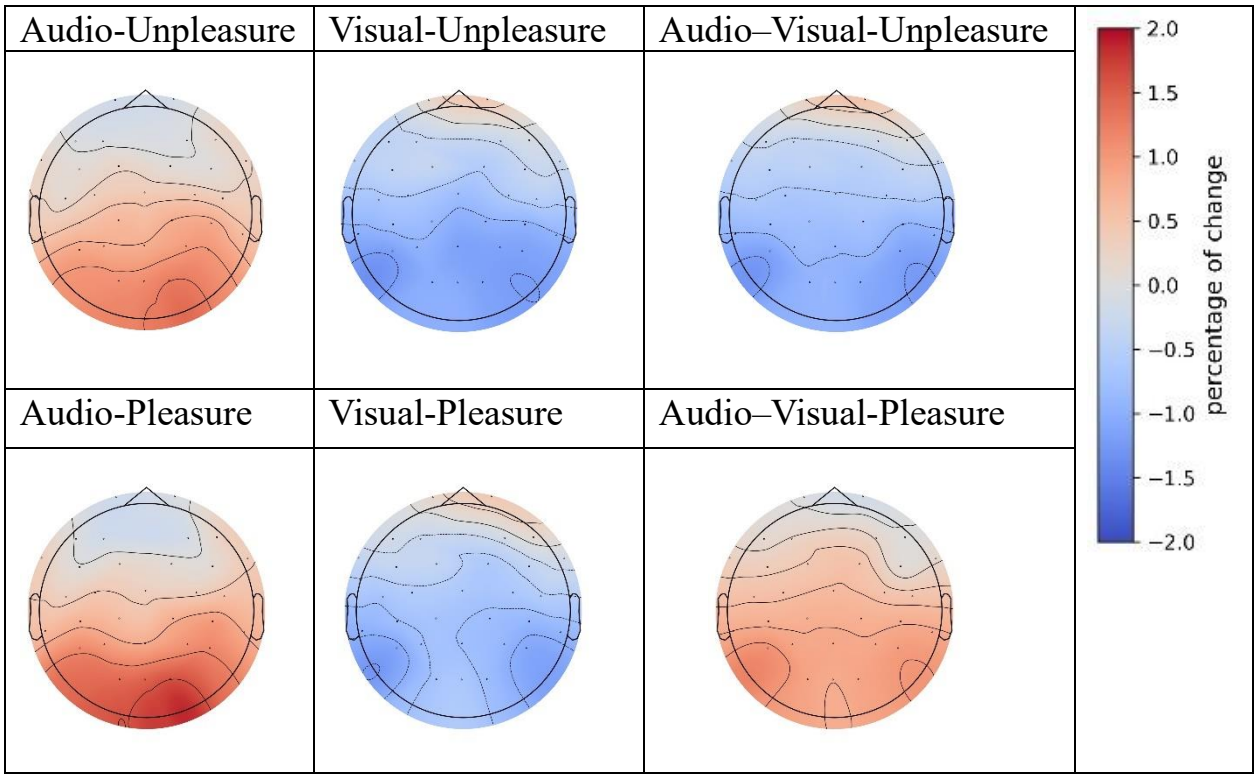


Figure 2.8. Topography distribution for change in PSD with pre-stimulated/stimulated conditions.

Table 2.3. Percentage changes in PSD with pre-stimulated/stimulated conditions, considering frequency bands and functional areas with p-values.

`	Band	Change in Percentage (%)				
		Frontal	Temporal	Central	Parietal	Occipital
Audio Unpleasure	Delta					
	Theta					
	Alpha					2.177^{*} (0.0745)
Visual Unpleasure	Delta					
	Theta					
	Alpha		-1.814^{**} (0.0434)		-2.449^{**} (0.039)	-2.845^{**} (0.020)
Audio-visual Unpleasure	Delta					
	Theta					
	Alpha		-1.514^{**} (0.0418)		-1.867^{**} (0.0412)	-2.176^{**} (0.0218)
Audio Pleasure	Delta					
	Theta					
	Alpha		1.465^{*} (0.0818)			2.696^{**} (0.0374)
Visual Pleasure	Delta					
	Theta					
	Alpha		-1.347^{*} (0.0506)		-1.759^{**} (0.0470)	-1.876^{**} (0.0382)
Audio-visual Pleasure	Delta					
	Theta					
	Alpha		0.359^{**} (0.0236)	0.727^{*} (0.0740)	0.473^{**} (0.0127)	0.070^{**} (0.0113)

There was a statistically significant difference between the stimulated and pre-stimulated PSD only in the frequency domain of the alpha wave and none in those of the delta and theta bands. Therefore, it chose only the alpha frequency domain and drew the topography of $Change\ of\ percentage_{PSD}$. The six conditions were as follows.

The occipital PSD increased by 2.177 in the presence of both auditory stimulation and unpleasure. Temporal, parietal, and occipital PSDs decreased by 1.814, 2.449, and 2.845, respectively, under visual stimulation and unpleasure moods. The PSD in the temporal, parietal, and occipital lobes decreased by 1.514, 1.867, and 2.176, respectively, under audio-visual stimulation and unpleasure moods. Temporal and occipital PSDs increased by 1.465 and 2.696, respectively, under auditory stimulation and pleasure. Temporal, parietal, and occipital PSDs decreased by 1.347, 1.759, and 1.876, respectively, under visual stimulation and pleasure. Under

audio-visual stimuli and pleasure emotions, parietal, parietal, and occipital PSDs increased by 0.359, 0.473, and 0.07, respectively. These results suggest that changes in the PSD in the brain are influenced by sensory and emotional factors and that the effects of these factors vary by brain region and specific sensory and emotional conditions.

In the pre-stimulated and stimulated conditions for comparison, only the alpha wave showed a statistically significant difference among the selected bands. This indicates that different frequency bands in the EEG are related to different brain processes. Alpha is associated with attention and relaxation, and changes in alpha activity are associated with different cognitive and emotional states [94]. Thus, PSD changes in the alpha band may be more sensitive to sensory and emotional conditions than those in the other bands, leading to the observations. The results show that changes in the PSD of the brain under different sensory and emotional conditions are complex and multifaceted. In terms of sensory conditions, the results suggest that auditory stimuli have different effects on the brain than visual stimuli. For example, PSD in the temporal and occipital regions increased in response to auditory stimulation, whereas PSD in the temporal, parietal, and occipital regions decreased in response to visual stimulation. In response to audio-visual stimuli, the PSD increased in the temporal regions but decreased in the parietal and occipital regions. These results suggest that different sensory inputs have different effects on brain function and that the combination of auditory and visual stimuli can produce unique patterns of PSD changes. In terms of emotional conditions, the results suggest that unpleasure and pleasure emotions affect the brain differently. Under the unpleasure condition, the PSD was reduced, with the greatest reduction in the occipital regions. In contrast, in the pleasure condition, the PSD increased, with the greatest increase in the occipital region. These results suggest that

emotions have effects on brain function and that different emotions can produce different patterns of PSD changes.

2.3.3 Emotion-related changes PSD analysis results

To explore the role of the emotional part of the stimulus in this experiment, it performed emotion-related PSD analyses. This study divided the control group into pleasure and unpleasure for each modality stimulation so that it could obtain a total of three pairs of comparisons: Audio Pleasure vs. Audio Unpleasure, Visual Pleasure vs. Visual Unpleasure, and Audio-Visual Pleasure vs. Audio-Visual Unpleasure. This study chose the start of stimulation as 0 s and an EEG of 0~30 s for 3000 ms in the stimulated phase for both pleasure and unpleasure conditions. The PSD of each functional area was calculated for the three theta, beta, and alpha bands. The percentage change was used to measure the change in the pleasure state compared to the unpleasure state. The calculation was as follows in equation (2-3):

$$Change\ of\ percentage_{PSD} = \frac{Pleasure\ stimulated_{PSD} - Unpleasure\ stimulated_{PSD}}{|Unpleasure\ stimulated_{PSD}|} \times 100\% \qquad (2\text{-}3)$$

Since the calculated Unp$leasure\ stimulated_{PSD}$ had a negative value, it used its absolute value to ensure the correction of results. The topography PSD calculation for 20 subjects were described in Figure 2.9. It calculated the respective results for the data from the 20 subjects and used the one-tailed t-test for statistical significance. Results with a p-value < 0.1 were considered statistically significant, and it denoted 0.05 < p < 0.1 as *, 0.01< p < 0.05 as **, and 0.01 > p as ***. The calculated results are shown in Table 2.4.

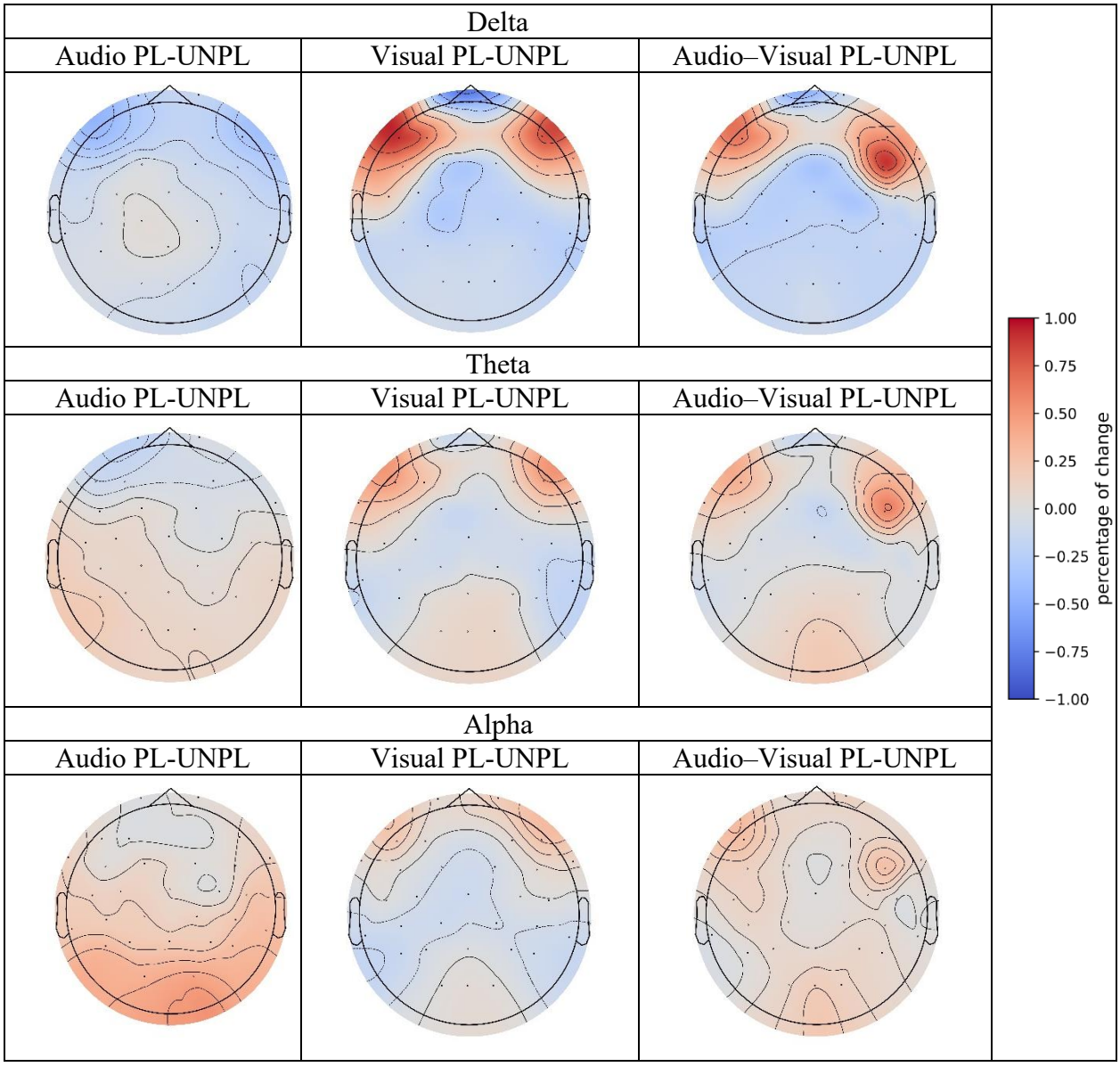


Figure 2.9. Topography distribution for changes in PSD with different emotional conditions.

Table 2.4. Changes in PSD with emotional status, considering frequency bands and functional areas.

Stimulus Types	Band	Change in Percentage (%)				
		Frontal	Temporal	Central	Parietal	Occipital
Audio PL-UNPL	Delta			-0.171^{**} (0.0221)		
	Theta		0.148^{*} (0.098)	0.174^{**} (0.046)	0.221^{**} (0.026)	-0.234^{**} (0.0477)
	Alpha		0.411^{***} (0.008)		0.362^{**} (0.024)	0.699^{***} (0.0016)
Visual PL-UNPL	Delta		-0.415^{**} (0.011)		-0.253^{*} (0.0796)	
	Theta		0.134^{*} (0.0956)			0.238^{**} (0.0596)
	Alpha				0.269^{**} (0.0378)	0.104^{**} (0.0240)
Audio-visual PL-UNPL	Delta	-0.4942^{**} (0.0449)		-0.443^{***} (0.0032)	-0.4889^{***} (0.027)	-0.432^{**} (0.011)
	Theta		0.123^{**} (0.0132)		0.104^{**} (0.0163)	
	Alpha		0.0492^{**} (0.0331)	0.031^{***} (0.0014)	0.034^{**} (0.041)	0.130^{***} (0.0018)

It observed that the results of the delta band were significantly different from those of theta and alpha. It concluded that in the delta band and parietal and occipital brain regions, the pleasure PSD was greater than the unpleasure PSD. However, in the theta and alpha bands and the parietal and occipital regions, pleasure PSD was lower than unpleasure PSD. The influence of each sensory stimulation modality differed among the three bands. The difference between theta with alpha patterns is evident from the topography diagram. Although the changing trend of theta is approximately the same as that of alpha, the degree of change in alpha is greater than that of theta; that is, alpha is more sensitive to changes in emotions.

Combining the results of the t-test and functional brain area calculations, it can observe the following. Delta band: The PSD of the delta wave changed under the audio and audio-visual modalities; however, no similar change was observed under the visual modality. In the audio

modality, the PSD of delta waves did not change significantly in the frontal and central regions but was slightly reduced by 0.171% in other regions. These results suggest that delta waves may be involved in emotional processing in certain contexts. Theta band: The PSD of the theta waves changed in all audio, visual, and audio-visual modalities, which in the audio modality, the PSD increased in the temporal, central, and parietal regions (0.148%, 0.174%, and 0.221%, respectively) and decreased in the occipital region (−0.234%). Under visual conditions, the PSD increased in the temporal and parietal regions (0.134% and 0.238%, respectively) but did not change significantly in other regions. In the audio-visual modality, the PSD increased in the central and temporal regions (0.123% and 0.104%, respectively). These results suggest that theta waves are sensitive to emotional states. Alpha band: The PSD of the alpha band changed in all audio, visual, and audio-visual modalities which the audio modality, the PSD increased in the central and parietal regions (0.362% and 0.699%, respectively) but did not change significantly in the temporal regions. In the visual mode, the PSD increased (0.269%) in the occipital region and decreased (−0.104%) in the temporal region, with no significant changes in other regions. In audio-visual condition, the PSD increased in the central and occipital regions (0.0492% and 0.034%, respectively) and decreased in the temporal and parietal regions (−0.031% and −0.130%, respectively).

2.3.4 Sensory-related changes PSD analysis results

Next, it focused on the influence of the sensory modality on PSD in the frequency domain. It used multimodality stimuli (audio-visual) and single-modality stimuli (audio and visual) in the experiment. This study would like to compare the changes in PSD between the multimodality stimulus and the two single-modality stimuli. Therefore, it performed comparisons in four categories: audio-visual unpleasure vs. audio unpleasure (AV-A-UNPL), audio-visual pleasure vs. audio pleasure (AV-A-PL), audio-visual unpleasure vs. visual unpleasure (AV-V-UNPL), and audio-visual pleasure vs. visual pleasure (AV-V-PL). Here, it used the percentage change to evaluate the change in PSD as well, which evaluated the change in PSD after adding additional stimuli (multimodal) compared with a single modality, which is calculated as follows in equation (2-4).

$$Change\ of\ percentage_{PSD} = \frac{multistimulated_{PSD} - Unistimulated_{PSD}}{|Unistimulated_{PSD}|} \times 100\% \quad (2\text{-}4)$$

Since the calculated $unitimulated_{PSD}$ had a negative value, it used its absolute value to ensure the correction of results. The topography PSD calculation for 20 subjects were described in Figure 2.10. For the data of 20 subjects, it calculated their respective results and used the one-tailed t-test to assess statistical significance. Results with a p-value < 0.1 were considered statistically significant, and it denoted 0.05 < p < 0.1 as *, 0.01< p < 0.05 as **, and 0.01 > p as ***. The calculated results are shown in Table 2.5.

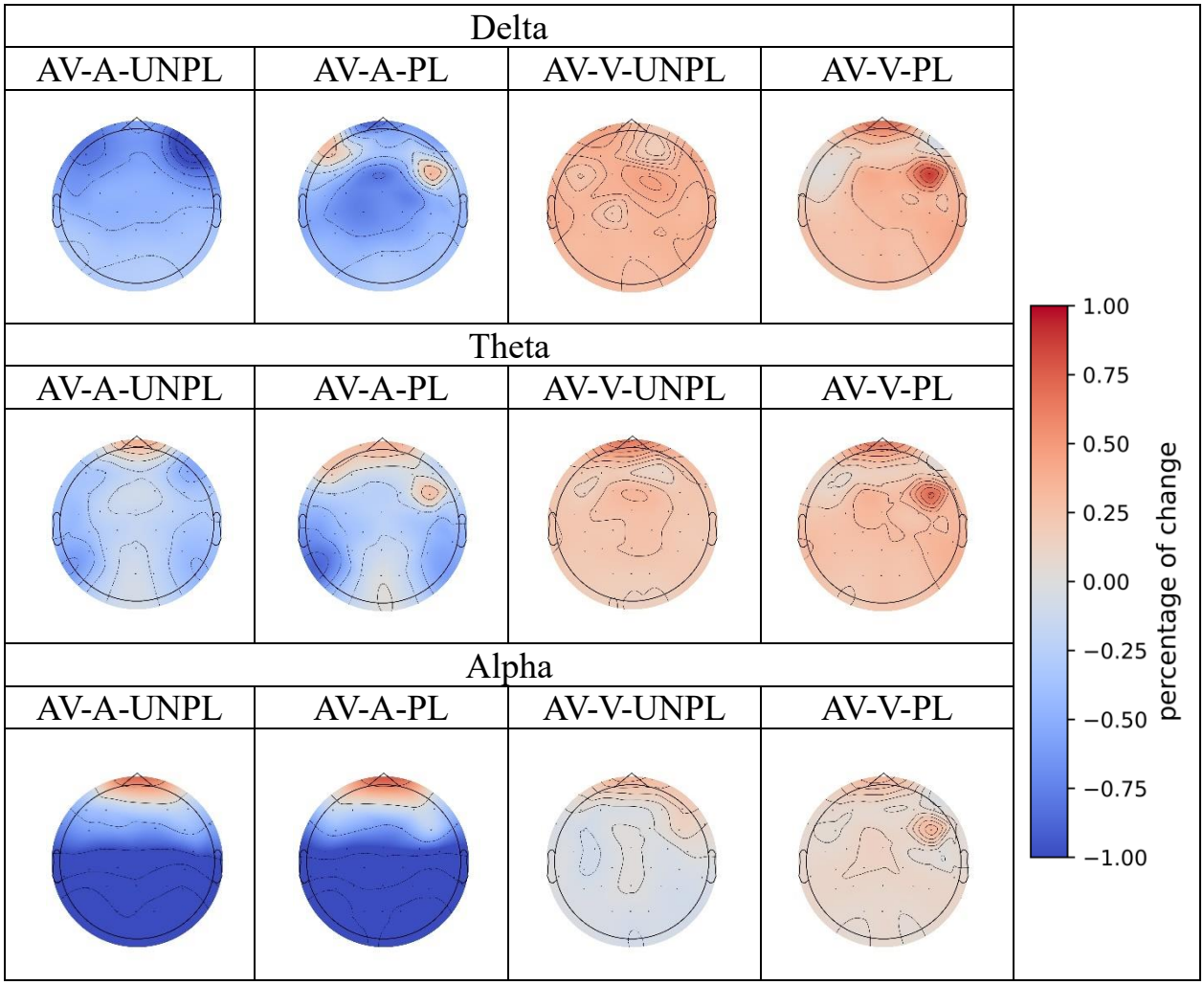


Figure 2.10. Topography distribution for the change in PSD with multi-stimulated/uni-stimulated pleasure/unpleasure conditions.

Table 2.5. Changes in PSD with respect to multi-stimulated/uni-stimulated and pleasure/unpleasure conditions, considering frequency bands and functional areas.

Stimulus types	Band	Change in Percentage (%)				
		Frontal	Temporal	Central	Parietal	Occipital
AV-A UNPL	Delta	-1.205^{***} (0.010)	-1.055^{***} (0.009)	-1.022^{***} (0.002)	-1.014^{***} (0.002)	-0.813^{***} (0.008)
	Theta	0.607^{*} (0.063)			0.369^{**} (0.040)	0.758^{***} (0.003)
	Alpha	0.309^{*} (0.095)	-2.752^{***} (0.0001)	-1.821^{***} (0.0001	-3.382^{**} (0.0001)	-4.441^{***} (0.0002)
AV-V UNPL	Delta		0.693^{***} (0.001)	0.633^{***} (0.003)	0.559^{***} (0.008)	0.605^{***} (0.004)
	Theta	0.838^{***} (0.0007)	0.369^{***} (0.0007)	0.380^{***} (0.0026)	0.362^{***} (0.002)	0.300^{**} (0.0027)
	Alpha	0.234^{**} (0.020)	-0.141^{**} (0.030)	-0.141^{*} (0.075)	-0.182^{**} (0.045)	-0.198^{**} (0.040)
AV-A PL	Delta	-1.528^{**} (0.020)	-0.721^{*} (0.082)	-1.309^{***} (0.004)	-1.381^{***} (0.003)	-1.088^{**} (0.011)
	Theta		0.123^{**} (0.0132)			0.768^{**} (0.030)
	Alpha		-3.1379^{***} (0.0001)	-1.942^{***} (0.0004	-3.811^{**} (0.0002)	-5.060^{***} (0.0001)
AV-V PL	Delta	0.525^{*} (0.095)	0.408^{*} (0.054)	0.306^{*} (0.095)	0.335^{*} (0.080)	0.405^{**} (0.043)
	Theta		0.359^{***} (0.0017)	0.306^{**} (0.014)	0.311^{***} (0.004)	0.313^{***} (0.005)
	Alpha	0.197^{*} (0.077)				

Therefore, it combined the previous results to arrive at the following analysis.

Delta (1–4 Hz):In the unpleasure state, adding visual to audio stimuli led to a statistically significant decrease in delta activity across all brain regions ranging from 0.8126% to 1.2049%, respectively. Adding audio to visual stimuli led to a statistically significant increase in use activity in the central, parietal, and occipital regions, ranging from 0.5585% to 0.6931%, respectively. In the pleasure state, adding visual to audio stimuli led to a statistically significant decrease in activity in the central, parietal, and occipital regions ranging from 1.3092% to 1.5284%. Adding audio to visual stimuli led to a statistically significant increase in activity in the frontal, temporal, and occipital regions ranging from 0.3061% to 0.5248%, respectively. Theta (4–8 Hz):In the unpleasure state, adding visual to audio stimuli led to a statistically significant increase in theta activity in the frontal and parietal regions, ranging from 0.1514% to 0.7576%, respectively, with a statistically significant p-value of less than 0.05 in the occipital region. Adding audio to visual stimuli led to a statistically significant increase in activity in all regions

ranging from 0.2996% to 0.8381%, respectively. In the pleasure state, adding visual to audio stimuli led to a statistically significant increase in activity in the frontal and parietal regions ranging from 0.2488% to 0.7675%, respectively. Adding audio to visual stimuli led to a statistically significant increase in activity in the temporal and occipital regions ranging from 0.3063% to 0.4241%, respectively. Alpha (8–13 Hz):In the unpleasure state, adding visual to audio stimuli led to a statistically significant increase in alpha activity in the frontal, parietal, and occipital regions ranging from 0.3093% to 4.4406%, respectively. Adding audio to visual stimuli led to a statistically significant increase in activity in all regions ranging from 0.1413% to 0.2344%, respectively. In the pleasure state, adding visual to audio stimuli led to a statistically significant increase in activity in the frontal and parietal regions ranging from 0.2535% to 5.0590%, respectively. Adding audio to visual stimuli led to a statistically significant increase in activity in the temporal and occipital regions ranging from 0.0259% to 0.4241%, respectively

2.3.5 Emotion and sensory-related statistical analysis result

From the results of PSD changes mentioned in 2.3.3 and 2.2.4, sensory patterns and emotion patterns under multimodal stimulation correspond with sensory-related changes and emotion-related changes. The numeric and topographic results indicated the changes in each decomposed band and area. To further explore comparative characteristics of emotion and sensory patterns of EEG under multimodal stimulation, emotion/sensory-related statistical analysis with respect to spectral, spatial, and quantitative aspects will be investigated in this part.

The spatial characteristics of emotion and sensory patterns comparison with respect to "Within area standard deviation (std) of PSD changes." Standard deviation is a viable metric for evaluating EEG features and indicating EEG characters [95]. Previously obtained results of PSD changes could be viewed from each of the five brain functional areas, standard deviation naturally represented the degree of variation within a group of data. Hence, the difference in spatial representation is available to be revealed by calculating the "Within area std of PSD changes.", which could be obtained from PSD changes with normalization, and calculating std within five functional areas. Within area std of PSD changes resulting from sensory and emotion are comparable for spatial characteristics of emotion and sensory patterns. In Figure 2.11, the results of within area std of PSD changes results from sensory and emotion are represented for Minimum: the position of the lower end of the box, lower quartile: the lower limit of the bottom edge of the box; median: the line in the middle of the box; upper quartile: the upper limit of the top edge of the box; maximum: the position of the upper end of the box, dot is abnormal data.

The results in Figure 2.11 showed that Within bands std as mean: emotion (0.12) > sensory (0.07) with significant difference. It indicated that the degree of variation in five areas which corresponds to emotion-related changes, were greater than sensory related changes.

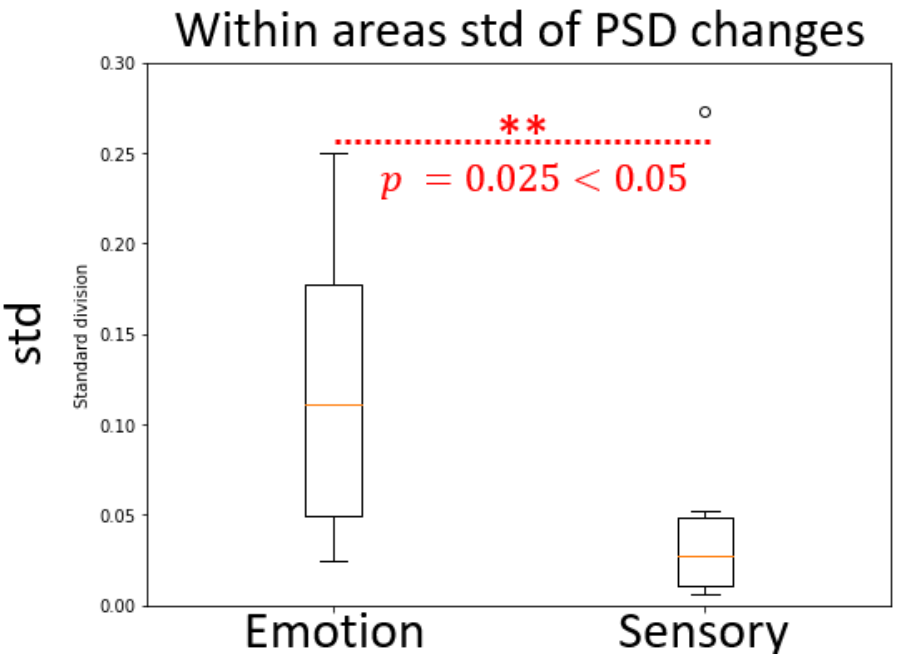


Figure 2.11. Within area std of PSD changes, left is emotion-related changes and right is sensory-related changes. Statistical significance from emotion and sensory were obtained as p-value.

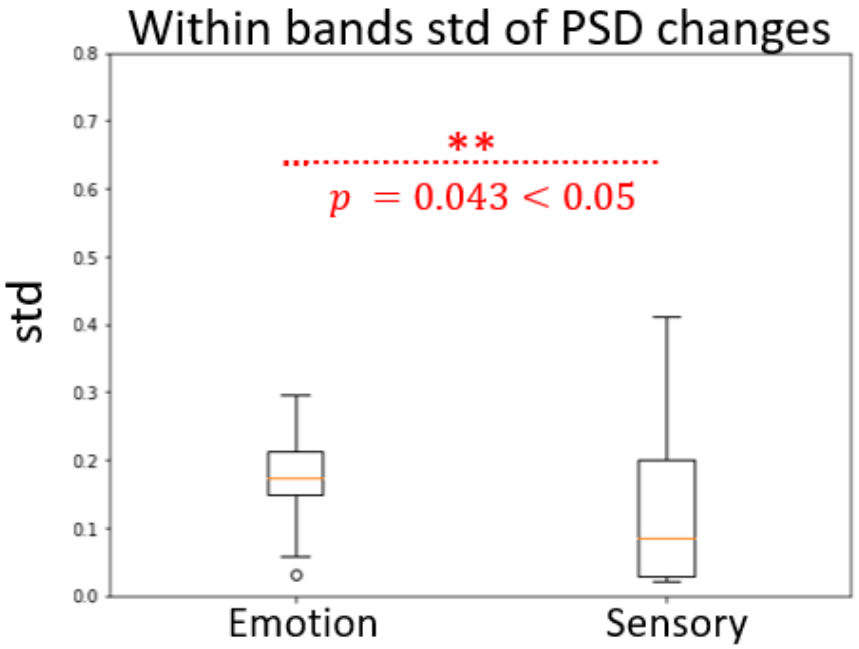


Figure 2.12. Within band std of PSD changes, left is emotion-related changes, right is sensory-related changes. Statistical significance from emotion and sensory were obtained as p-value.

The spectral characteristics of emotion and sensory patterns comparison with respect to "Within band standard deviation (std) of PSD changes." Previously obtained results of PSD changes could be viewed from each of the three bands, standard deviation naturally represented the degree of variation within a group of data. Hence, the difference in spectral characteristics is available to be revealed by calculating the "Within band std of PSD changes.", which could be obtained from PSD changes with normalization, and calculating std within three bands. Within band std of PSD changes results from sensory and emotion are comparable for spectral characteristics of emotion and sensory patterns. In Figure 2.12, the results of within band std of PSD changes results from sensory and emotion are represented.

The results in Figure 2.12 showed that Within bands std as mean: emotion (0.17) > sensory (0.13) with significant difference. It indicated that the degree of variation in three bands which corresponds to emotion-related changes, were greater than sensory related changes.

Quantitative comparison of emotion and sensory patterns, which refer to the emotion-related changes and sensory-related changes of EEG under multimodal stimulation. The metrics of quantitative comparison is the sum of absolute value of significant changes, aim to quantitatively represent statistical significance validated PSD changes. Each "Sum of absolute value" is from five functional allocated areas in each band and each band in each sensory or emotion analysis pair would have one "Sum". The results are represented in Figure 2.13.

The results in Figure 2.13 showed that mean of each emotion pair(2.06) < sensory pair(12.08). It indicated that PSD changes from sensory-related changes are quantitative greater than emotion related changes.

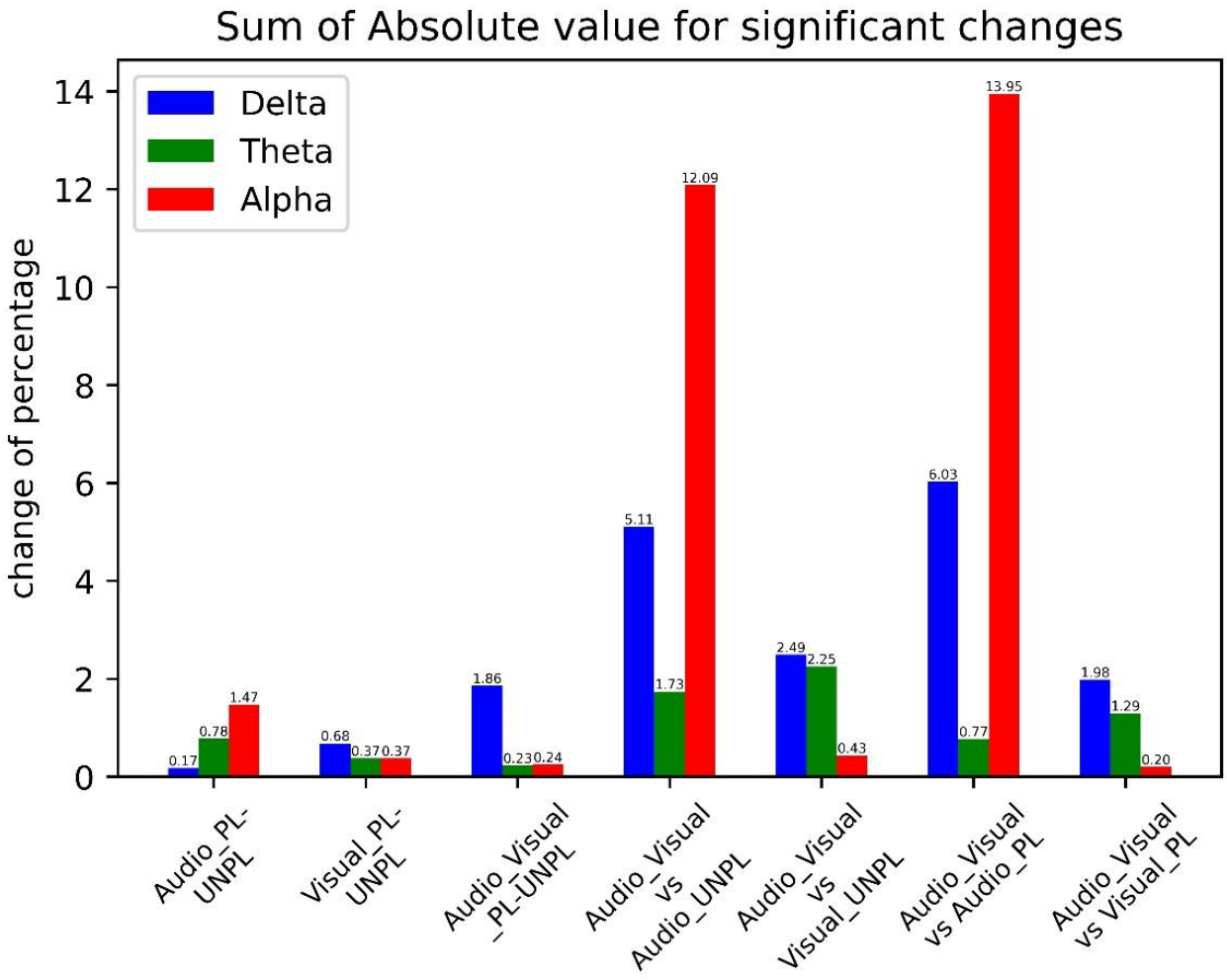


Figure 2.13. Sum of the absolute value for significant changes corresponding to emotional change pairs and sensory change pairs. The three on the left are emotional changes, and the four on the right are sensory-related changes.

2.4 Discussion

As for the emotion-related changes, based on the results in Table 2.4 and Figure 2.9, results indicated that the change in emotional state had a significant impact on the change in theta activity. The PSD in the central and parietal regions significantly increased under the unpleasure condition, while the theta activity in the temporal and parietal regions decreased under the pleasure condition. This suggests changes in theta activity, which is the demonstration of emotional processing in the brain. The specific regions affected by changes in theta activity may vary according to the emotional state. Furthermore, the results indicated that the change in emotional state also had a significant impact on the change in alpha activity. The PSD of the occipital region significantly increased with respect to the pleasure emotion, whereas the alpha activity of the temporal region significantly decreased with respect to the unpleasure emotion. These results suggest that alpha activities may be involved in emotional processing, particularly in the occipital region, which is thought to be involved in visual processing. The specific regions affected by changes in alpha activity may depend on emotional state and sensory input [96].Findings in the delta band suggest that changes in the emotional state may be less sensitive to changes in the delta activity than in other bands. The PSD in the frontal regions was significantly lower in the pleasure emotion, suggesting that changes in delta activity may be related to emotional processing within this brain region [97]. The specific regions affected by changes in delta activity may vary according to the emotional state and sensory input. Thus, the effect of emotion on PSD could be discussed in that the analysis found that the emotional state of the participants had a significant impact on PSD, with all three frequency bands (delta, theta, and alpha) showing changes in activity that were dependent on the emotional state. Emotion perception shows serial dependence within and between sensory modalities [98]. This suggests

that emotions can modulate neural activity in specific brain regions and frequency bands, which possibly have implications for understanding emotional processing and regulation.

As for the changes in PSD in response to emotional stimuli across different frequency bands, focusing on the emotional states of pleasure and unpleasure. This explored the changes in PSD across different sensory inputs and functional brain regions to better understand the interconnection between emotional states and brain functions. The findings suggest that different frequency bands exhibit different patterns of change in response to emotional stimuli and that the specific affected regions also vary relative to the emotional state and sensory input. As for the sensory-related changes, based on the results in Table 2.5 and Figure 2.10, the three bands showed significant changes in activity when audio and visual stimuli were combined compared to when they were presented separately. In addition, all three bands showed changes in activity that were dependent on the emotional state of participants. Moreover, some differences were observed between bands. Delta activity consistently decreased when visual stimuli were added to audio stimuli across all regions and emotional states. Theta activity significantly increased in the frontal and parietal regions when audio stimuli were combined with visual stimuli, regardless of the emotional state. The alpha activity showed more complex changes, with different regions and emotional states showing different patterns. These results suggest that the addition of different sensory inputs can modulate neural activity in specific brain regions. The pattern of changes in neural activity in response to different sensory inputs varies across the different frequency bands and brain regions [99].

2.4.1 Emotion/Sensory-related changes spatial character discussion

Based on the std of PSD changes within the brain functional area and Figure 2.11, the spatial variation of emotion-related changes was larger than that of sensory-related changes. This finding suggests there are differences existing in the way emotions and sensory patterns take the role of influencing brain activity. Previous studies have indicated individual pattern analysis, in EEG of the indication of EEG spatial characters with the emotion-related changes with emotion pattern [100] and sensory-related changes with sensory patterns [101]. The results of this research have more comprehensive results for comparing the spatial characters of emotion and sensory patterns.

Emotion-related changes showed more diverse regional changes. This suggests that the impact of emotion on brain activity pretends to be broader and less localized. From this perspective, the effects of emotion are not limited to specific brain regions but involve broader interactions between regions. This broader influence may reflect the interaction between emotion and cognition, emotion and emotion regulation, which in turn affects the overall activity pattern of the brain [102]. In contrast, the spatial characteristics of sensory-related changes are relatively consistent and localized. This may indicate the relatively fixed neural circuits and pathways in the brain for sensory information processing [103].

2.4.2 Emotion/Sensory-related changes frequency-varied character discussion

According to the std of the power spectral density (PSD) changes within the frequency band and Figure 2.12, it found that the frequency change characteristics of emotion-related changes were larger than sensory-related changes. The results of spatial characters comparison suggest that emotions and senses take different ways of modulating brain activity distinctly in different frequency ranges. Previous studies have shown the individual influence of sensory and emotion patterns from the viewpoint of different frequency bands, frequency change characteristics the neurological feedback as EEG, the emotion-related changes [104] and sensory-related changes [105] had been discussed, respectively. Moreover, this study gives a further explanation of sensory and emotion patterns' spectral character under multimodal stimulation.

Emotion-related changes showed more diversified variation in spectral rather than sensory-related changes. This suggests that emotion regulation of brain activity frequency pretends to be more diverse. The role of emotion may involve brain oscillations in different frequency ranges and reflect that emotional processing involves multiple frequency ranges of the brain [106]. Conversely, the spectral change characteristics of sensory-related changes are relatively consistent. This implies that the frequency regulation of sensory information processing in the brain is relatively fixed and stable. The transmission and processing of sensory information may be more concentrated in specific frequency ranges, which is possibly related to the encoding and processing of perceptual information, provides a certain degree of regularity and reliability for the brain's processing of external sensory input [107].

2.4.3 Emotion/Sensory related changes quantitative discussion

Based on the sum of absolute values of changes and Figure 2.13, it found that the degree of sensory-related changes was significantly higher than that of emotion-related changes. The results from this research indicate that there are clear quantitative differences in the influence of sensory and emotional effects on brain activity. The previous study focused on the influence of emotion or sensory-related changes with EEG under emotion stimulations [104] and sensory stimulations [105] respectively. From the analytic results of this study, the comparative evaluation of sensory and emotion-related changes, in the name of sensory and emotion patterns, had been obtained.

Sensory-related changes showed more significant differences in spectral effects, which had been shown as PSD changes in this research. This suggests the implication that the processing of sensory information has a more direct and significant impact on the spectral distribution of the brain. The priority of the degree from spectral changes of sensory-related changes was higher than that of emotion-related changes, which reflects the important role of sensory information processing in the brain. The inner reception and determination of sensory information involve a wider range of neural networks and these changes show a more quantitative impact. This difference in the degree of changes is possibly related to the efficiency and priority of the brain in processing sensory information [108]. When people look back at the results emotion-related changes were relatively lower degree of changes. This means that emotion has a relatively weak impact on brain spectral contribution. The role of emotion is more reflected in the functional connections of the brain and the regulation of neural networks, rather than directly affecting changes in spectral distribution. The lower degree of emotion-related

changes at spectral changes possibly indicates the moderate role of emotion processing in brain activity and the slow and gradual nature of emotion regulation processes [109].

2.5 Summary

This chapter investigated the impact of emotional states and stimulus modality on brain activity with feature analysis of sensory and emotion pattern under multimodal stimulation. The participants' EEG data were collected under six conditions: two types of emotion (Pleasure) /Unpleasure)* three stimuli modality (Audio /Visual /Audio-visual) with self-assessment metric. Validation of subjective emotion results confirmed Participants can recognize pleasure/unpleasure regardless of types of sensory. Analysis of spectral and spatial characters of EEG with PSD: Sub-band analysis combined with PSD allowed detailed study within specific frequency bands; Functional area allocation combined with PSD determined activities in each functional allocation of brain areas. The results of EEG analysis under multimodal emotion stimulation indicated: Spatial character: Emotion-related changes exhibit greater diversity in regional changes; Spectral character: Emotion-related changes exhibit greater diversity in sub-band; Quantitative character: Sensory-related changes exhibit greater spectral impact quantitatively. Overall, our study provides a comprehensive understanding of the neural mechanisms underlying emotional processing across sensory modalities, mainly via EEG pattern analysis of sensory and emotion. These findings have important implications for understanding the neural mechanisms and gave the feature analytic basement for cross-sensory EEG emotion recognition.

3. Cross-sensory EEG Emotion Recognition

3.1 Introduction

Emotion recognition is identifying and interpreting human emotions based on cues such as facial expressions, vocal intonations, physiological signals, and behavioral patterns [110]. It involves using algorithms and machine learning to analyze these cues and classify them into different emotional states, such as happiness, sadness, anger, or fear. Emotion recognition is applicable in various fields, including psychology, human-computer interaction, and healthcare [111]. EEG is a neuroimaging technique that measures the electrical activity in the brain using electrodes placed on the scalp. It records collective electrical signals generated by neuron firings in the brain. EEG provides valuable insights into brain activity and is widely used in neuroscience research, clinical diagnostics, and BCI systems [3]. EEG signals can be analyzed to detect patterns associated with specific mental states, cognitive processes, and emotional responses. EEG, owing to its non-invasiveness and high sensitivity to various emotions, has recently gained increasing attention as a physiological signal for emotion recognition [112]. The a-BCI system combines the principles of EEG and emotion recognition to create a direct communication pathway between the brain and an external device or computer system. This allows individuals to control external devices or interact with computer systems using their emotional states as inputs. The a-BCI systems use EEG signals to detect and interpret emotional states, translated into commands or actions for connected devices [113].

The current trend in EEG-based a-BCI focuses on improving the accuracy and reliability of emotion recognition using EEG signals [114]. The general concept of EEG a-BCI comprises data acquisition, preprocessing, feature extraction, classification/regression, model evaluation, and emotion recognition. The mean-focused parts are feature extraction and

classification/regression. Recently, various machine learning-inspired methods have been applied in EEG emotion recognition. Liu et al. [115] proposed an EEG emotion recognition model that combines an attention mechanism and a pre-trained convolution capsule network to effectively recognize emotions, enhancing emotion-related information in EEG signals. Li et al. [116] presented the Frontal Lobe Double Dueling Deep Q Network (FLD3QN), a model inspired by the Papez circuit theory and reinforcement learning neuroscience, utilizing EEG signals from the frontal lobe to enhance emotion perception. Padhmashree and Bhattacharyya [117] presented a novel four-stage method for human emotion recognition using multivariate EEG signals, achieving exceptional performance by incorporating deep residual networks and time-frequency-based analysis. Wei et al. [118] proposed that the Transformer Capsule Network (TC-Net) achieves state-of-the-art EEG-based emotion recognition performance on DEAP and DREAMER datasets. This demonstrates its effectiveness in capturing global contextual information through an EEG Transformer and Emotion Capsule modules. Cui et al. [119] integrated signal complexity, spatial brain structure, and temporal context through 4D feature tensors, Convolutional Neural Networks, and Bidirectional Long-Short Term Memory, achieving high accuracy (94% for DEAP and 94.82% for SEED datasets) by deep decoding EEG signals and extracting key emotional features. Algarni et al. [120] developed a deep learning-based approach for EEG-based emotion recognition, contributing to improved accuracy in diagnosing psychological disorders by achieving high accuracies (99.45% valence, 96.87% arousal, 99.68% liking) through a multi-phase process involving data selection, feature extraction, selection, and classification using a stacked bi-directional Long Short-Term Memory (Bi-LSTM) Model. To enhance Human-Computer Interaction, Islam et al. [121] propose PCC-CNN, which is a deep machine-learning-based Convolutional Neural Network (CNN) model for emotion recognition using EEG signals,

where Pearson's Correlation Coefficient featured images are employed for channel correlation analysis in sub-bands. Peng et al. [122] introduced a unified framework, GFIL (graph-regularized least square regression with feature importance learning), for EEG-based emotion recognition and highlighted the significance of the Gamma band and prefrontal/central region channels in emotion recognition. Huang et al. [123] presented an EEG-based emotion detection system that utilized short EEG segments of 1s, incorporating a novel feature extraction algorithm—asymmetric spatial filtering—into a filter bank framework. Chen et al. [124] proposed an emotion recognition method using EEG signals, which involved extracting the energy means of detail coefficients as feature values and using a support vector machine (SVM) for classification. They demonstrated the validity of the feature values and provided a theoretical basis for implementing effective human-computer interaction. Wu et al. [14] highlighted the importance of Riemannian feature extraction in EEG-based emotion recognition, demonstrating that the proposed independent component analysis with Riemannian manifold and long short-term memory networks (ICRM-LSTM) model outperformed existing methods by effectively addressing the uncertain ordering in independent component analysis (ICA). Wang et al. [125] proposed a domain-adaptation symmetric positive definite (SPD) matrix network (daSPDnet) that effectively captured shared emotional representations among different individuals using Riemannian feature extraction. Based on the previous EEG emotion recognition studies, the filter bank was highly robust, and Riemannian feature extraction is an emerging field for effective feature extraction.

Recent research has revealed that multimodal stimulation with cross-sensory emotions is crucial in human-computer interaction and affective computing (AC). Ranasinghe et al. [126] demonstrated that wearable accessories such as head-mounted displays (HMD), with wind and

thermal stimuli, significantly improve sensory and realism factors, enhancing the sense of presence compared to traditional virtual reality (VR) experiences. Zhu et al. [127] explored the emerging field of multimodal sentiment analysis, which integrates text, visual, and audio information to infer sentiment polarity, aiming to provide researchers with insights and inspiration for developing effective models in this field. Calvo and D'Mello [128] provided an overview of recent progress in AC, focusing on affect detection. This highlights the need for an integrated examination of emotional theories from multiple disciplines to develop effective practical AC systems. Wang et al. [129] introduced the multimodal emotion database (MED4), encompassing EEG, photoplethysmography, speech, and facial images for emotion recognition research, demonstrated the superiority of EEG signals in emotion recognition and proposes fusion strategies that combine speech and EEG data to significantly enhance accuracy and robustness.

Tsiourti et al. [130] examined how incongruity in emotional expressions displayed by humanoid robots affects their recognition and response. Their research underscored the negative impact of such incongruence on the robot's appeal and credibility. A retrospective examination of previous studies in this field is necessary to comprehend the role of cross-sensory modalities in EEG-based emotion recognition. A critical limitation of applying the EEG-based aBCI is that the required system training and test data are highly sensory-dependent. Additionally, the current EEG-based aBCI system can only be based on one type of sensation from emotion simulation. For example, if an EEG-based aBCI is trained using EEG data from auditory stimuli, its operation is limited to auditory stimulation.

Similarly, its operation is limited to visual stimulation if it is trained using EEG data from visual stimuli. If an EEG-based aBCI is trained using EEG data from audio-visual stimuli (e.g.,

video stimuli), its operation requires audio-visual stimulation. In real-world applications, emotional stimuli are inherently multimodal [131]. Video stimuli contain auditory and visual modalities, resulting in a compound multimodal representation. When exposed to a partial sensory component of a multimodal emotional stimulus, individuals naturally perceive the same emotional category [132]. For instance, when individuals receive the audio, the visual, or both (audio-visual) components for the same video stimulus, the emotional category should be the same and identifiable using the EEG-based aBCI. Nevertheless, owing to sensory differences, the features extracted from EEG data can vary for different sensory modalities, even if they originate from the same stimulus source (e.g., the same video material). The key challenge in this study was effectively extracting features and mitigating the differences between different sensory modalities in EEG data. The cross-sensory EEG emotions were lacking in earlier research. However, the cross-sensory theme was highly related to transfer learning in EEG-based emotion recognition. This is valuable for tracing the previous comprehensive research on transfer learning in EEG-based emotion recognition. Existing transfer learning research has mainly focused on cross-subject, cross-session, and cross-dataset. Cimtay et al. [133] introduced a novel multimodal emotion recognition system using facial expressions, galvanic skin response (GSR), and electroencephalogram (EEG) data, achieving high accuracy rates and surpassing reference studies in subject-independent recognition. Li et al. [134] presented the self-organized graph neural network (SOGNN) for cross-subject EEG emotion recognition, achieving state-of-the-art performance by dynamically constructing graph structures for each signal. CLISA (Contrastive Learning method for Inter-Subject Alignment), a method developed by Shen et al. [135], leveraging contrastive learning to minimize inter-subject differences and improve cross-subject EEG-based emotion recognition by extracting aligned spatiotemporal representations from EEG

time series. Domain adaptation with adversarial adaptive processes has recently gained increasing attention, achieving state-of-the-art performance. Wang et al. [136] proposed a multimodal domain adaptive variational autoencoder method by learning shared cross-domain latent representations and reducing distribution differences, demonstrating superior performance in emotion recognition with small labeled multimodal data. Guo et al. [137] proposed a multi-source domain adaptation with a spatiotemporal feature extractor for EEG emotion recognition, effectively reducing cross-subject and cross-session, demonstrating its powerful generalization capacity. He et al. [138] proposed a method that combines temporal convolutional networks and adversarial discriminative domain adaptation for EEG-based cross-subject emotion recognition, effectively addressing the domain-shift challenge. Sartipi and Cetin [139] proposed an approach that combined transformers and adversarial discriminative domain adaptation for cross-subject EEG-based emotion recognition, achieving improved classification results for valence and arousal.

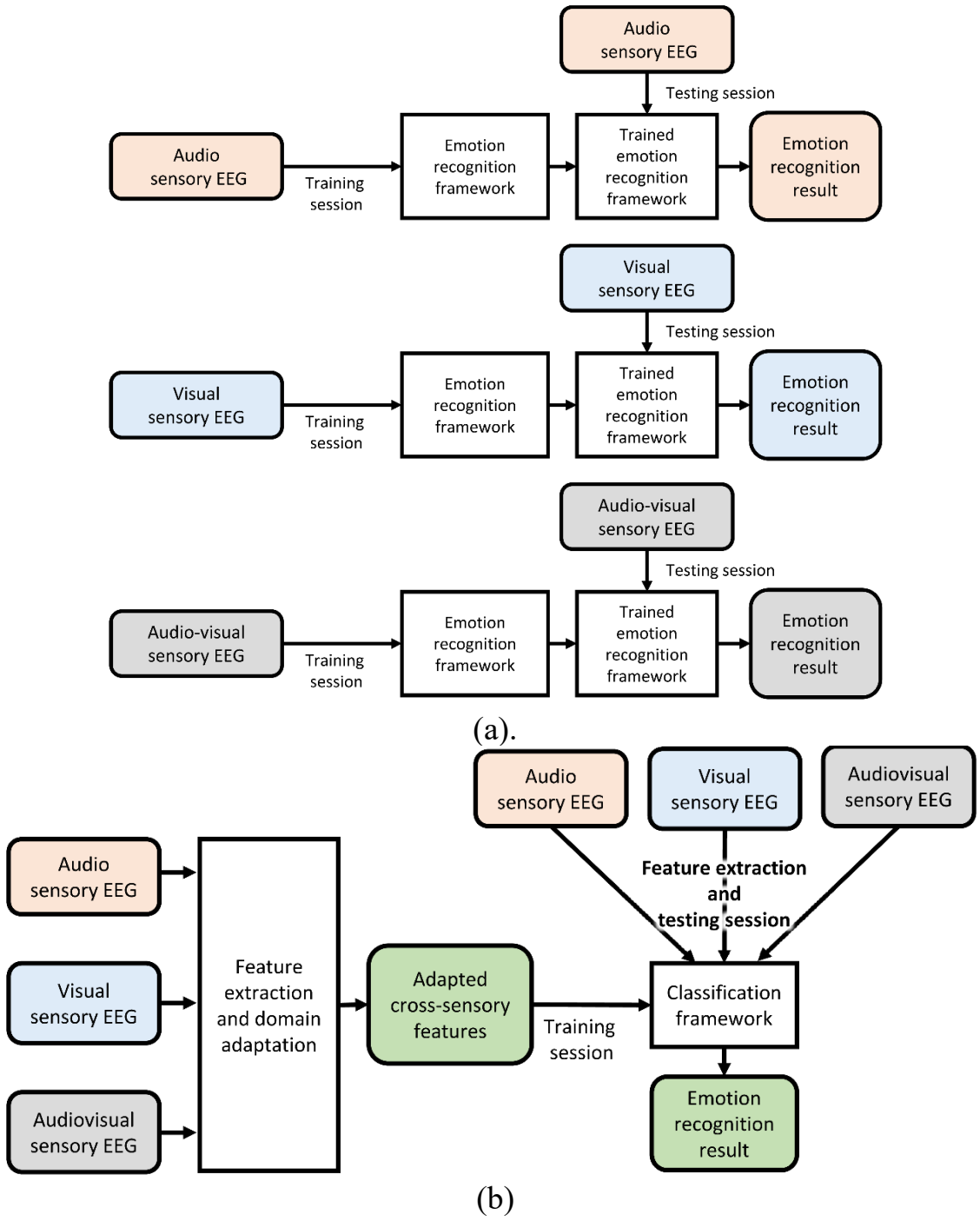


Figure 3.1. Demonstration of previous EEG-based aBCI (a) and the proposed cross-sensory EEG-based aBCI (b). The stimuli source was video stimulation.

This study aimed to effectively extract features and mitigate the differences in EEG data between sensory modalities using domain adaptation techniques. The previous EEG-based aBCI and proposed cross-sensory EEG-based aBCI are illustrated in Figure 3.1. Emotional EEG data can be used to train the aBCI system regardless of the sensory modality. Establishing a robust emotion recognition framework is essential to address this issue.

Hence, inspired by previous research on general emotion recognition, transfer learning-based emotion recognition, and domain adversarial adaptation, this chapter conducted experiments to collect cross-sensory emotion EEG data with a video as the source stimulus. Audio, visual, and audio-visual sensory-inspired EEG data were obtained from 20 participants. Subsequently, filter bank adversarial domain adaptation Riemann methods (FBADR) have been proposed for cross-sensory EEG emotion recognition. Specifically, the proposed methods use filter banks and Riemann methods for feature extraction. An adversarial domain adaptation method inspired by conditional Wasserstein generative adversarial networks was explored for domain adaptation. The classification was conducted using an ensemble of SVMs with a meta-classifier. Experimental results from the proposed FBADR revealed its state-of-the-art performance in cross-sensory emotion recognition with high robustness. This study could also be recognized as pioneering research in cross-sensory emotion recognition with a video stimulus as multimodal emotion stimulation using two categories of emotion: pleasure and displeasure. The results of this study can be further applied to EEG-based aBCI for theoretical analysis and practical applications.

3.2 Materials and Methods

3.2.1 Data acquisition paradigm

Cross-sensory EEG emotion recognition experiments were conducted using EEG data from self-designed multimodal emotion stimulation experiments [140] in Chapter 2. Twenty emotion videos: ten of pleasure and ten of unpleasure–sourced from the New Standardized Emotional Film Database for Asian Culture [78] were explored as multimodal emotion stimulation for inducing cross-sensory EEG emotion data during the experiment. Self-assessments were performed after stimulation in each trial. The participants were twenty healthy participants with no previous mental or physical injury and currently native Chinese speakers. In the experiment, two computers were utilized, one was dedicated to controlling the stimulation process, while the other is used for recording EEG data and manually inspecting the recorded information. As the stimulation control window could potentially interfere with the window for EEG data recording, it employ two separate computers for these tasks to ensure the smooth progress of the experiment.

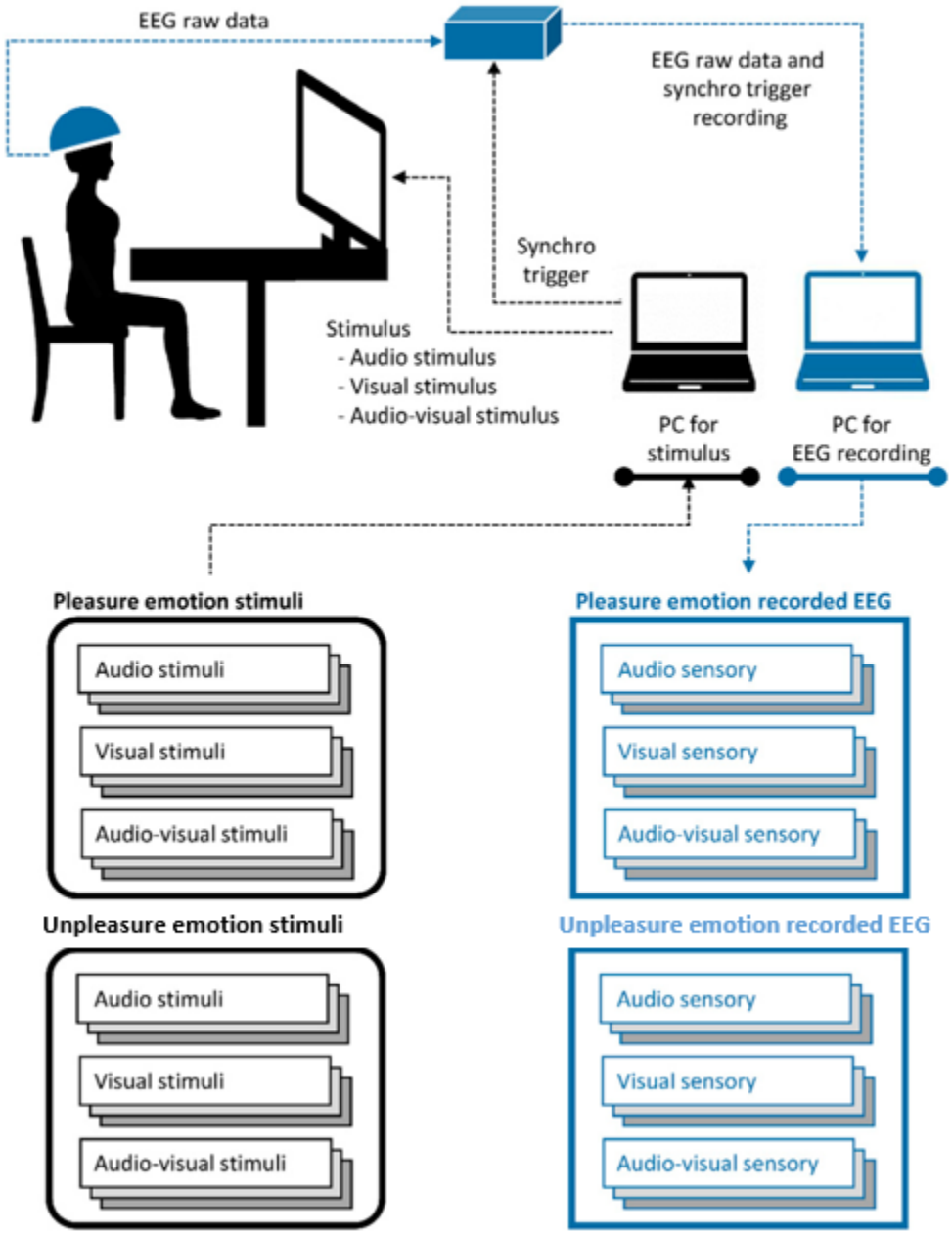


Figure 3.2. Description for cross-sensory EEG emotion data acquisition.

There are three stimulus modalities for inducing their respective sensory modalities: audio, visual, and audio-visual. The audio stimulus modality provided audio information from the emotion videos, the visual stimulus modality provided visual information from the emotion videos, and the audio-visual modality provided audio and visual information from the emotion videos. Hence, it had two types of emotions (pleasure or displeasure) and three stimulus modalities (audio, visual, and audio-visual) for inducing EEG data. The collected cross-sensory emotion EEG data can be described as follows: Pleasure EEG (audio /visual /audio-visual sensory) and unpleasure EEG (audio /visual /audio-visual sensory). The experimental setup is illustrated in Figure 3.2. Twenty healthy participants were recruited for the experiment. The experimental details and data validation are described in Chapter 2. Detailed information on the cross-sensory EEG emotion data is presented in Table 3.1.

Table 3.1. Detailed information on the cross-sensory EEG emotion data.

Cross-sensory EEG emotion data information	
Number of participants	20 (Native Chinese speaker)
Sex	11 males, 9 females
Age	24.7 ± 1.9 years
Number of channels	32 channels
Sampling rate	500 Hz
Experimental stimulus conditions	Audio Pleasure/visual Pleasure/audio-visual pleasure/audio unpleasure/visual unpleasure/audio-visual unpleasure
Collected EEG data	Each participant: 10 trials with 30 s duration for each condition

3.2.2 Data preprocessing

The data preprocessing is described in Figure 3.3. Preprocessing cross-sensory EEG emotion data for cross-sensory emotion recognition involved the following steps. An applied finite impulse response (FIR) bandpass filter facilitated EEG signal extraction between 1 Hz and 50 Hz [141][142], with a subsequent notch filter employed to eliminate 50 Hz mains power interference. Baseline correction was conducted within the initial 1000 ms before stimulation onset. The primary focus of the investigation was the EEG data spanning 0–30 seconds during stimulation, utilizing a 5-second window [173-145]. A total of 7200 samples were analyzed, comprising 60 samples per participant for each of the six cross-sensory EEG data types, each sample containing 2500 temporal features derived from 500 × 5 data points. The EEG data consisted of 32 channels, referencing FCz. Excluding FCz, 31 channels were utilized for further framework implementation. The above preprocessing steps were executed by the MNE [89] library in Python.

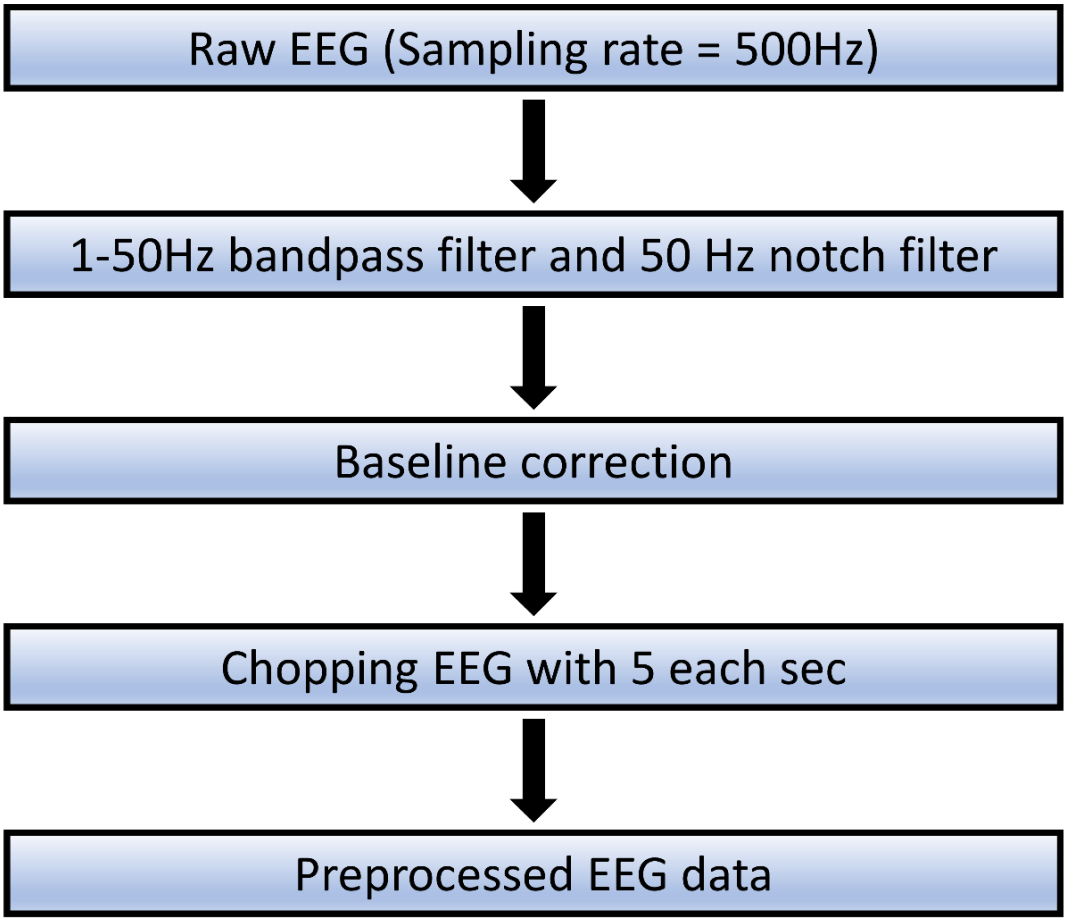


Figure 3.3. Workflow for EEG data preprocessing.

3.2.3 Filter bank Riemannian feature extraction

Emotional EEG signals contain information of various frequency ranges. By categorizing the signals into different frequency bands, the features within each band can be better captured. The frequency bands are associated with different neural activities related to emotional states. Therefore, separating them helps extract emotion-related features accurately. Different frequency ranges of EEG signals are associated with specific neural activity. For instance, the low-frequency range (1–4 Hz) is often associated with relaxation and resting, whereas the high-frequency range (30–50 Hz) is associated with attention and arousal states [146]. Dividing the signals into different frequency bands allowed for a more accurate representation of the influence of different neural activities on emotional states. Consider the balance of sub-bands length and the results in Chapter 2, in this Chapter, it further divided the beta into two sub-bands:13-20 Hz (Beta 1) and 20–30 Hz (Beta 2), to better distinguish different neural activities and EEG signal characteristics [147][148].In the beta band, the lower frequency range (13–20 Hz) is typically associated with attention, cognitive processing, and emotion regulation, whereas the higher frequency range (20–30 Hz) tends towards motor control and perception [149]. Different neural activities and features can be captured more precisely by separating them. Therefore, it used IIR bandpass filters to divide the preprocessed 1–50 Hz EEG signals into six sub-bands: 1–4 Hz (Delta: band 1), 4–8 Hz (Theta: band 2), 8–13 Hz (Alpha: band 3), 13–20 Hz (Beta 1: band 4), 20-30 (Beta 2: band 5), and 30–50 Hz (Gamma: band 6).

The Riemannian space method has good generalizability in transfer learning. By learning Riemannian features from the source domain, transfer learning can be performed in the target domain, reducing the sample requirements and improving the classification performance [150].

Riemannian space methods can capture shared structures and features between source and target domains, enabling knowledge and model transfer.

Extracting Riemannian features from EEG signals involves computing the covariance matrix (CM) in SPD form [151]. EEG signals are typically represented as multi-dimensional time-series data, in which each observation at a given time corresponds to a voltage recorded at different electrode locations. Transforming these time series into a matrix representation is necessary for feature extraction and classification. It can capture the correlation and coactivity between different electrodes in the EEG signals by computing the covariance matrix. The covariance matrix is a symmetric positive-definite, describing the spatial covariance relationships between the different electrodes. Hence, the dimensions of the covariance matrix are related to the number of electrodes. Therefore, computing the covariance matrix in its SPD form is necessary for the Riemannian feature extraction of EEG signals. This allows the transformation of EEG signals into points on the Riemannian manifold and facilitates feature extraction and classification analysis using the geometric structure of the manifold. x and y represent EEG signals from the two channels, with N as the window length or time point of the signals. Preprocessing EEG signals had a signal mean of zero by filtering the direct components; therefore, the CM between both channels can be obtained using Equation (3-1).

$$Cov(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i y_i) \tag{3-1}$$

Let the preprocessed EEG trials be $X_i \in \mathbb{R}^{K*N}$, $i = 1, \dots, t$, K is the number of electrodes, t is the total number of trials. The corresponding CM for X_i can be determined using Equation (3-2):

$$CM_i = \frac{1}{N-1} X_i X_i^T \quad (3\text{-}2)$$

T is the transpose operation of the matrix.

It used the oracle approximating shrinkage [151] regularization method to ensure that all covariance matrices were regularized (in the symmetric positive-definite form). The shape of the CM corresponding to each EEG trial depended on the number of signal channels. There were K signal channels for each EEG trial; thus, the shape of the CM was (K, K). The number of rows and columns in this matrix equals the number of signal channels, representing the correlations and variances between each channel. The series of SPD CMs are denoted as $sym^+ = \{CM \in \mathbb{R}^{K\times K}, x^T CMx > 0, CM = CM^T, \forall x \in \mathbb{R}^K$ and x is a non-zero vector.

CMs naturally belong to the Riemannian manifold rather than the Euclidean space [152]. Transformative operations are required to apply operations suitable for Euclidean space.

However, using the original CMs as features for classification is challenging because they reside on the Riemannian manifold $\mathbb{RM}$, not in Euclidean space. The features extracted from the Riemannian tangent space have better discriminative properties and are easier to handle than the original Riemannian space features [153]. This is primarily due to their linear properties and enhanced discriminability. In tangent space, traditional linear methods such as the Euclidean distance and linear classifiers can be directly applied, as the space is linear, and linear operations have simpler expressions. Additionally, the projection operations in the tangent space highlight the differences between the sample classes, improving feature discriminability. In contrast, the original Riemannian space may have smaller differences between sample classes, leading to lower discriminability of the features. It will utilize the transformation from the $CM \in sym^+$at the Riemannian space—$\mathbb{RM}$ of CMs to the tangent space—$\mathbb{TS}$ to address this issue. Figure 4 illustrates the two-dimensional Riemannian manifold and tangent spaces.

Based on singular value decomposition, the operation $logm$ is denoted as $logarithm$ of a matrix [154] in equation (3-3), where D is a diagonal matrix and P is a non-singular matrix.

$$S_i = log_{CM}(CM_i) = CM^{\frac{1}{2}}(logm(CM^{-\frac{1}{2}}CM_iCM^{-\frac{1}{2}}))CM^{\frac{1}{2}} \quad (3\text{-}3)$$

$log(D)$ can be obtained by taking the logarithm of the elements of the diagonal matrix D.

The Riemannian space point CM_i to the tangent space can be described as in equation (3-4)

$$S_i = log_{CM}(CM_i) = CM^{\frac{1}{2}}(logm(CM^{-\frac{1}{2}}CM_iCM^{-\frac{1}{2}}))CM^{\frac{1}{2}} \quad (3\text{-}4)$$

The operation $expm$ is denoted as $exponential$ of a matrix in equation (3-5)

$$expm(CM) = e^{CM} = \sum_{n=0}^{\infty}\frac{1}{n!}CM^n \quad (3\text{-}5)$$

Inversely, the projected tangent space points can be transferred into the Riemannian space using $expm$, which can be described as in equation (3-6)

$$CM_i = log_{CM}(S_i) = CM^{\frac{1}{2}}(expm(CM^{-\frac{1}{2}}S_iCM^{-\frac{1}{2}}))CM^{\frac{1}{2}} \quad (3\text{-}6)$$

The distance between CM and CM_i are denoted as $\delta_R(CM, CM_i)$ in Figure 3.4.

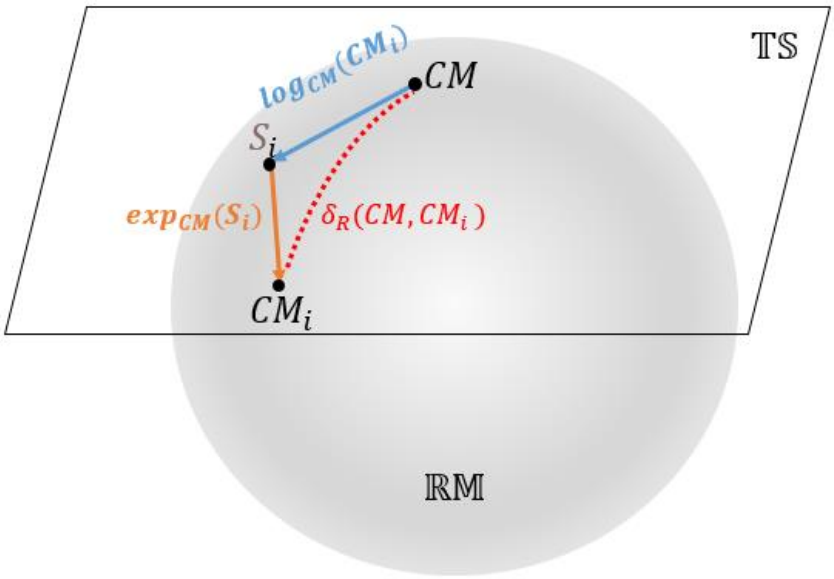


Figure 3.4. Riemannian manifold and tangent space.

Features extracted from the Riemannian tangent space have better discriminative properties and are easier to handle than the original Riemannian space features. This is primarily due to their linear properties and enhanced discriminability. The Euclidean distance and linear classifiers can be directly applied in the tangent space because the space is linear, and linear operations have simpler expressions. Additionally, the projection operations in the tangent space highlight the differences between the sample classes, improving feature discriminability. In contrast, the original Riemannian space may have smaller differences between sample classes, with lower feature discriminability. The features of this study are the tangent space CMs. Therefore, from Equations (3-3) and (3-4), the original Riemannian space CMs——$CM_i^{\mathbb{RM}} = CM_i$ can be transformed into tangent space CMs—$CM^{\mathbb{TS}}$—using Equation (3-7), known as the Log-Euclidean mean covariance [14].

$$CM^{\mathbb{TS}} = expm[\frac{1}{t}\sum_i^t logm(CM_i^{\mathbb{RM}})] \quad (3\text{-}7)$$

The set of Riemannian tangent features in each frequency band could be obtained using $CM_i^{\mathbb{TS}} \in \mathbb{R}^{K\times K}, where\ i = 1, \dots, t$. There are further domain adaptation and classification requirements; therefore, it flattened each trail feature from the two-dimensional Riemannian tangent features—$CM_i^{\mathbb{TS}} \in \mathbb{R}^{K\times K}$ as $CM_i^{\mathbb{TS}\ flatten} \in \mathbb{R}^{F\times 1}, F = K * K$. Therefore, the workflow for feature extraction using the filter bank and Riemannian methods is illustrated in Figure 3.5.

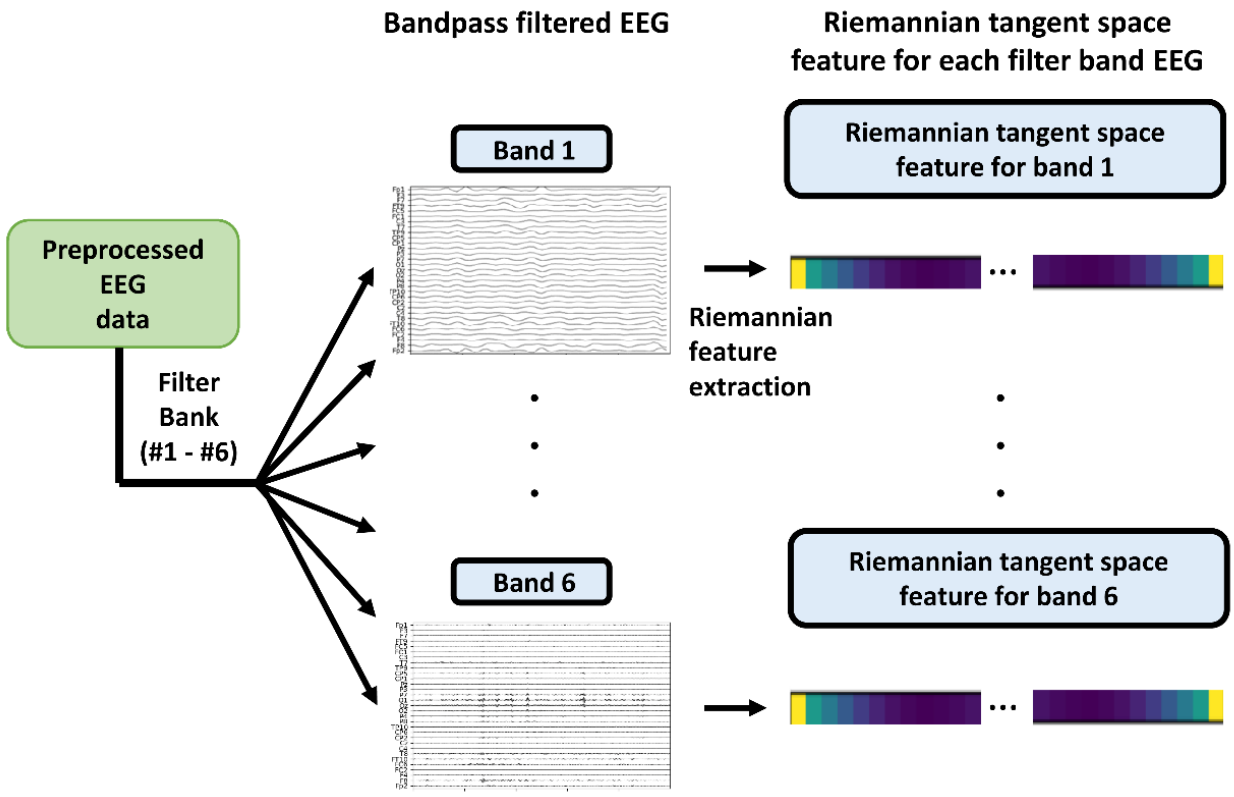


Figure 3.5. Workflow for Filter bank Riemannian feature extraction.

3.2.4 Adversarial domain adaptation

This study mainly focused on Riemannian tangent space features for feature interpolation and analysis, which can be obtained as described in Section 3.2. Features from one of the three sensory modalities can be recognized as the target domain to fulfill the cross-sensory emotion recognition requirements. Thus, the features of the remaining two sensory modalities were recognized as the source domains. For instance, when the features from audio sensory modalities constituted the target domain. Conversely, those from the visual and audio-visual sensory modalities constituted the source domain. Reducing the discrepancy between the source and target domains and making the source domain features proximal to those of the target domain are crucial to domain adaptation. Therefore, it employed an adversarial domain adaptation approach to achieve domain adaptation from the source to the target domain.

Adversarial domain adaptation evolved from generative adversarial networks (GANs) [155] and addressed domain adaptation challenges. GANs consist of generator and discriminator networks trained in an adversarial manner to generate realistic data samples. Adversarial domain adaptation leverages the adversarial training concept of GANs to address domain adaptation, which involves transferring knowledge from a source domain with labeled data to a target domain. Adversarial domain adaptation has two key components: a feature extractor and a domain discriminator. The feature extractor learns the shared representation of the input data, whereas the domain discriminator differentiates the source from the target domains based on the learned features. Adversarial domain adaptation involves minimizing the distribution discrepancy between the source and target domains through a feature extractor, making the features more similar across domains. Simultaneously, the domain discriminator maximizes the distribution discrepancies to distinguish the domains. Adversarial training helps feature extractors generate

domain-invariant representations, enabling knowledge transfer and performance improvement in the target domain. Hence, adversarial domain adaptation builds upon the adversarial training concept of GANs and extends it to address domain adaptation challenges. Minimizing feature distribution discrepancy and maximizing domain discrimination facilitate knowledge transfer from a labeled source domain to an unlabeled target domain, enabling adaptation and improved performance in the target domain.

The original GAN algorithm was first introduced in 2014 [155]. The main idea was a minimax process with two elements: generator—G and discriminator—D. The generator can create real-like fake data by inputting random noise into the generator. The discriminator distinguishes between the generated real-like fake data and the real data. This minimax process aims to train the generator to produce real-like fake data to trick the discriminator into recognizing the generated data as real. In the domain adaptation field, the feature adaptor replaces the original generator by inputting the source domain features into the adaptor to generate source domain features that are target domain-like from the domain adaptor. The proposed adversarial domain adaptation framework comprises a feature adaptor and a discriminator.

With the loss function L, the entire minimax process based on the original GAN for adversarial domain adaptation can be described as follows:

$$L_{\min_{AD}\max_{DS}} = \mathbb{E}_{x_a \sim X_a}[\log(DS(x_a))] + \mathbb{E}_{r \sim z_r(r)}[\log(1\text{-}DS(AD(r))] \quad (3\text{-}8)$$

In equation (3-8), X_a represents the target domain features and $DS(x_a)$ represents the discriminator for calculating the x probability of the target domain distribution. The input source domain features are z_r, $AD(r)$ is an adaptor used with the input r source domain features, outputting the targeted domain-like features.

The features from the target and source domains were extracted using the Riemannian tangent space methods. Previous research has indicated that the original GAN loss is vulnerable to collapse during training sessions. Meanwhile, the emotion-corresponding labels in this study were pleasure and displeasure, regardless of sensory modality. Feature adaptation can fit each emotion label corresponding to the target domain rather than the emerged target domain. Inspired by Wasserstein generative adversarial networks [156][157] and adding label information as conditions, the proposed adversarial domain adaptation framework is indicated in Figure 3.6.

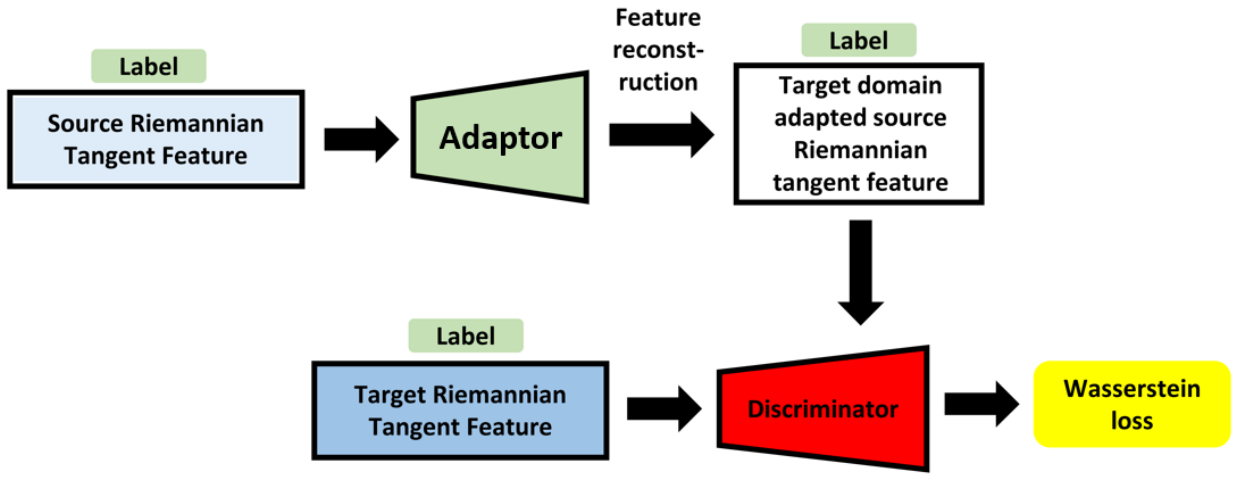


Figure 3.6. Illustration for proposed adversarial domain adaptation framework.

The detailed structures and parameters of the proposed domain adaptor and discriminator are presented in Tables 3.2 and 3.3, respectively.

$$Lw_{\min_{AD}\max_{DS}} = -\mathbb{E}_{x_a\sim X_a}[DS(x_a|y_a)] + \mathbb{E}_{r\sim z_r(r)}\left[DS\left(AD(r|y_r)\right)\right] + \lambda\mathbb{E}_{\hat{x}\sim\hat{X}}[(\|\nabla_{\hat{x}}DS(\hat{x}]y_r)\|_2 - 1)^2 \quad (3\text{-}9)$$

The last term in Equation (3-9) is a penalty element. The data point was sampled from a straight line between the target domain X_a and the adapted source domain X_{AD} is $\hat{X}$, where x_a denotes the data from X_a. The hyperparameter λ controls the trade-off between the target domain and the gradient penalty.

The detailed parameters and functional components of the adaptor are listed in Table 3.2, with one dense layer, one reshaped layer, and two convolutional layers. The leaky ReLU function had a good performance in GANs [158][159]. Therefore, this study chose a leaky ReLU activation layer, followed by a dense layer and the first convolutional layers. It chose a tanh activation layer attached to the last deconvolutional layer for the final adaptor output. In each convolutional layer, batch normalization was introduced to increase the solution speed of the gradient descent and avoid overfitting.

The loss function of proposed framework could be described in (3-9). The detailed parameters and functional parts, specifically explaining each layer, are described in Table 3.3. A leaky ReLU activation layer followed each convolutional layer, and batch normalization was performed after each convolutional layer. The final output of the discriminator uses the sigmoid function. The kernel size for each convolutional layer in the adaptor was three, and the convolutional layer in each adaptor was two, based on previous GAN-related studies [160-162] After implementing the proposed adversarial domain adaptation networks, the adapted source domain features could be obtained, and the target and source domain features were cross-sensory features used together to train the classifiers.

Table 3.2. Adaptor structure.

Detailed parameters in the adaptor				
Layer	Kernel size	Output shape	Activation function	Batch normalization
Input	-	961	-	-
Reshape	-	961*1	-	-
Conv1D	3	961*32	Leaky ReLU	YES
Conv1D	3	961*8	Leaky ReLU	YES
Flatten	-	7688	-	-
Dense	-	961	Tanh	-

Table 3.3. Discriminator structure.

Detailed parameters in the discriminator				
Layer	Kernel size	Output shape	Activation function	Batch normalization
Input	-	961	-	-
Dense	-	32	Leaky ReLU	-
Reshape	2	32*1	-	-
Conv1D	2	32*32	Leaky ReLU	YES
Conv1D	2	32*64	Leaky ReLU	YES
Flatten	-	2048	-	-
Fully connected	-	256	Leaky ReLU	-
Fully connected	-	64	Leaky ReLU	-
Dense	-	1	Sigmoid	-

3.2.5 Classification strategy

After executing domain adaptation, the adapted source and target domain features were used with the corresponding bands. The adapted source and target domain features are recognized as Riemannian domain target features; the SVM with polynomial kernels had a considerable discriminative performance for these specific features [163].

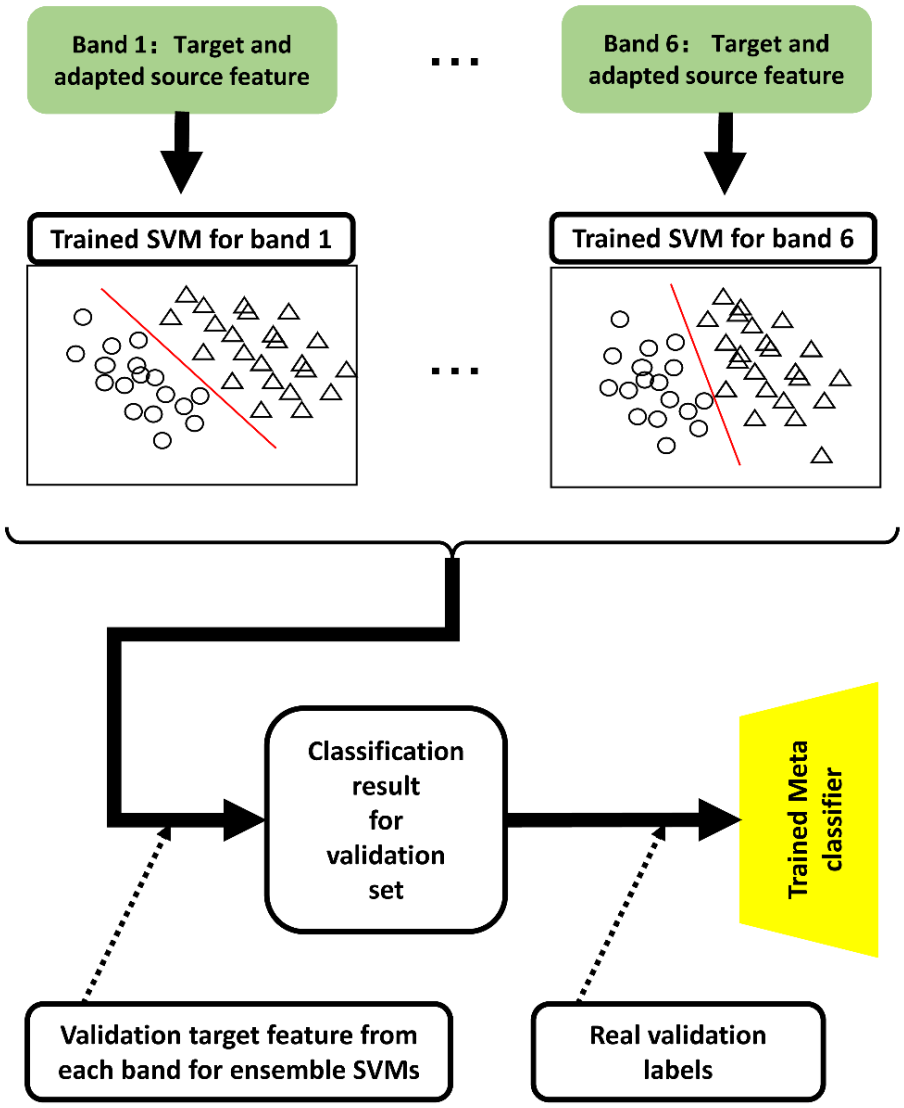


Figure 3.7. Description for ensemble SVMs classification strategy.

Therefore, this study explored polynomial kernel SVMs for each band feature to obtain trained SVM classifiers for emotion recognition. Moreover, six trained SVMs corresponded with the six bands to ensemble the six SVMs' results into final emotion recognition outputs. The proposed classification strategies are illustrated in Figure 3.7.

The target and adapted source features for each band constituted the training data corresponding to each SVM classifier. After the training session, six SVMs were trained for each of the six bands.

Each predicted output with the trained SVMs is identical to each label. Therefore, the transformational relationship between the six SVM outputs and the final classification result cannot be adjusted using the training set directly. Inspired by the blending of stacked methods for ensemble learning [164], it introduced a validation set for blending the transformative relationship between the six SVM outputs into the final classification result.

A set of validation data was explored to ensemble the six SVMs. A set of predicted validation data labels was obtained by inputting the validation target features from each band into each corresponding SVM. The predicted validation data labels are denoted as $L_i^{bandj}, j = 1, \dots, 6,\ i = 1 \dots n_v$, n_v is the number of trials for the validation set. The L_i^{bandj} is the input feature for the meta-learner and the real validation labels serve as the input label information for the meta-classifier.

Two classifiers were used for the blending methods. The basic classifiers obtained the predicted multiple outputs from the original input features, and a meta-classifier for receiving multiple outputs from basic classifiers and providing the final classification results. In this study, the basic classifiers were SVMs. A logistic regression (LR) model was selected as the meta-classifier for the final susceptibility prediction because of the effectiveness of simple linear models [165].

Therefore, LR was trained to ensemble the outputs from the six SVMs. In the test session, emotion recognition results for the test data were obtained by inputting the filter bank Riemannian feature from the target and adapted source features for the proposed trained SVM-meta learner framework. Based on sections 3.1, 3.2, and 3.3, the training process of proposed FBADR could be described as Algorithm 1.

Algorithm 1: Training process of proposed FBADR

1. Filter bank

Determine sub-bands EEG signals banks $X_{i|bandj} \in R^{K*N}$, j =1…6 from filter bank method.

2. Riemannian method

Compute the covariance matrix for each sample $CM_{i|sub}$ from equation 2.

Compute the CM^{TS} as Riemannian tangent space features from equation 5.

3. Adversarial domain adaptation

Determine adapted source domain feature X_{AD} from equation 7.

4. Ensembled SVMs classifiers training

Train SVMs with target and adapted source feature and labels from band 1 to band 6.

Determine L_i^{bandj} by inputting the validation target features into trained SVM correspondingly.

Train meta learner with L_i^{bandj} and real validation labels.

3.3 Results

3.3.1 Experimental description

The experiment can be described in three sessions: feature extraction, adversarial domain adaptation, and classifier training and validation, as illustrated in Figure 3.8. For example, the target domain EEG was considered audio sensory, and the source domain EEG as visual and audio-visual sensory. All experiments were executed on Windows 10 and Python 3.6, with an Intel i7-10875H CPU and an NVIDIA RTX, the 2080s GPU.

Feature extraction was performed using the MNE, SciPy, and NumPy libraries. In the feature extraction session, the target domain EEG first underwent six bndpass filters to obtain a target domain EEG with six bands. Then, based on the well-known transfer learning metric for cross-subject on EEG decoding—Leave-One-Subject-Out [166]—we chose "Leave-One-Sensory-Out." Specifically, we used one sensory modality as the target domain and the other two as source domains for each subject. The target domain data were split into a training, validation, and test set of 60%, 20%, and 20%, respectively, coordinated with each band's EEG data. The training, validation, and testing sets were used for Riemannian feature extraction. The source domain EEG served only as training data and was passed through filter banks and Riemannian feature extraction.

Adversarial domain adaptation frameworks were built and executed based on TensorFlow1.13.1, the Adam optimizer [167] with $Learning\ rate = 0.001$ was used for network optimization. In the adversarial domain adaptation session, the source domain features were adapted using the proposed adversarial domain adaptation methods coordinated with the extracted features from the target domain training set. The target and adapted source features from each band were obtained from an adversarial domain adaptation session. For each target

sensory modality subject, 100 training epochs with a batch size of 32 were used to train the framework. After completing the training session, a trained adaptor was used to generate the adapted source domain features.

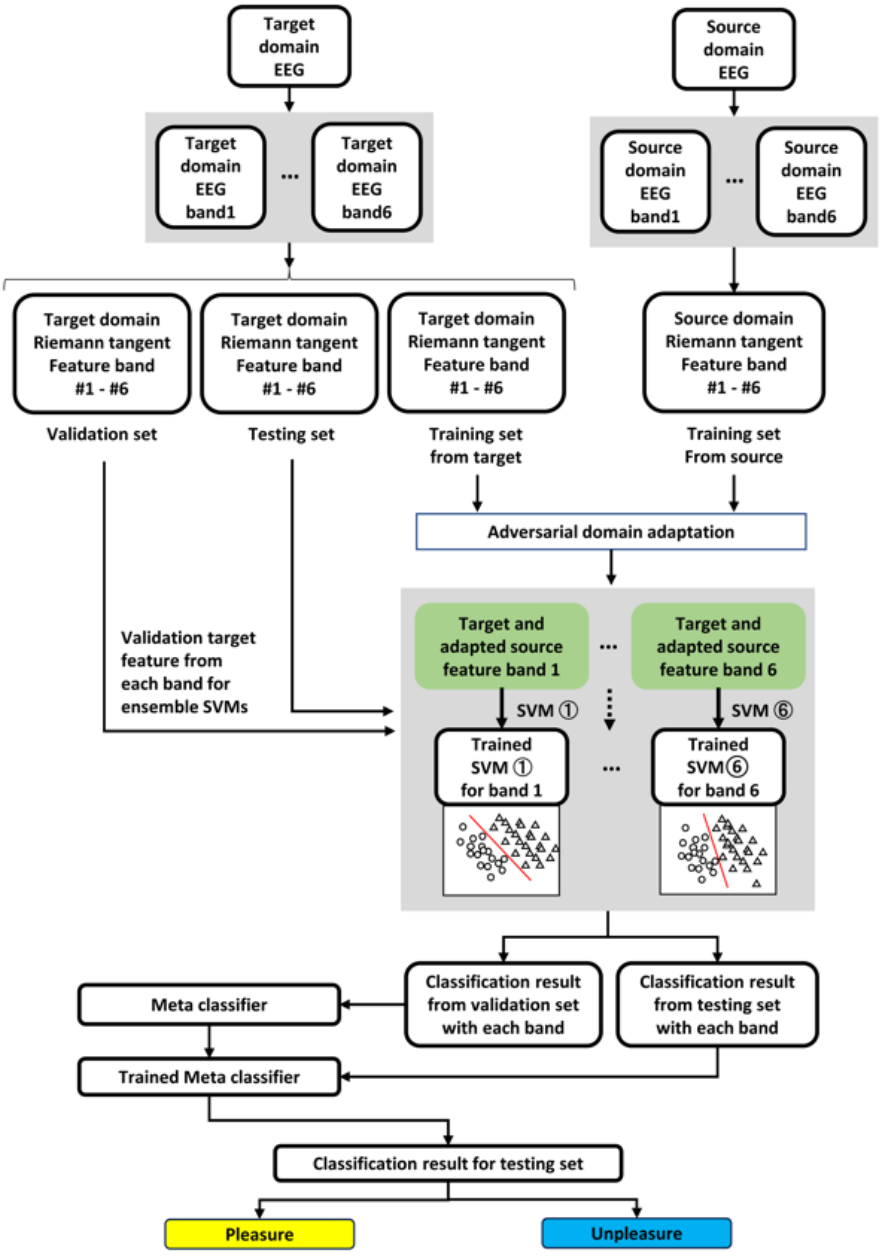


Figure 3.8. Workflow for entire cross-sensory EEG emotion recognition based on proposed FBADR

In the classification session, the frameworks were realized using Scikit-learn to build the SVM and meta-classifier LR. Using GridSearchCV in subjects 01, 02, and 03, SVM parameters were adjusted as $\mathrm{C} = 0.001, \mathrm{kernel} = \prime\mathrm{poly}\prime, \mathrm{gamma} = 10,$ and the LR-based meta classifier was used under the default parameters. In training and validating the classification session, the target and adapted source features from each band served as training features for each SVM. Emotion labels (pleasure or displeasure) constituted the training label information. Subsequently, six SVMs were trained through stacked ensemble methods using LR classifiers trained by inputting the classification results of the six SVMs from the validation set of target domain features as input features. The real label information of the validation set of target domain features was used as training label information. An ensemble of six final classification outputs was built. Finally, the testing set features were used with six trained SVMs to obtain the classification outputs, which underwent the trained LR, and the classification results from the testing set were obtained. It used 5-fold cross-validation methods to obtain the final classification results for each subject in each target sensory modality. Specifically, for 5-fold cross-validation, each subject's target domain features were randomly shuffled while maintaining the proportion of the label information. The shuffled data were divided into five equal subsets. For each subset, three of the remaining four constituted the training set, and one was the validation set. This subset was used as the test set. The allocated training, validation, and test sets were used for domain adaptation and classification, and classification results were obtained. This process was repeated five times until each subset was used as the test set. Therefore, all the subjects' target domain features were included in the model evaluation.

3.3.2 Adversarial Riemannian methods validation

The proposed adversarial domain adaptation method aimed to align the source-domain features to the target domain. The training process loss of the adaptor and discriminator for 20 subjects × 3 sensory modalities × 6 filter banks = 360 training sessions; the loss during the training session is indicated in Figure 3,9 to illustrate the validation of the proposed adversarial domain adaptation.

It fitted and highlighted the average loss curve for all 360 training sessions. Figure 3.9 indicates that the adaptor and training losses converged to zero under 100 training epochs in the most sessions. Hence, the purpose of minimum-maximum training for domain adaptation was confirmed.

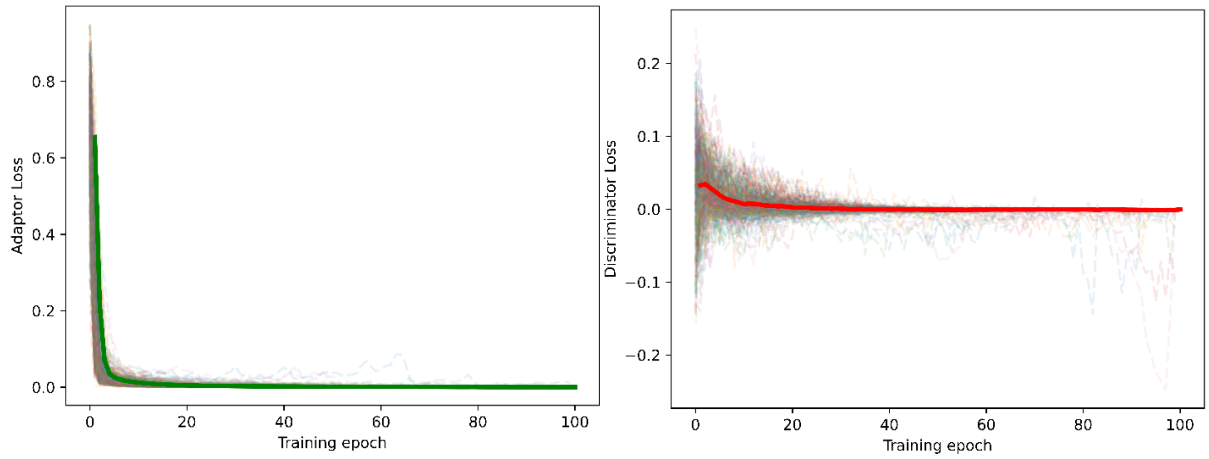


Figure 3.9. Loss curve under training sessions of the adaptor and discriminator.

Moreover, we compared the target and original/adapted source domain features to investigate the domain alignment from the proposed adversarial domain adaptation methods by visualizing two-dimensional t-stochastic Neighbor Embedding [168], as illustrated in Figure 3.10.

Red points indicate the target domain features; green points represent the source domain features. For (a), (b), (c), (d), (e), and (f), the left subfigures are the target and the original source domain features, and the right subfigures are the target and adapted source domain features. From the target domain features (a): sub 01; audio; 1–4 Hz, (b): sub 04; audio-visual; 4–8 Hz, (c): sub 17; visual; 8–13 Hz, (d): sub 11; visual; 13–20 Hz, (e): sub 05; audio; 20–30 Hz, and (f): sub 10; audio-visual; 30–50 Hz, the source domain features were the corresponding subject, filter banks, and the rest of two sensory conditions. Figure 10 reveals the success of domain adaptation by comparing the target and original/adapted source domain features. Via the proposed methods, the source domain Riemannian features have been adapted with the target domain Riemannian features.

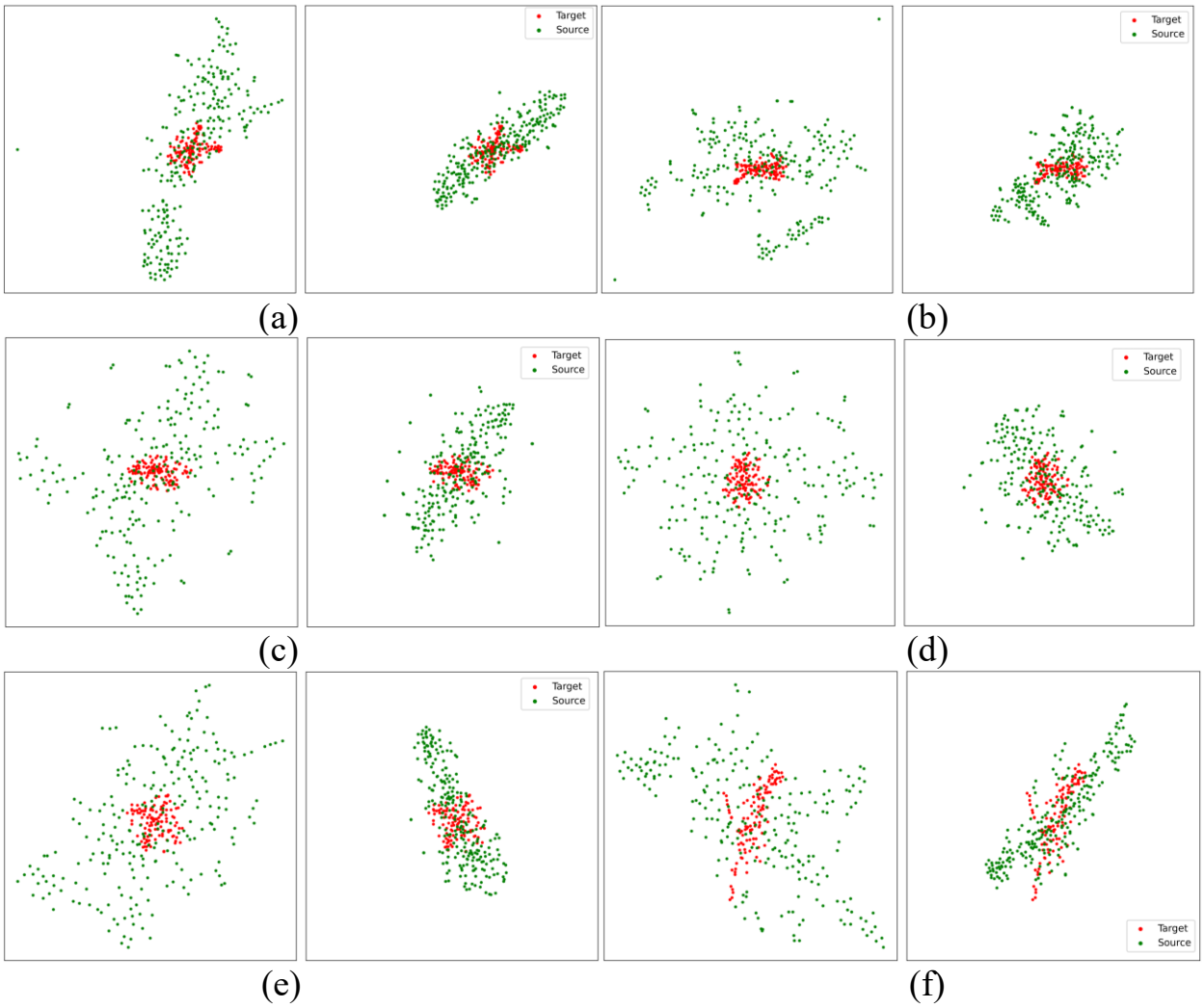


Figure 3.10. Visualization of target domain features and original/adapted source domain features.

3.3.3 FBADR emotion recognition and comparative results

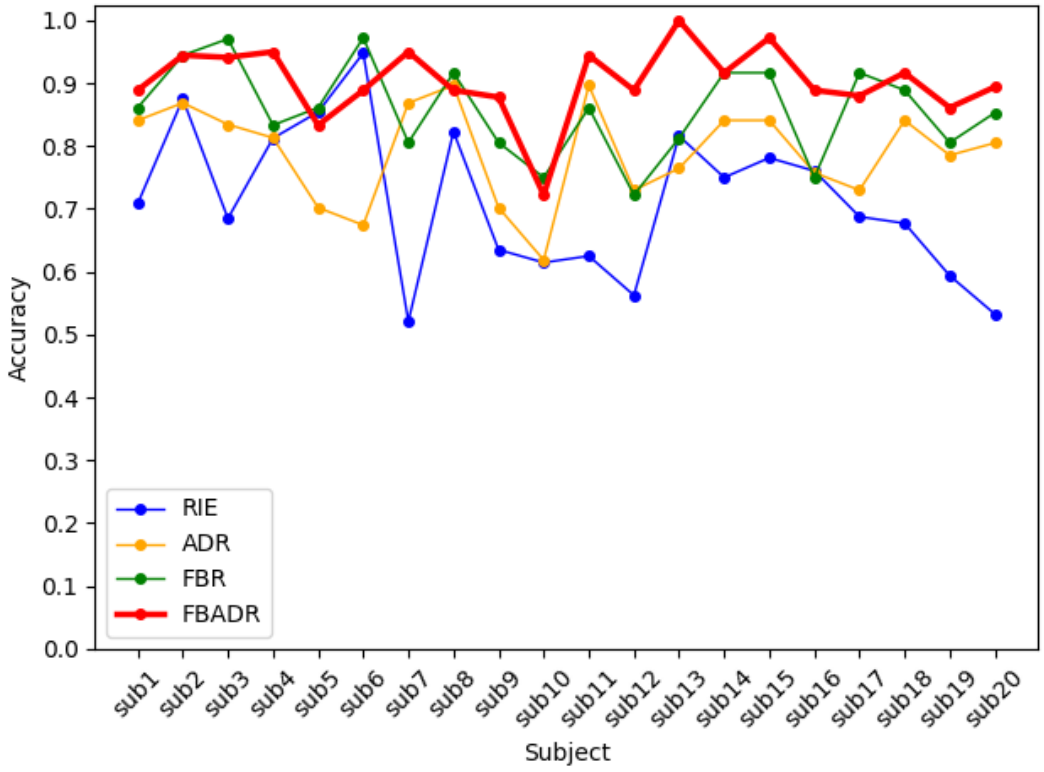
The cross-sensory emotion recognition results were obtained using the proposed FBADR methods (Figure 3.11). Meanwhile, for a comparative study, Riemannian methods (RIE), which use Riemannian feature extraction without filter banks and domain adaptation, adversarial domain Riemannian methods (ADR), filter bank Riemannian methods (FBR), and Riemannian feature extraction with the proposed adversarial domain adaptation, their classification results were used. RIE and ADR did not have multibank features; therefore, it used only SVM for classification, and the FBR classification strategy was the same as that of the proposed FBADR. The emotion labels with the proportion of $pleasure: unpleasure = 1:1$ for binary emotion recognition; thus, the accuracy could be utilized for model evaluations. The accuracy can be determined using Equation (3-10):

$$ACC = \frac{TP+TN}{TP+FN+TN+FP} \tag{3-10}$$

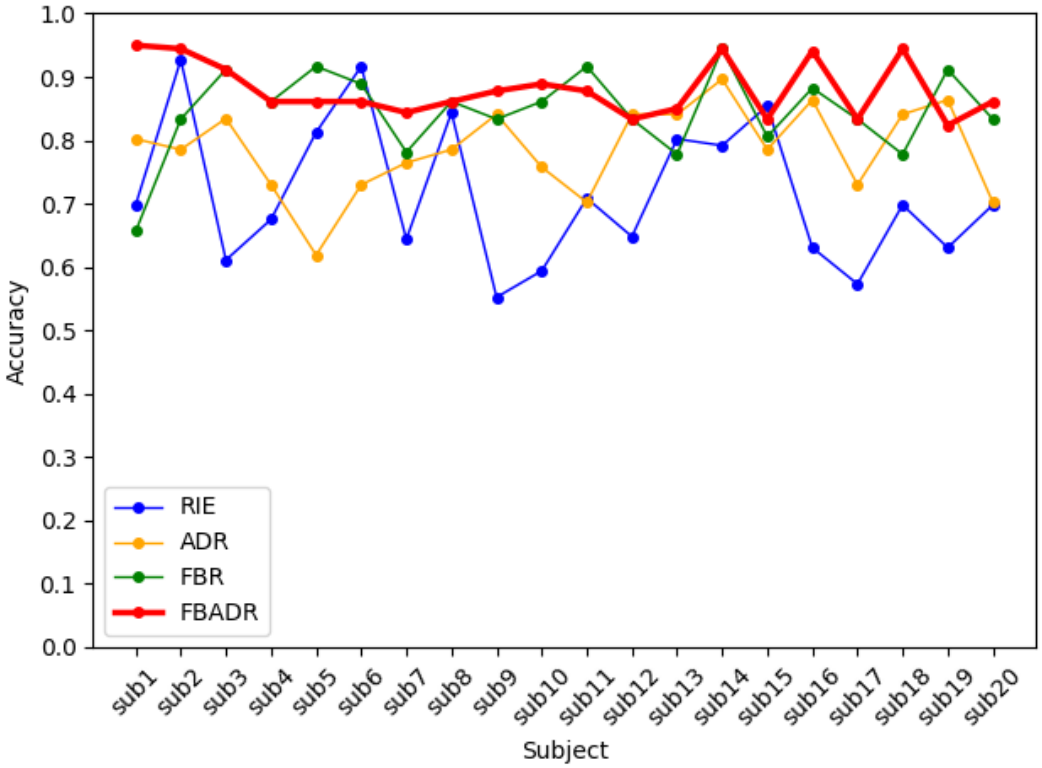
Results classified pleasure as true positive (TP) and classified unpleasure as true negative (TN), false positive (FP), or false negative (FN).

The proposed FBADR methods had the best classification results in cross-sensory emotion recognition, with an average accuracy of $89.01\% \pm 5.06\%$ (Figure 3.11). RIE, ADR, and FBR had average accuracies of $71.11\% \pm 11.31\%$, $79.47\% \pm 6.80\%$and $84.79\% \pm 7.83\%$, respectively. Additionally, the average computational cost in training and test sessions for RIE, ADR, FBR, and FBADR was 0.03s, 1.22s, 0.21s, and 7.87s for each subject. Figure 3.11(a), (b) and (c) show that FBADR performs best in audio, visual and audio-visual sensory modalities. One-way ANOVA was used to determine significant differences between each method, with p

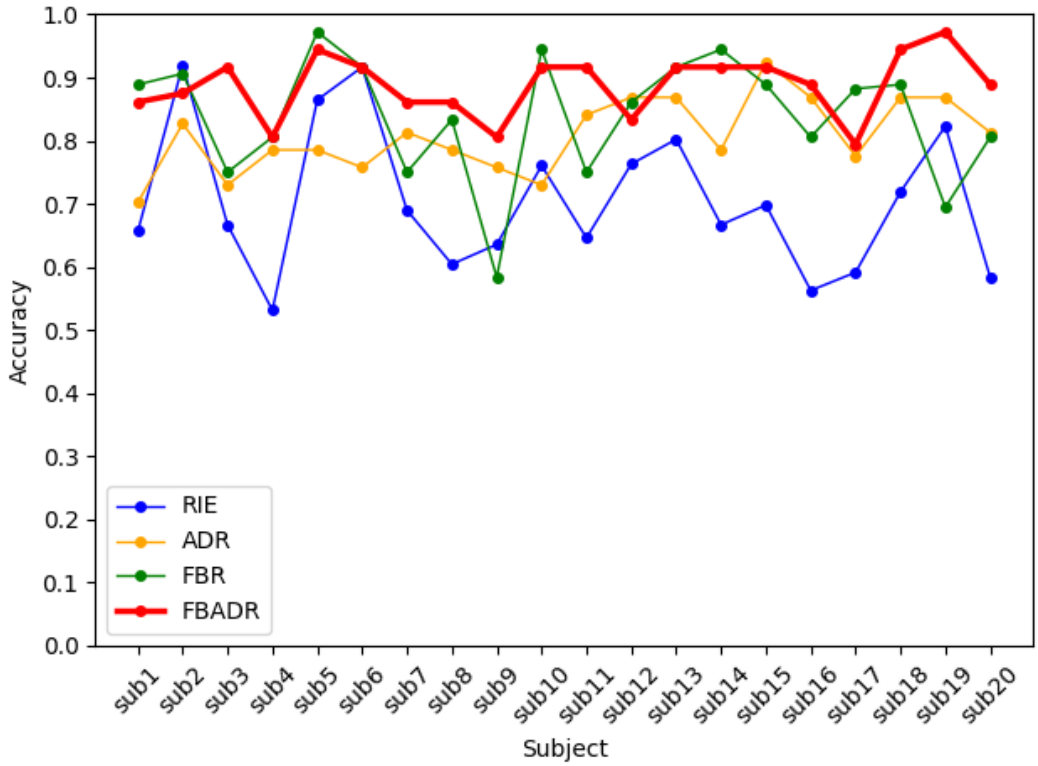
values < 0.001 for the RIE-FBR, RIE-ADR, RIE-FBADR, ADR-FBADR and FBR-FBADR pairs in Figure 3.11 (d) is represented by ***. Filter combination learning and adversarial domain adaptation methods can also improve the accuracy of cross-sensory emotion recognition.



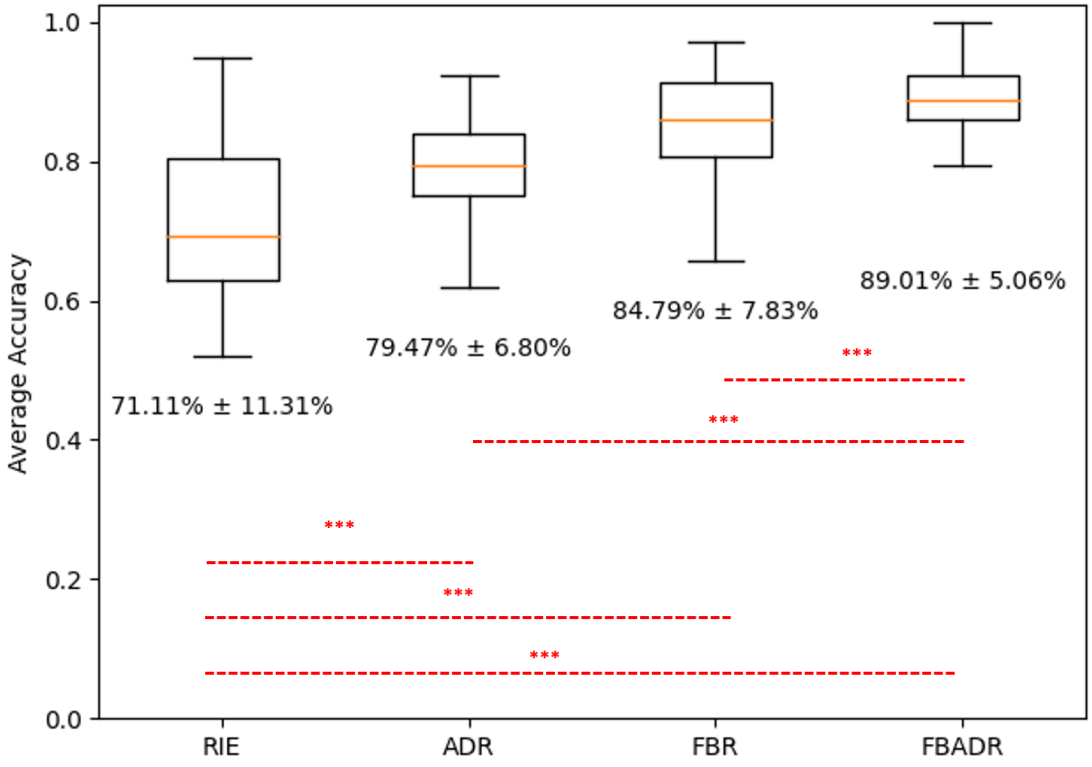
(a)


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Accuracy
RIE
ADR
FBR
FBADR
sub1 sub2 sub3 sub4 sub5 sub6 sub7 sub8 sub9 sub10 sub11 sub12 sub13 sub14 sub15 sub16 sub17 sub18 sub19 sub20
Subject

(b)


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Accuracy
RIE
ADR
FBR
FBADR
sub1 sub2 sub3 sub4 sub5 sub6 sub7 sub8 sub9 sub10 sub11 sub12 sub13 sub14 sub15 sub16 sub17 sub18 sub19 sub20
Subject

(c)



(d)

Figure 3.11. Cross-sensory emotion recognition results from RIE, ADR, FBR, and FBADR. The classification results for each subject with the target domain as audio, visual, and audio-visual are (a), (b), and (c), respectively; (d) is the average accuracy of 20 subjects and three sensory modalities, the text in (d) indicate the average ± std of the accuracies.

Meanwhile, in order to assess the impact of various SVM kernels on the performance of the proposed FBADR method, it incorporated different types of kernels including polynomial, linear, and radial basis function (RBF) kernels into the SVM framework. Specifically, for the implementation of FBADR, it utilized SVMs with polynomial, linear, and RBF kernels, setting the parameters as C=0.001 and gamma=10. The outcomes obtained from employing these three distinct kernels are outlined in Table 3.4.

Table 3.4. Average classification results from different SVM kernels of proposed FBADR

Kernel	Mean Accuracy
Linear	$72.49\ \% \pm 9.85\%$
RBF	$81.91\ \% \pm 8.25\%$
Polynomial	$89.01\% \pm 5.06\%$

The results from Table 3.4 indicated that the polynomial kernel-based SVM classifier performed best in the Riemannian feature-based classification task [163].

3.3.4 Baseline methods comparison

Cross-sensory EEG emotion recognition is an emerging challenge in the EEG aBCI field; therefore, no related studies exist. To obtain a comprehensive performance evaluation, it chose six recently published EEG decoding frameworks for comparative study:

KNN [169], an EEG emotion recognition method that utilizes entropy and energy, was calculated as features after being divided into four frequency bands using discrete wavelet transform and of K-nearest neighbor (KNN) classifier. Frequency band features from the Gamma band were used as a baseline.

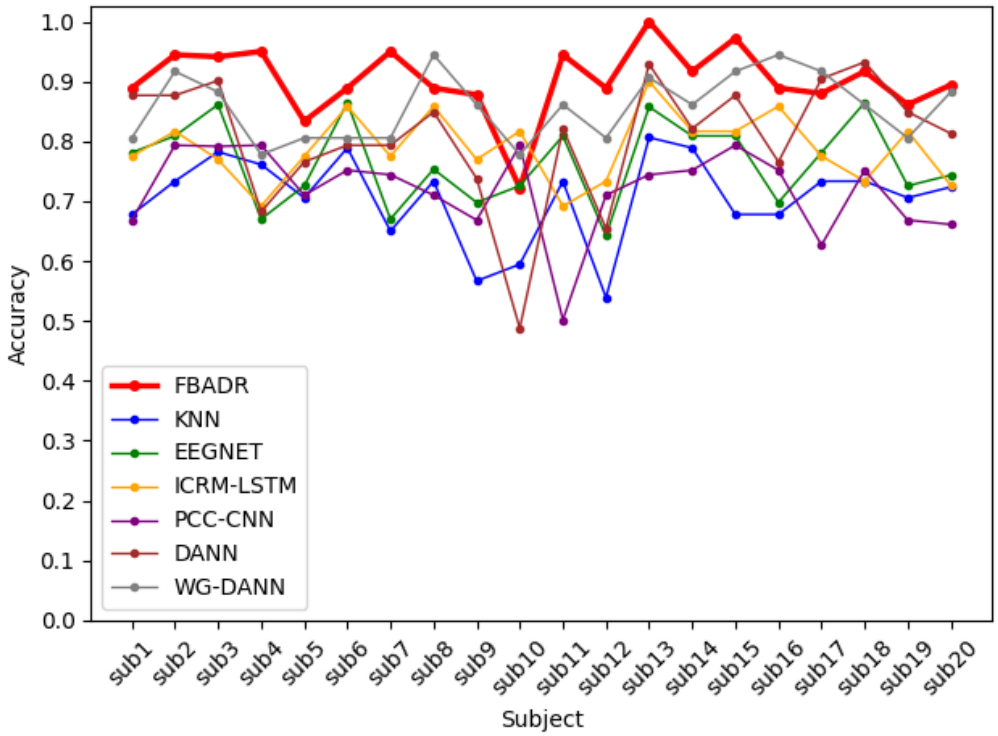
EEGNET [170], and end-to-end EEG decoding framework based on neural networks, has been widely adopted in emotion recognition, motor imagery, and other BCI fields.

ICRM-LSTM [14], a model for EEG-based emotion recognition by combining the independent component analysis (ICA), the Riemannian manifold (RM), and the long short-term memory network (LSTM).

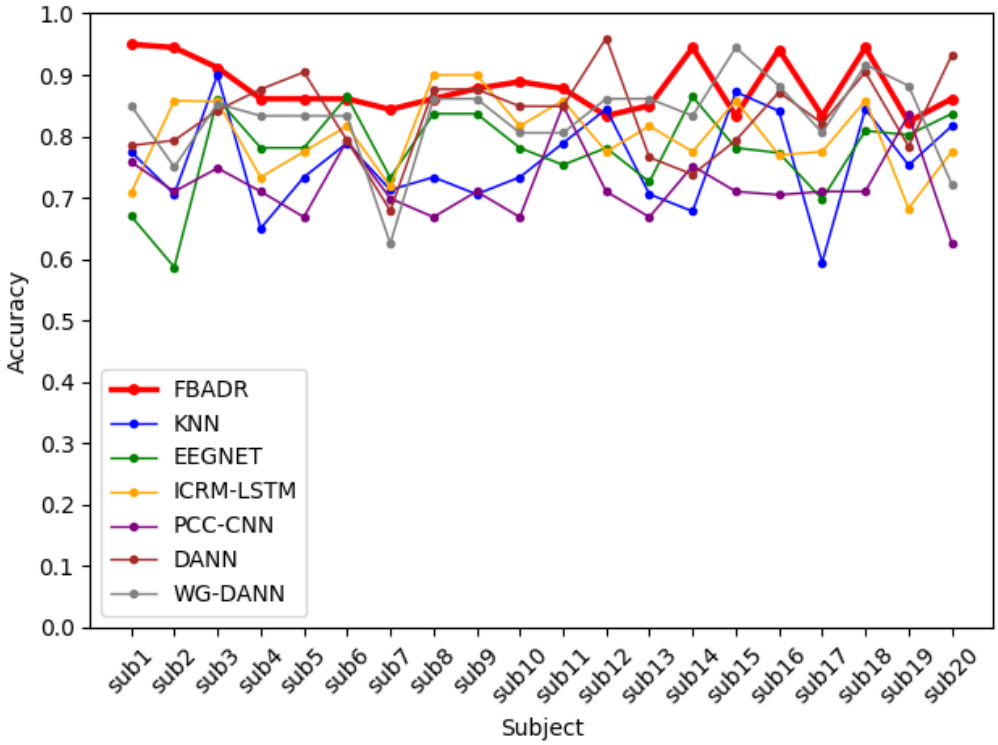
PCC-CNN [121], an EEG Emotion recognition framework with Pearson's Correlation Coefficient (PCC), featured images of channel correlation of EEG sub-bands and the CNN model to recognize emotion.

DANN [171], a transfer learning-based EEG emotion recognition framework with adversarial domain adaptation neural networks for cross-subject EEG emotion recognition framework.

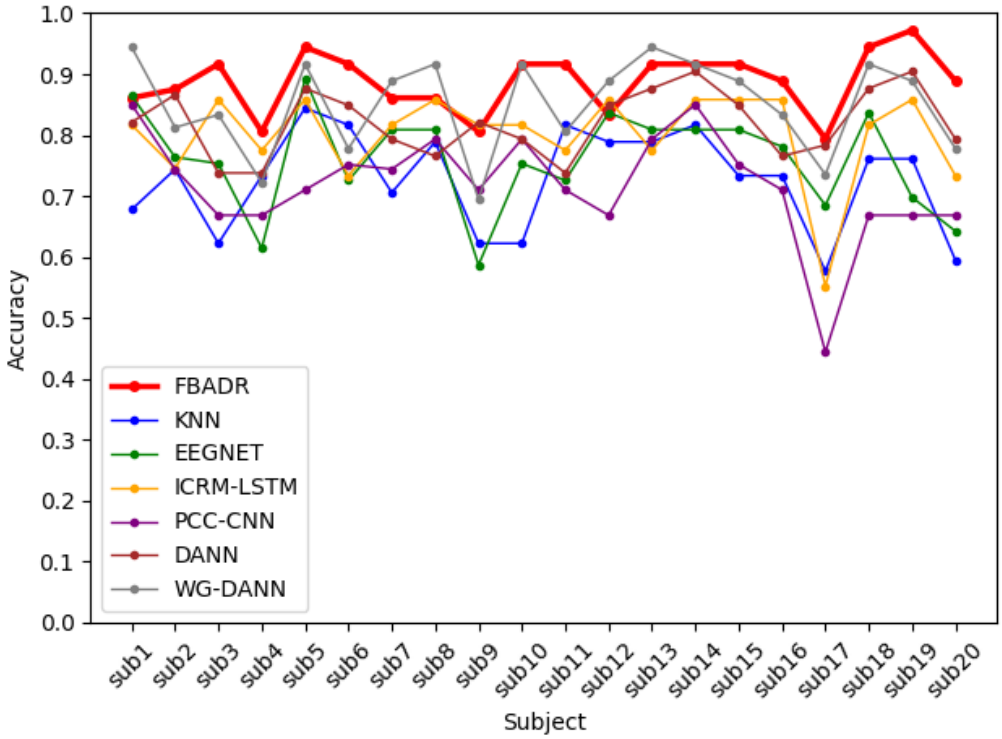
WG-DANN [172], a transfer learning-based EEG emotion recognition framework, consists of GANs-like components and a two-step training procedure with pre-training and adversarial training with Wasserstein GAN gradient penalty loss for cross-subject EEG emotion recognition framework


Accuracy
Subject
FBADR
KNN
EEGNET
ICRM-LSTM
PCC-CNN
DANN
WG-DANN
sub1 sub2 sub3 sub4 sub5 sub6 sub7 sub8 sub9 sub10 sub11 sub12 sub13 sub14 sub15 sub16 sub17 sub18 sub19 sub20

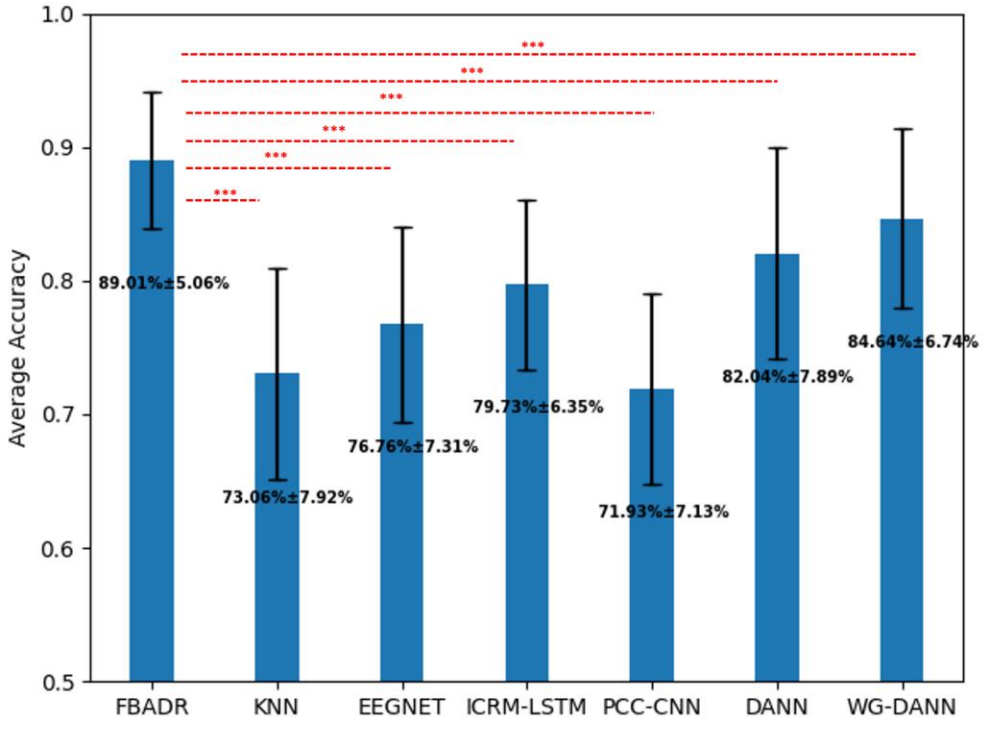
(a)


Accuracy
Subject
FBADR
KNN
EEGNET
ICRM-LSTM
PCC-CNN
DANN
WG-DANN
sub1 sub2 sub3 sub4 sub5 sub6 sub7 sub8 sub9 sub10 sub11 sub12 sub13 sub14 sub15 sub16 sub17 sub18 sub19 sub20

(b)


Accuracy
Subject
FBADR
KNN
EEGNET
ICRM-LSTM
PCC-CNN
DANN
WG-DANN
sub1
sub2
sub3
sub4
sub5
sub6
sub7
sub8
sub9
sub10
sub11
sub12
sub13
sub14
sub15
sub16
sub17
sub18
sub19
sub20

(c)


Average Accuracy
89.01%±5.06%
73.06%±7.92%
76.76%±7.31%
79.73%±6.35%
71.93%±7.13%
82.04%±7.89%
84.64%±6.74%
FBADR
KNN
EEGNET
ICRM-LSTM
PCC-CNN
DANN
WG-DANN

(d)

Figure 3.12. Cross-sensory emotion recognition results from proposed FBADR, KNN, EEGNET, ICRM-LSTM, PCC-CNN, DANN, and WG-DANN. (a)(b)(c) were the classification results for each subject with the target domain as audio, visual, and audiovisual, respectively. (d) was the averaged accuracy of 20 subjects and 3 sensory modalities, the text in (d) are average $\pm$ std of accuracy.

All the baseline results were the target domain as one sensory modality and the source domain as the other two for each subject, with the same 5-fold cross-validation training strategy as in the proposed FBADR. The input data for the baseline methods were proposed using the workflow described in Section 3.2.2. The original algorithms from the baseline methods did not contain a domain adaptation process (KNN, EEGNET, ICRM-LSTM, PCC-CNN), and the training data were used directly as original features from the source and target domains. The baseline methods only required the training and testing sets; therefore, the training data used comprised 80% of the target and all of the source domain data. The testing data included 20% of the target domain data, implemented with 5-fold cross-validation, similar to the proposed FBADR classification strategy. As EEG emotion recognition mainly focuses on improving the classification accuracy of EEG-based emotion recognition BCI systems, it compared the classification accuracy of the cross-sensory EEG data for the proposed FBADR and baseline methods. The classification accuracy of the baseline comparison is presented in Figure 3.12. The EEGNET was reimplemented from https://github.com/vlawhern/arl-eegmodels, and the rest of the baseline methods were reimplemented based on original papers.

Proposed FBADR frameworks has the best performance among all baseline methods, proposed Riemannian feature-based framework showed SOTA performance in cross-sensory emotion recognition from Figure 3.13, which reveals that the proposed FBADR method had the best classification results in the context of cross-sensory emotion recognition compared to the

baseline methods, with an improvement of approximately 5% in average accuracy of the best performance of the baseline methods and a lower standard deviation. One-way ANOVA was used to determine the statistical difference between each method, and a $\mathrm{p}-\mathrm{value}\ of\ <0.001$ for the pairs of FBADR to the six baseline methods is denoted as *** in Figure 3.12(d). The Average accuracy and statistical difference between the proposed FBADR and baseline methods in the experiments indicated the viability of the proposed FBADR method for cross-sensory emotion recognition.

3.3.5 Robustness verification

Gaussian noise was introduced into the experimental data to further validate the robustness of cross-sensory emotion recognition based on the proposed FBADR. Specifically, with the average power of the original data as P_{signal}, the average power of noise as P_{noise} , the signal-to-noise ratio (SNR) can be obtained using Equation (3-11)

$$SNR = 10log_{10}(\frac{P_{signal}}{P_{noise}}) \tag{3-11}$$

Normally distributed Gaussian noise and the probability density function can be represented by Equation (3-12):

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{(\frac{x^2}{2})} \tag{3-12}$$

Coordinated with P_{signal} , the P_{noise} varied with $SNR = 30\ dB, 20\ dB, 10\ dB, 1\ dB, -0.1\ dB$, with the lower signal-to-noise ratio representing a higher power ratio in the noised signals. The temporal presentation of the noised signals is indicated in Figure 3.13.

It used the proposed FBADR on the five-group noisy cross-sensory data, and the average classification accuracies for the original data and the five-group noisy cross-sensory data are shown in Figure 3.14.

Figure 3.14 indicates that with the noised data of $SNR = 30\ dB, 20\ dB, 10\ dB, 1\ dB, and\ -0.1\ dB$, the average accuracy reached 88.77%, 88.66%, 88.53%, 87.30%, and 81.83%, respectively. For further statistical analysis One-way ANOVA was used to determine the statistical difference between the accuracy of the original data and the five-group noised data, $p - values$ were determined from the five pairs as follows:

original - SNR: 30 dB = 0.83 , original - SNR: 20 dB = 0.75 , original - SNR: 10 dB = 0.70 , original - SNR: 1 dB = 0.12, original - SNR: −0.1 dB = 2.9×10^{-10}. Moreover, with the test session in each of 5s, the average $time\ cost = 3ms$, indicated the viability for real time application.

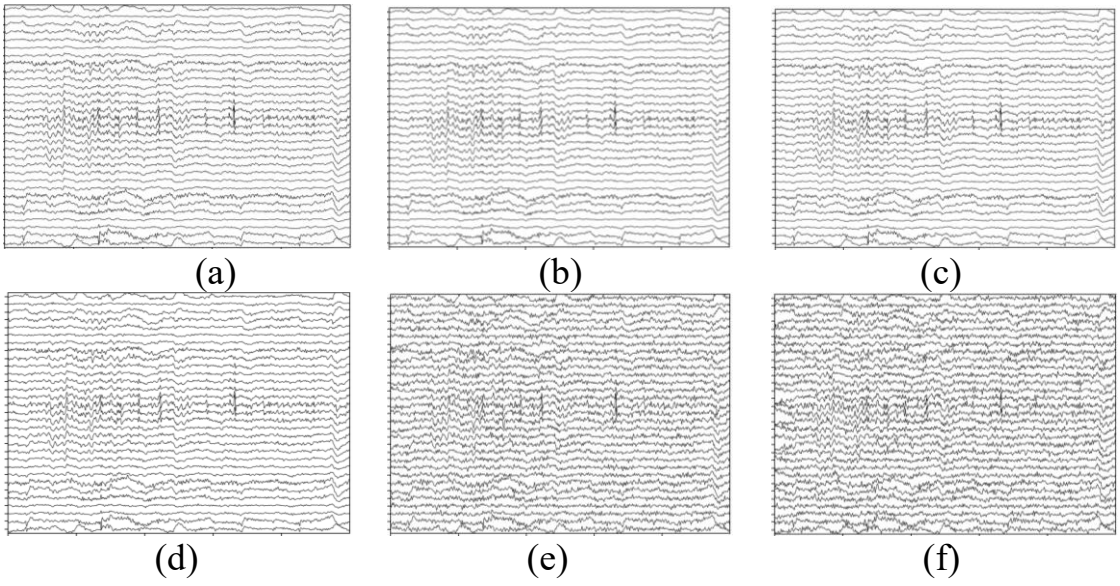


Figure 3.13. Temporal presentation of the noised signals, data are from the first 5 s of sub 01 audio-happy condition with 1–50 Hz filtered. (a) Original data without noise. (b) Noised data with $SNR = 30\ dB$. (c) Noised data with $SNR = 20\ dB$. (d) Noised data with $SNR = 10\ dB$. (e) Noised data with $SNR = 1\ dB$. (f) Noised data with $SNR = -0.1\ dB$.

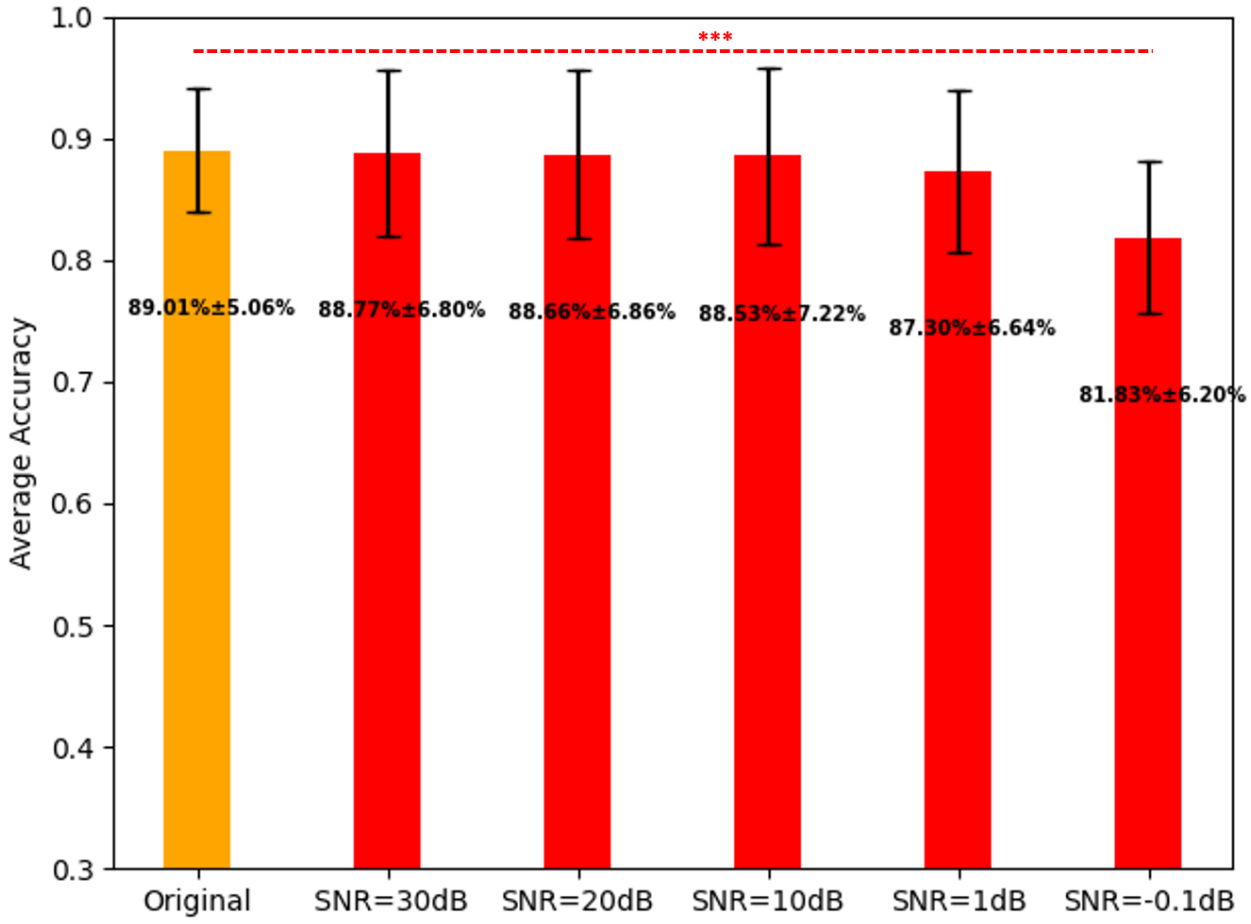


Figure 3.14. Average classification for the original data and five-group noised data with proposed FBADR, the text average $\pm$ std of accuracy.

3.4 Discussion

Generally, we introduced a novel framework to address the challenge of cross-sensory EEG emotion recognition in multimodal emotion stimulation as three sensory modalities: audio /visual /audio-visual with two emotion states: pleasure or unpleasure. To accomplish this, we conducted self-designed experiments involving multimodal emotion simulations to acquire cross-sensory emotion EEG data. The proposed approach—filter bank adversarial domain adaptation Riemann method—leverages Riemannian tangent space methods and filter bank techniques to effectively extract features from cross-sensory emotion data. A key innovation of this study is applying adversarial domain adaptation to mitigate domain differences in cross-sensory situations, enhancing emotion recognition performance. Specifically, we employed adapted features from the source and target domains to train the ensemble SVM classifiers. This integration of adversarial learning and ensemble learning methodologies successfully addressed the challenges associated with cross-sensory EEG emotion recognition, enabling the accurate binary classification of pleasurable and unpleasurable emotions. This study analyzed comparative classification results and demonstrated that the proposed FBADR framework achieved state-of-the-art performance in cross-sensory emotion recognition, attaining an average accuracy of $89.01\% \pm 5.06\%$. Notably, this level of accuracy was the highest among comparable approaches, accompanied by a low standard deviation. Furthermore, this study assessed the robustness of the framework by introducing Gaussian noise, which indicated that the framework was highly resilient to noise interference. Therefore, it would like to further discuss the results in this chapter, in the name of baseline methods comparison, comparative classification and robustness verification.

3.4.1 Baseline methods comparison discussion

Baseline methods comparison results from Figure 3.12 showed that proposed FBADR methods have the best classification average among all baseline methods, and ICRM-LSTM has the best results with the Riemannian feature in non-domain adaptation methods. By employing the filter bank ensemble approach, it is possible to decompose the original signal into different frequency sub-bands, aiding in capturing information from various frequency ranges [173]. Under emotion stimulation, the brain's responses to different frequencies have the possibility to vary, potentially relating to emotion regulation and sensory processing [106]. Implementing filter bank ensemble techniques better captures these frequency-specific variations, hence enhancing sensitivity to emotion-related changes. Furthermore, integrating decoding results from each filter bank through ensemble learning could improve the final classification results. With the comparative spatial characteristic of emotion and sensory-related changes, the covariance matrix-based Riemannian tangent methods, and channel covariance matrix-based Riemannian feature extraction methods could capture the more varied spatial feature from emotion-related changes [174]. By combining with adversarial domain adaptation, it is possible to execute adversarial transfer of features between different sensory inputs, enabling the model to better adapt to diverse data distributions. The domain adaptation has the strength to reduce the impact of different feature distributions [172][173], which are audio, visual, and audio-visual sensory differences in this study. It could benefit the model's generalization ability in cross-sensory EEG emotion recognition.

The novelty of this study is based on the novel utilization and integrated framework of Filter Bank Riemannian features with adversarial domain adaptation. The proposed method introduces a perspective methodology to cross-sensory EEG emotion recognition. Proposed

model with filter bank feature extraction, Riemannian methods, and adversarial domain adaptation effectively captures and differentiates emotion-related and decodes it which leads to state-of-art decoding results in cross-sensory EEG emotion recognition. Moreover, ensemble learning with the adapted filter bank Riemannian features could eliminate the influence of individual-specific frequency band variations potentially caused by individual differences, resulting in reduced standard deviation. From the baseline methods comparison, this study proposed and achieved EEG-based emotion recognition by introducing a comprehensive framework for cross-sensory EEG emotion recognition that overcomes the limitations of previous sensory-dependent approaches.

3.4.2 Comparative classification discussion

Comparative classification results from Figure 3.11 show that only adversarial domain adaptation, filter bank ensemble learning, or both affect the framework performance. Previous research has shown that filter combination learning [164][173] and adversarial domain adaptation in feature methods [172], these feature methods play an important role in the background of cross-sensory EEG emotion recognition. Filter bank ensemble learning combined with adversarial domain adaptation enhances Riemannian feature-based classification. From the evaluation of computational cost, compared with RIE, FBADR mainly increases the computational cost because the neural network-based domain adaptation training is associated with approximately six times the sub-bands of repeated operations, thus achieving better EEG decoding results.

Filter bank ensemble learning is an effective feature extraction method that can extract emotion-related feature information from cross-sensory emotion EEG data under multimodal stimulation. By exploring six filter banks within the frequency range of 1-50Hz, and utilizing meta classification method in ensemble learning, the channel covariance matrix-based Riemannian features can be made more distinguishable, thereby improving the accuracy of emotion recognition. Adversarial domain adaptation is an advanced domain adaptation method that can improve the model's generalization ability on the target domain by establishing a mapping relationship between the source domain and the target domain, it explored the adversarial domain adaptation with leave-one-sensory-out approach. In emotion recognition tasks, the source domain and the target domain often have different distribution characteristics, which requires adversarial domain adaptation methods to solve the distribution differences between domains, thereby improving the performance of the model in cross-sensory situations.

The comparison between FBADR and FBR indicates the success of implementation. Hence, filter bank ensemble learning, and adversarial domain adaptation, the usage of these two feature methods that can be combined with each other to realize and improve cross-sensory EEG emotion recognition. Moreover, for ensemble learning structure, both basic classifier and meta classifier could be adjusted to test the viability of further improvement. The number of basic classifiers is equal to number of filter banks. In future works, the different arrangements of filter banks would be associated with different number of basic classifiers. Meanwhile, the meta classifiers would be potential applicable for arranged into multi-layer structure, which are also worthful for future investigations.

3.4.3 Robustness verification discussion

The robustness verification result in Figure 3.14 shows that the proposed framework can maintain an average ACC of approximately 88% when the SNR ⩾1dB. Previous research has shown that in EEG-BCI applications, noise resistance is crucial to ensure system performance and reliability [175]. Noise can come from various sources, including electromagnetic interference, physiological interference, and environmental interference [176]. These interferences will affect the quality and reliability of EEG signals, and in turn, affect the performance and stability of the brain-computer interface system. Therefore, as for further applying the cross-sensory EEG emotion recognition system in EEG-aBCI, the impact of noise must be fully considered. The system's noise resistance, in the proposed FBADR methods, the robustness of feature methods of filter bank ensemble learning [173], and covariance matrix-based Riemannian feature extraction [174] mainly ensure that the proposed methods, which could be operated stably and reliably in practical applications.

The SNR ⩾ 1dB is considered a resealable viability for EEG noise robustness verification [177]. Since in EEG-BCI application, the ratio between signal and noise often changes dynamically [178], choosing SNR ⩾ 1dB can ensure that the signal has sufficient strength relative to the noise so that the system can recognize emotional information more stably. In addition, verification within this SNR range can also better simulate the noise situation in actual application scenarios, making the verification results more reliable and practical. The emotion recognition framework based on filter bank ensemble learning and adversarial domain adaptation proposed in this article has good noise resistance and can maintain a high level of recognition accuracy under the condition of SNR ⩾ 1dB. Noise resistance is one of the important indicators for evaluating the performance and stability of BCI systems. By performing

noise resistance verification on the proposed cross-sensory EEG emotion recognition method, the performance of the method in the face of different noise interferences can be verified, thus providing strong support for its reliability in practical applications. This provides strong support for its feasibility in EEG BCI applications.

Overall, this study contributes to the field of EEG emotion recognition by offering a comprehensive framework that effectively addresses the challenges in cross-sensory scenarios. Incorporating adversarial domain adaptation and ensemble learning techniques with filter-banked Riemannian features enables accurate emotion classification, while its robustness strengthens its potential application in EEG-based a-BCI.

3.5 Summary

This chapter introduced a novel framework to address the challenge of cross-sensory EEG emotion recognition in multimodal emotion stimulation as three sensory modalities: audio/visual/audio-visual with two emotion states: pleasure or unpleasure. To accomplish this, we conducted self-designed experiments involving multimodal emotion simulations to acquire cross-sensory emotion EEG data. Our proposed approach—filter bank adversarial domain adaptation Riemann method—leverages Riemannian tangent space methods and filter bank techniques to effectively extract features from cross-sensory emotion data. The feature methods were developed based on the feature analytic results in Chapter 2. A key innovation of our study is applying adversarial domain adaptation to mitigate domain differences in cross-sensory situations, enhancing emotion recognition performance. Specifically, we employed adapted features from the source and target domains to train the ensemble SVM classifiers. This integration of adversarial learning and ensemble learning methodologies successfully addressed the challenges associated with cross-sensory EEG emotion recognition, enabling the accurate binary classification of pleasurable and unpleasurable emotions.

We analyzed comparative classification results and demonstrated that our proposed FBADR framework achieved a state-of-the-art performance in cross-sensory emotion recognition, attaining an average accuracy of 89.01%±5.06%.89.01%±5.06%. Notably, this level of accuracy was the highest among comparable approaches, accompanied by a low standard deviation. Furthermore, we assessed the robustness of our framework by introducing Gaussian noise, which indicated that the framework was highly resilient to noise interference.

Overall, our study contributes to the field of EEG emotion recognition by offering a comprehensive framework that effectively addresses the challenges in cross-sensory scenarios.

Incorporating adversarial domain adaptation and ensemble learning techniques with filter-banked Riemannian features enables accurate emotion classification, while its robustness strengthens its potential for real-world applications.

4. General Discussion

4.1 Comprehensive result discussion.

In Chapter 2 and Chapter 3, this research aimed to solve the general problem of cross-sensory EEG emotion recognition under multimodal stimulation. In Chapter 2, this study focused on the emotion and sensory pattern with respect to emotion-related changes and sensory-related changes of EEG under multimodal stimulation. To explore the neurological cue of multimodal stimulation from EEG, it analyzed and obtained the feature analytic results from emotion EEG data under multimodal stimulation in Chapter 2 as Experiment 1. The key analytic results from Chapter 2 as feature characters' finding for emotion and sensory patterns are mainly three points with respect to spatial, spectral, and quantitative aspects respectively:

I. Emotion-related changes exhibit greater diversity rather than sensory-related changes in regional changes.

II. Emotion-related changes exhibit greater diversity rather than sensory-related changes in the sub-band.

III. Sensory-related changes exhibit greater spectral impact quantitatively rather than emotion-related changes.

The above-mentioned feature analytic results not only brought a comparative study for revealing emotion and sensory patterns of EEG under multimodal stimulation, which could neurologically explain the cross-sensory emotion perception. More importantly, the results could serve as feature analytic basements for the realization of cross-sensory EEG emotion recognition by the emotion and sensory-related changes features processing, to extract and decode emotion

from cross-sensory emotion EEG data in the next step. In Chapter 3. by designing and executing Experiment 2, as the cross-sensory EEG emotion recognition method contains filter bank ensemble learning with spatial characteristics, adversarial domain adaptation with spectral characteristics, and covariance matrix-based Riemannian feature extraction with spectral quantitative characteristics, this study proposed FBADR method. By using cross-sensory Emotion EEG data mentioned and acquired in Chapter 2 from Experiment 1, it realized the cross-sensory EEG emotion recognition. Baseline comparison, comparative classification, and robustness verification were taken for testing and validation of the proposed FBADR methods, which indicated the state-of-the-art results in comparison to related research and similar approaches. In contrast to previous studies in EEG a-BCI research [110][114][115], this research represents a comprehensive end-to-end investigation, starting from the fundamental experimental design, data collection, feature analysis, and culminating in the development of an emotion recognition framework. This approach sets this research apart in the field of affective EEG-BCI studies. Generally, this research represents a comprehensive investigation within the field of EEG-BCI studies. It contains experimental design, data acquisition, feature analysis, and the development of an end-to-end cross-sensory EEG emotion recognition framework, setting it apart from previous research efforts. Moreover, previous study indicated the cross-domain problems (cross-dataset, cross-day, cross subject) in physiological signals (ECG, fMRI, EMG…) analyzing and decoding [179-181] Chapter 2 and 3 proposed a feature analytical-based workflow for solving cross-sensory EEG emotion recognition, which is also potential viable for cross-domain problems in physiological signals.

4.2 Feature analysis of EEG under multimodal emotion stimulation

In the area of human perception and emotional response, the impact of multimodal stimuli is an important aspect. The inspection and understanding of specific types of stimulation is important in the field of neurological study of EEG. Though the psychological phenomenon of multimodal emotion stimulation has been recognized as cross-sensory emotion perception [45], how the brain processes and experiences phenomenon still lacks relative investigation. In this study from Chapter 2, the EEG features analytic results under multimodal emotion stimulation are supportive evidence for explaining cross-sensory emotion perception. The spatial, spectral, and quantitative feature characteristics of emotion and sensory patterns brought a comparative understanding of the role of emotion and sensory components from multimodal emotion stimulation. The results filled the previous blank in emotion and sensory-related research in the field of EEG feature investigation.

With the sensory-related changes, it showed more prevalent spectral variation in the sense of quantitative changes. Compared with emotion, the sensory is more primary and front in the process of experiencing the external world [182]. This natural sensory makes humans receive and process a huge amount of information surrounding us and also enables humans to maintain high sensitivity to adapt and feedback on the response from the sensory system [183]. On the other side, the emotion patterns in response to multimodal stimulation are different from sensory patterns. Compared with sensory-related changes, which showed more immediate and obvious reactions, emotion-related changes exhibit wider spectral variations in different frequency ranges. Emotion comes from more diverse aspects of influence, not only from mere sensory inputs. Emotion is subjected to multiple factors, including internal thoughts, memories, and previous

experiences [184]. The generation of emotion not only comes from sensory inputs but is also mixed with the internal cognitive process. Henceforth, the process of emotion is leading the variations of spectral changes to be wider and more diversified. Moreover, the diversification of emotion-related changes is more obvious rather than sensory-related changes. Emotion is rooted deeply in the mind and brain, which is associated with different functional areas within the brain [185]. Different emotions can activate distinct brain regions and neural pathways, resulting in distinct regional alterations [186]. This diversity in regional changes reflects the unique signature of each emotion, providing a characterized feature for each emotion processing. In contrast, sensory-related changes may exhibit more standardized neural responses, as they are primarily concerned with the initial processing of external sensory information [187] before it reach higher-level cognitive and emotional centers. Investigating for cross-domain task for physiological signals is crucial problem, ex: cross-day hand gesture for EMG [179], cross-session cardiovascular reactivity for fMRI [180], cross-subject mental stress for ECG [181]. Chapter 2 brought ideas for revealing cross-domain task physiological signals, with pattern analysis for deserved feature and cross-domain feature.

4.3 The model effectiveness for cross-sensory EEG emotion recognition and application

This part from Chapter 3 is the realization of cross-sensory emotion recognition under multimodal stimulation, which introduces a novel method focusing on multimodal emotion stimulation with audio, visual, and audio-visual of multimodal stimulation, and two emotion states as pleasure and unpleasure. The effectiveness of this framework for cross-sensory EEG emotion recognition is mainly based on the inspiration from the feature analysis of Chapter 2 and its potential applications could be concluded that: The proposed method achieved cross-sensory EEG emotion recognition to overcome the previous sensory-dependent methods. This innovation is essential because emotions are sophisticated, and different sensory inputs contribute to human emotional experiences. Under multimodal stimulation, input sensory changes (ex: mute, black screen) are normal [188], by integrating audio, visual, and audio-visual modalities this research enabled continuous emotion recognition like a human by one framework.

The use of filter bank techniques and Riemannian tangent space feature methods for feature extraction is a strong point of the method. These techniques can effectively capture the relevant information from EEG data, considering the spectral and spatial information in EEG signals [14]. This enhances the framework's ability to identify emotions accurately. The integration of adversarial domain adaptation is a significant advancement. It helps to reduce domain differences in cross-sensory situations, making the framework more adaptable and transferable. This is crucial because EEG data can vary significantly between individuals and scenarios [189]. The application of adversarial learning to adapt features from different domains demonstrates a strong understanding of the generalization of domain adaptation techniques [172]. The use of ensemble SVM classifiers further improves the framework's performance. Ensemble

learning methods with filter banks are known for their ability to improve classification accuracy by combining multiple models. In this case, it helps in achieving accurate emotion classification, which is a critical goal in emotion recognition tasks [173]. The reported average accuracy of 89.01% with a low standard deviation is crucial for conclusion with comparison for other relative models. It indicates that the framework can consistently and accurately classify emotions across different sensory modalities. The robustness demonstrated against Gaussian noise interference suggests that it has the potential for real-world applications, where noisy EEG data is common. The framework's effectiveness and robustness make it potential for various real-world applications. Together with the experimental and feature analytic parts in Chapter 2, it offered the realization of cross-sensory EEG emotion recognition under multimodal stimulation. The decoding for cross-domain conditions is key point for further applications of physiological signals [190]. Chapter 3 brought ideas for decoding cross-domain tasks for physiological signals with feature extraction of deserved pattern and reduce cross-domain differences by domain adaptation.

4.4 limitations and future works

This study achieved an average accuracy of around 89% in EEG-based emotion recognition, indicating room for further improvement. Some individual results fell below 80%, regardless of baseline methods and comparative classification. Additionally, the exploration of different types of emotions and sensory inputs is still in its fundamental stage. The relationship between remaining problems and future works could be considered as Figure 4.1.

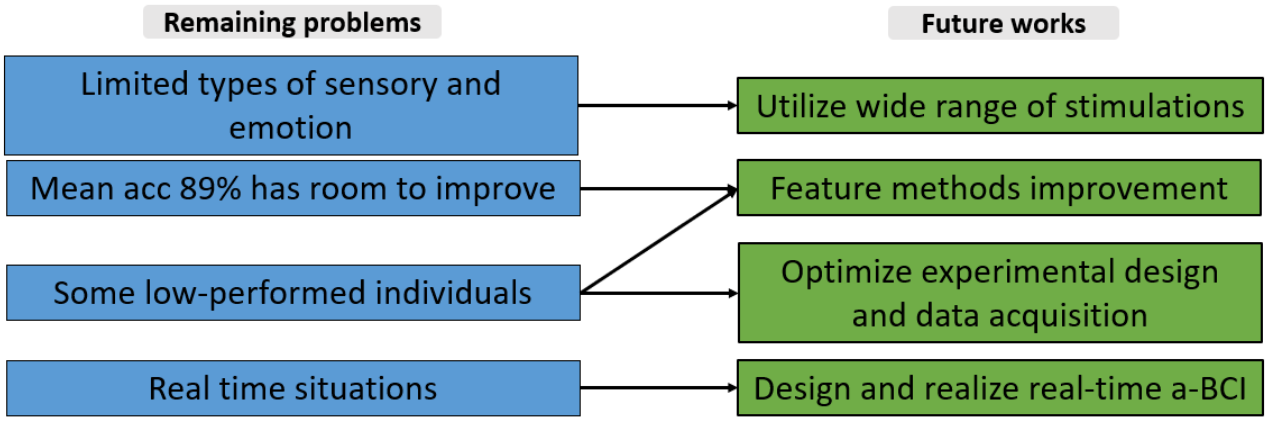


Figure 4.1. Relationship between remaining problems and future works.

Therefore, future work should consider several aspects to enhance this study's findings and applicability. Specifically, the basic step will be the increase of stimulations, for both the number of stimulation and types of stimulation. Then, corresponding with technological improvements for feature methods, we would further consider optimization of experimental design and data acquisition Finally, the long-term consideration will be the real-time application.

Model generalization and improvements: To enhance model generalization, future studies should include a broader range of video stimulations and recruit more participants. The multimodal nature of human sensory experience necessitates the inclusion of wide range of

emotion and sensory types in future work. Importantly, the cross-sensory emotion perception is not restricted to specific types of emotion and sensory [45]. For instance, incorporating olfactory (smell) and somatosensory (touch) stimuli, along with a wider range of emotional states such as anger and relaxation, could provide a more comprehensive analytic basement. Advanced algorithms for feature extraction and decoding, such as optimized channel selection and diffusion-based methods, should also be explored to further improve the robustness and accuracy of the proposed method.

EEG acquisition and experimental design: To further ensure stimulation-induced emotion EEG, improvements in experimental design and participant performance are essential. Optimizing the experimental setup can help reveal potential confounding factors and ensure more reliable data acquisition. This includes refining the EEG acquisition diagram and ensuring that the stimuli used further effectively induce the target emotional responses and reduce individual differences.

Real-time applications: For future applications, the benefits of cross-sensory EEG recognition could be further revealed with development real-time systems. These systems could be applied in various multimodal stimulation scenarios such as virtual VR, a-BCI-driven applications, and real-time gaming monitoring. The ability to accurately recognize emotions regardless of changes in sensory input would enhance user experience and system performance in these applications. For instance, in VR, real-time emotion recognition could adapt the environment to better suit the user's emotional state, while in gaming, it could provide real-time feedback to enhance gameplay and user engagement. The main challenge for real-time application is to design and build an effective BCI system for utilizing humans’ EEG with emotion to regulate external devices. The system implementation diagram is critical to real-time

applications. The online system application would be taken into priority, the main time cost will be two aspects: information transfer cost and data processing cost. Since the amount of information in the a-BCI is not a lot, normal WIFI or 5G situation could guarantee low time-delay communication. For data processing, which had been mentioned in section 3.3, the proposed method have relatively low computational cost. Furthermore, the data transfer will be more frequent, like twice in each second, for rapid and reliable a-BCI feedback.

Generally, future work on EEG-based emotion recognition would like to focus on enhancing model generalization through diverse stimuli and participant inclusion, refining experimental design. Additionally, real-time applications cooperated with cross-sensory recognition capabilities should be explored to create more adaptive and responsive systems. By integrated the mentioned aspect, this study could be further improved in field of EEG emotion recognition and a-BCI applications.

5. Conclusion

Cross-sensory EEG emotion recognition under multimodal stimulation, as an important issue in the research of EEG-based a-BCI, this study reveals characteristics of sensory and emotion patterns through conducting experimental design and feature analysis and proposed cross-sensory EEG emotion recognition framework (FBADR) applied and achieved SOTA performance. For this article, it has the following summary.

In Chapter 1, this study have an in-depth investigation on the research background related to cross-sensory EEG emotion recognition under multimodal stimulation, the research issues involved in emotion recognition and previous study, discussion of the EEG-based a-BCI system, as well as the purpose and research methods and strategies of this study.

In Chapter 2, it studied the basic EEG fundamentals of emotion EEG analysis under multimodal emotion stimulation. A multimodal emotion stimulation experiment based on audio-visual (video) was designed and implemented with six experimental conditions in total through two emotional changes (pleasure/unpleasure) and three groups of stimulation modalities (audio/visual/audio-visual). In Chapter 2, for the first time, a comparative study of emotion and sensory, two factors that affect cross-sensory emotion recognition, was conducted in emotion EEG analysis under multimodal stimulation. By using PSD-based spectral analysis, combined with brain functional area analysis and sub-band analysis, it comprehensively and systematically explored the relative emotion-related changes and sensory patterns of emotion pattern and sensory pattern under six experimental conditions. And in this chapter, this study reveals for the first time the comparative feature analysis of spectral, spatial and quantitative characteristics of emotion and sensory related patterns in emotion EEG analysis under multimodal emotion stimulation, that is, sensory and emotion analytic results brought EEG understanding for cross-

sensory emotion perception. At the same time, the feature analysis in Chapter 2 also provides strong support for subsequent implementation of cross-sensory EEG emotion recognition.

In Chapter 3, this study focuses on the framework design and implementation of cross-sensory EEG emotion recognition from the experiment of Chapter 2. Through the feature analysis of emotion pattern and sensory pattern based on Chapter 2, combined with the specific feature methods of EEG-related features from past related research, it can achieve the extraction and recognition of emotional components of cross-sensory EEG under multimodal stimulation. That is, for emotion changes showed frequency-varied, filter bank (band decomposition) with ensemble learning (sub-band results ensemble) effectively capture varied emotional components; emotion changes showed regional diversified, Riemannian feature extraction from covariance matrix (channel covariance calculation) can capture these diversities; also sensory-related changes showed greater spectral impact quantitatively, adversarial domain adaptation reduces the influence from sensory-related changes. Therefore, this study propose a framework with Filter Bank Riemannian Feature and Adversarial Domain Adaptation for cross-sensory emotion recognition. In this chapter, it is described the specific process of how to implement FBADR and the processing and application details of the cross-sensory EEG data. In the experiment of Chapter 3, mainly through comparative classification, baseline methods comparison and robustness verification, the SOTA results as 89.01%±5.06% accuracy of the proposed FBADR method for cross-sensory emotion recognition and its applicability in the a-BCI field were obtained.

In Chapter 4, it comprehensively discussed the separate and comprehensive roles played in solving the cross-sensory EEG emotion recognition problem in Chapter 2 and Chapter 3. That is, compared with past studies, this study has achieved end-to-end cross-sensory EEG emotion

recognition for the first time, and in this chapter, it discuss the feature analysis of cross-sensory EEG emotion recognition under multimodal stimulation and the effectiveness of proposed modal. .

In summary, this study investigated a crucial problem for cross-sensory EEG emotion recognition under multimodal stimulation. Conducting experimental design and feature analysis reveals characteristics of sensory and emotion patterns. Proposed cross-sensory EEG emotion recognition framework (FBADR) applied and achieved SOTA performance. Hence, this research indicated a comprehensive flow for inspecting and realizing the cross-sensory EEG emotion recognition under multimodal stimulation, widening the potential application for a-BCI.

Reference

1. Spezialetti, M., Cinque, L., Tavares, J. M. R., & Placidi, G. (2018). Towards EEG-based BCI driven by emotions for addressing BCI-Illiteracy: a meta-analytic review. *Behaviour & Information Technology*, *37*(8), 855-871.
2. Mridha, M. F., Das, S. C., Kabir, M. M., Lima, A. A., Islam, M. R., & Watanobe, Y. (2021). Brain-computer interface: Advancement and challenges. Sensors, 21(17), 5746.
3. Schalk, G., & Leuthardt, E. C. (2011). Brain-computer interfaces using electrocorticographic signals. *IEEE reviews in biomedical engineering*, *4*, 140-154.
4. Othmani, A., Brahem, B., & Haddou, Y. (2023). Machine learning-based approaches for post-traumatic stress disorder diagnosis using video and eeg sensors: A review. IEEE Sensors Journal.
5. Pinilla, A., Garcia, J., Raffe, W., Voigt-Antons, J. N., Spang, R. P., & Möller, S. (2021). Affective visualization in virtual reality: An integrative review. *Frontiers in Virtual Reality*, *2*, 630731.
6. Houssein, E. H., Hammad, A., & Ali, A. A. (2022). Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. *Neural Computing and Applications*, *34*(15), 12527-12557.
7. Shoumy, N. J., Ang, L. M., Seng, K. P., Rahaman, D. M., & Zia, T. (2020). Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals. Journal of Network and Computer Applications, 149, 102447.
8. Liu, Y., Sourina, O., & Nguyen, M. K. (2011). Real-time EEG-based emotion recognition and its applications. *Transactions on Computational Science XII: Special Issue on Cyberworlds*, 256-277.
9. Mladenović, J. (2021). Standardization of protocol design for user training in EEG-based brain–computer interface. *Journal of Neural Engineering*, *18*(1), 011003.
10. Landau, O., Puzis, R., & Nissim, N. (2020). Mind your mind: EEG-based brain-computer interfaces and their security in cyber space. *ACM Computing Surveys (CSUR)*, *53*(1), 1-38.
11. Sarma, P., & Barma, S. (2020). Review on stimuli presentation for affect analysis based on EEG. *IEEE Access*, *8*, 51991-52009.
12. Ko, K. E., Yang, H. C., & Sim, K. B. (2009). Emotion recognition using EEG signals with relative power values and Bayesian network. *International Journal of Control, Automation and Systems*, *7*, 865-870.
13. Duan, R. N., Zhu, J. Y., & Lu, B. L. (2013, November). Differential entropy feature for EEG-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 81-84). IEEE.
14. Wu, M., Hu, S., Wei, B., & Lv, Z. (2022). A novel deep learning model based on the ICA and Riemannian manifold for EEG-based emotion recognition. *Journal of Neuroscience Methods*, *378*, 109642.
15. Yohanes, R. E., Ser, W., & Huang, G. B. (2012, August). Discrete wavelet transform coefficients for emotion recognition from EEG signals. In *2012 annual international conference of the IEEE engineering in medicine and biology society* (pp. 2251-2254). IEEE.

16. Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T. L., Jeng, S. K., Duann, J. R., & Chen, J. H. (2010). EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, *57*(7), 1798-1806.
17. Mehmood, R. M., & Lee, H. J. (2016). A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns. *Computers & Electrical Engineering*, *53*, 444-457.
18. Liu, Y. J., Yu, M., Zhao, G., Song, J., Ge, Y., & Shi, Y. (2017). Real-time movie-induced discrete emotion recognition from EEG signals. *IEEE Transactions on Affective Computing*, *9*(4), 550-562.
19. Planalp, S. (1999). Communicating emotion: Social, moral, and cultural processes. Cambridge University Press.
20. Layder, D. (2004). Emotion in social life: The lost heart of society. Emotion in Social Life, 1-136.
21. Fredrickson, B. L. (2013). Love 2.0: Creating happiness and health in moments of connection. Penguin.
22. Seyfert, R. (2012). Beyond personal feelings and collective emotions: Toward a theory of social affect. Theory, culture & society, 29(6), 27-46.
23. Loving, T. J., Heffner, K. L., & Kiecolt-Glaser, J. K. (2006). Physiology and interpersonal relationships. The Cambridge handbook of personal relationships, 385-405.
24. Tremoleda, J. L., Kerton, A., & Gsell, W. (2012). Anaesthesia and physiological monitoring during in vivo imaging of laboratory rodents: considerations on experimental outcomes and animal welfare. *EJNMMI research*, *2*(1), 1-23.
25. Seligman, M. E. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. Simon and Schuster..
26. Ezzameli, K., & Mahersia, H. (2023). Emotion recognition from unimodal to multimodal analysis: A review. *Information Fusion*, 101847.
27. Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. *Sensors*, *20*(3), 592.
28. Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research*, *33*(3), 217-238.
29. Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281-291.
30. Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological psychiatry*, *54*(5), 504-514.
31. Basu, S., Chakraborty, J., & Aftabuddin, M. (2017, October). Emotion recognition from speech using convolutional neural network with recurrent neural network architecture. In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)* (pp. 333-336). IEEE.
32. Ng, H. W., Nguyen, V. D., Vonikakis, V., & Winkler, S. (2015, November). Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 443-449).

33. Hibbeln, M., Jenkins, J. L., Schneider, C., Valacich, J. S., & Weinmann, M. (2017). How is your user feeling? Inferring emotion through human–computer interaction devices. Mis Quarterly, 41(1), 1-22.
34. El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern recognition*, *44*(3), 572-587.
35. Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, *128*(2), 203.
36. Feidakis, M. (2016). A review of emotion-aware systems for e-learning in virtual environments. *Formative assessment, learning data analytics and gamification*, 217-242.
37. Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, *2*(5), 377.
38. Egger, M., Ley, M., & Hanke, S. (2019). Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, *343*, 35-55.
39. Stemmler, G., & Wacker, J. (2010). Personality, emotion, and individual differences in physiological responses. *Biological psychology*, *84*(3), 541-551.
40. Taylor, S. F., Kang, J., Brege, I. S., Tso, I. F., Hosanagar, A., & Johnson, T. D. (2012). Meta-analysis of functional neuroimaging studies of emotion perception and experience in schizophrenia. *Biological psychiatry*, *71*(2), 136-145.
41. Brunswik, E. (2023). *Perception and the representative design of psychological experiments*. Univ of California Press.
42. Engelmann, J. B., & Pogosyan, M. (2013). Emotion perception across cultures: the role of cognitive mechanisms. *Frontiers in psychology*, *4*, 118.
43. Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. Neuroimage, 37(4), 1445-1456.
44. Cappe, C., Rouiller, E. M., & Barone, P. (2012). Cortical and thalamic pathways for multisensory and sensorimotor interplay.
45. Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice, and touch: comparisons and convergence. Trends in cognitive sciences, 21(3), 216-228.
46. Massaeli, F., Bagheri, M., & Power, S. D. (2023). EEG-based detection of modality-specific visual and auditory sensory processing. *Journal of Neural Engineering*, *20*(1), 016049.
47. Kim, M. K., Kim, M., Oh, E., & Kim, S. P. (2013). A review on the computational methods for emotional state estimation from the human EEG. *Computational and mathematical methods in medicine*, *2013*.
48. Berthold-Losleben, M., Habel, U., Brehl, A. K., Freiherr, J., Losleben, K., Schneider, F., ... & Kohn, N. (2018). Implicit affective rivalry: A behavioral and fMRI study combining olfactory and auditory stimulation. *Frontiers in behavioral neuroscience*, *12*, 313.
49. Castiblanco Jimenez, I. A., Gomez Acevedo, J. S., Olivetti, E. C., Marcolin, F., Ulrich, L., Moos, S., & Vezzetti, E. (2022). User engagement comparison between advergames and traditional advertising using EEG: Does the user's engagement influence purchase intention?. *Electronics*, *12*(1), 122.

50. Bazzani, A., Ravaioli, S., Trieste, L., Faraguna, U., & Turchetti, G. (2020). Is EEG suitable for marketing research? A systematic review. *Frontiers in Neuroscience*, *14*, 594566.
51. Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., ... & Yang, X. (2018). A review of emotion recognition using physiological signals. *Sensors*, *18*(7), 2074.
52. He, Z., Li, Z., Yang, F., Wang, L., Li, J., Zhou, C., & Pan, J. (2020). Advances in multimodal emotion recognition based on brain–computer interfaces. *Brain sciences*, *10*(10), 687.
53. Torres, E. P., Torres, E. A., Hernández-Álvarez, M., & Yoo, S. G. (2020). EEG-based BCI emotion recognition: A survey. *Sensors*, *20*(18), 5083.
54. Murugappan, M., Juhari, M. R. B. M., Nagarajan, R., & Yaacob, S. (2009). An Investigation on visual and audiovisual stimulus based emotion recognition using EEG. *International Journal of Medical Engineering and Informatics*, *1*(3), 342-356.
55. Xing, B., Zhang, H., Zhang, K., Zhang, L., Wu, X., Shi, X., ... & Zhang, S. (2019). Exploiting EEG signals and audiovisual feature fusion for video emotion recognition. *IEEE Access*, *7*, 59844-59861.
56. Jiang, J., Bailey, K., & Xiao, X. (2018). Midfrontal theta and posterior parietal alpha band oscillations support conflict resolution in a masked affective priming task. *Frontiers in human neuroscience*, *12*, 175.
57. Mu, Y., Han, S., & Gelfand, M. J. (2017). The role of gamma interbrain synchrony in social coordination when humans face territorial threats. *Social cognitive and affective neuroscience*, *12*(10), 1614-1623.
58. Schelenz, P. D., Klasen, M., Reese, B., Regenbogen, C., Wolf, D., Kato, Y., & Mathiak, K. (2013). Multisensory integration of dynamic emotional faces and voices: method for simultaneous EEG-fMRI measurements. *Frontiers in Human Neuroscience*, *7*, 729.
59. Li, W., Li, Y., & Cao, D. (2021). The effectiveness of emotion cognitive reappraisal as measured by self-reported response and its link to EEG alpha asymmetry. *Behavioural Brain Research*, *400*, 113042.
60. Yu, W., Wu, X., Chen, Y., Liang, Z., Jiang, J., Misrani, A., ... & Yang, L. (2021). Pelvic pain alters functional connectivity between anterior cingulate cortex and hippocampus in both humans and a rat model. *Frontiers in systems neuroscience*, *15*, 642349.
61. Sonkusare, S., Nguyen, V. T., Moran, R., van der Meer, J., Ren, Y., Koussis, N., ... & Guo, C. (2020). Intracranial-EEG evidence for medial temporal pole driving amygdala activity induced by multi-modal emotional stimuli. *Cortex*, *130*, 32-48.
62. Ahirwal, M. K., & Kose, M. R. (2020). Audio-visual stimulation based emotion classification by correlated EEG channels. *Health and Technology*, *10*(1), 7-23.
63. Balconi, M., & Vanutelli, M. E. (2016). Vocal and visual stimulation, congruence and lateralization affect brain oscillations in interspecies emotional positive and negative interactions. *Social neuroscience*, *11*(3), 297-310.

64. Balconi, M., & Carrera, A. (2011). Cross-modal integration of emotional face and voice in congruous and incongruous pairs: The P2 ERP effect. *Journal of Cognitive Psychology*, *23*(1), 132-139.
65. Gallotto, S., Sack, A. T., Schuhmann, T., & de Graaf, T. A. (2017). Oscillatory correlates of visual consciousness. *Frontiers in psychology*, *8*, 258566.
66. Craddock, M., Poliakoff, E., El-Deredy, W., Klepousniotou, E., & Lloyd, D. M. (2017). Pre-stimulus alpha oscillations over somatosensory cortex predict tactile misperceptions. *Neuropsychologia*, *96*, 9-18.
67. Klasen, M., Kreifelts, B., Chen, Y. H., Seubert, J., & Mathiak, K. (2014). Neural processing of emotion in multimodal settings. Frontiers in human neuroscience, 8, 822.
68. Bugos, J. A., Bidelman, G. M., Moreno, S., Shen, D., Lu, J., & Alain, C. (2022). Music and visual art training increase auditory-evoked theta oscillations in older adults. *Brain Sciences*, 12(10), 1300.
69. Zimmer, U., Wendt, M., & Pacharra, M. (2022). Enhancing allocation of visual attention with emotional cues presented in two sensory modalities. *Behavioral and Brain Functions*, 18(1), 10.
70. Ross, B., Dobri, S., Jamali, S., & Bartel, L. (2022). Entrainment of somatosensory beta and gamma oscillations accompany improvement in tactile acuity after periodic and aperiodic repetitive sensory stimulation. *International Journal of Psychophysiology*, 177, 11-26.
71. Magosso, E., De Crescenzio, F., Ricci, G., Piastra, S., & Ursino, M. (2019). EEG alpha power is modulated by attentional changes during cognitive tasks and virtual reality immersion. Computational intelligence and neuroscience, 2019.
72. Cantero, J. L., Atienza, M., Salas, R. M., & Dominguez-Marin, E. (2002). Effects of prolonged waking-auditory stimulation on electroencephalogram synchronization and cortical coherence during subsequent slow-wave sleep. Journal of Neuroscience, 22(11), 4702-4708.
73. Baumgartner, T., Esslen, M., & Jäncke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. International journal of psychophysiology, 60(1), 34-43.
74. Jääskeläinen, I. P., Sams, M., Glerean, E., & Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage*, *224*, 117445.
75. Mulders, D., de Bodt, C., Lejeune, N., Courtin, A., Liberati, G., Verleysen, M., & Mouraux, A. (2020). Dynamics of the perception and EEG signals triggered by tonic warm and cool stimulation. *PLoS One*, *15*(4), e0231698.
76. Murugappan, M., & Murugappan, S. (2013, March). Human emotion recognition through short time Electroencephalogram (EEG) signals using Fast Fourier Transform (FFT). In *2013 IEEE 9th International Colloquium on Signal Processing and its Applications* (pp. 289-294). IEEE.

77. Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention–a possible marker of engagement?. *Frontiers in human neuroscience*, *6*, 112.
78. Deng, Y., Yang, M., & Zhou, R. (2017). A new standardized emotional film database for Asian culture. *Frontiers in psychology*, *8*, 288291.
79. Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS one*, *11*(2), e0148037.
80. Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.
81. Stancin, I., Cifrek, M., & Jovic, A. (2021). A review of EEG signal features and their application in driver drowsiness detection systems. *Sensors*, *21*(11), 3786.
82. Rahman, M. M., Sarkar, A. K., Hossain, M. A., Hossain, M. S., Islam, M. R., Hossain, M. B., ... & Moni, M. A. (2021). Recognition of human emotions using EEG signals: A review. *Computers in Biology and Medicine*, *136*, 104696.
83. Liao, D., Shu, L., Liang, G., Li, Y., Zhang, Y., Zhang, W., & Xu, X. (2019). Design and evaluation of affective virtual reality system based on multimodal physiological signals and self-assessment manikin. *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, *4*(3), 216-224.
84. Knyazev, G. G. (2007). Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neuroscience & Biobehavioral Reviews*, *31*(3), 377-395.
85. Mensen, A., Marshall, W., & Tononi, G. (2017). EEG differentiation analysis and stimulus set meaningfulness. *Frontiers in psychology*, *8*, 274988.
86. de Haan, E. H., Corballis, P. M., Hillyard, S. A., Marzi, C. A., Seth, A., Lamme, V. A., ... & Pinto, Y. (2020). Split-brain: what we know now and why this is important for understanding consciousness. *Neuropsychology review*, *30*, 224-233.
87. Ding, L., Shou, G., Yuan, H., Urbano, D., & Cha, Y. H. (2014). Lasting modulation effects of rTMS on neural activity and connectivity as revealed by resting-state EEG. IEEE Transactions on Biomedical Engineering, 61(7), 2070-2080.
88. Ong, Z. Y., Saidatul, A., & Ibrahim, Z. (2018, August). Power spectral density analysis for human EEG-based biometric identification. In *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)* (pp. 1-6). IEEE.
89. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *neuroimage*, *86*, 446-460.
90. Uzun, S. S., Yildirim, S., & Yildirim, E. (2012, January). Emotion primitives estimation from EEG signals using Hilbert Huang Transform. In *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics* (pp. 224-227). IEEE.

91. Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Gentili, C., ... & Valenza, G. (2019). Real vs. immersive-virtual emotional experience: Analysis of psycho-physiological patterns in a free exploration of an art museum. *PloS one*, *14*(10), e0223881.
92. Girden, E. R. (1992). *ANOVA: Repeated measures* (No. 84). sage.
93. Suk, H. J., & Irtel, H. (2010). Emotional response to color across media. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, *35*(1), 64-77.
94. Lagopoulos, J., Xu, J., Rasmussen, I., Vik, A., Malhi, G. S., Eliassen, C. F., ... & Ellingsen, Ø. (2009). Increased theta and alpha EEG activity during nondirective meditation. *The journal of alternative and complementary medicine*, *15*(11), 1187-1192.
95. Übeyli, E. D. (2009). Statistics over features: EEG signals analysis. *Computers in Biology and Medicine*, *39*(8), 733-741.
96. Grasso, P. A., Pietrelli, M., Zanon, M., Làdavas, E., & Bertini, C. (2020). Alpha oscillations reveal implicit visual processing of motion in hemianopia. *Cortex*, *122*, 81-96.
97. Izhar, L. I., Babiker, A., Rizki, E. E., Lu, C. K., & Abdul Rahman, M. (2022). Emotion self-regulation in neurotic students: A pilot mindfulness-based intervention to assess its effectiveness through brain signals and behavioral data. *Sensors*, *22*(7), 2703.
98. Van der Burg, E., Toet, A., Brouwer, A. M., & Van Erp, J. B. (2021). Serial dependence of emotion within and between stimulus sensory modalities. *Multisensory research*, *35*(2), 151-172.
99. Von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *International journal of psychophysiology*, *38*(3), 301-313.
100. Alarcao, S. M., & Fonseca, M. J. (2017). Emotions recognition using EEG signals: A survey. *IEEE transactions on affective computing*, *10*(3), 374-393.
101. Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in psychology*, *2*, 10747.
102. Mueller, S. C. (2011). The influence of emotion on cognitive control: relevance for development and adolescent psychopathology. *Frontiers in psychology*, *2*, 327.
103. Lestienne, R. (2001). Spike timing, synchronization and information processing on the sensory side of the central nervous system. *Progress in neurobiology*, *65*(6), 545-591.
104. Yoon, H. J., & Chung, S. Y. (2011, October). EEG spectral analysis in valence and arousal dimensions of emotion. In *2011 11th International Conference on Control, Automation and Systems* (pp. 1319-1322). IEEE.

105. Poza, J., Gómez, C., Gutiérrez, M. T., Mendoza, N., & Hornero, R. (2013). Effects of a multi-sensory environment on brain-injured patients: Assessment of spectral patterns. *Medical engineering & physics*, *35*(3), 365-375.
106. Ertl, M., Hildebrandt, M., Ourina, K., Leicht, G., & Mulert, C. (2013). Emotion regulation by cognitive reappraisal—the role of frontal theta oscillations. *NeuroImage*, *81*, 412-421.
107. Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in cognitive sciences*, *16*(7), 390-398.
108. Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, *54*(5), 677-696.
109. Gyurak, A., Gross, J. J., & Etkin, A. (2011). Explicit and implicit emotion regulation: A dual-process framework. *Cognition and emotion*, *25*(3), 400-412.
110. Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*, *59*, 103-126.
111. Bota, P. J., Wang, C., Fred, A. L., & Da Silva, H. P. (2019). A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE access*, *7*, 140990-141020.
112. Hramov, A. E., Maksimenko, V. A., & Pisarchik, A. N. (2021). Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states. *Physics Reports*, *918*, 1-133.
113. Mühl, C., Allison, B., Nijholt, A., & Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Computer Interfaces*, *1*(2), 66-84.
114. Gu, X., Cao, Z., Jolfaei, A., Xu, P., Wu, D., Jung, T. P., & Lin, C. T. (2021). EEG-based brain-computer interfaces (BCIs): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. *IEEE/ACM transactions on computational biology and bioinformatics*, *18*(5), 1645-1666.
115. Liu, S., Wang, Z., An, Y., Zhao, J., Zhao, Y., & Zhang, Y. D. (2023). EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network. *Knowledge-Based Systems*, *265*, 110372.
116. Li, D., Xie, L., Wang, Z., & Yang, H. (2023). Brain emotion perception inspired eeg emotion recognition with deep reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*.
117. Padhmashree, V., & Bhattacharyya, A. (2022). Human emotion recognition based on time–frequency analysis of multivariate EEG signal. *Knowledge-Based Systems*, *238*, 107867.

118. Wei, Y., Liu, Y., Li, C., Cheng, J., Song, R., & Chen, X. (2023). TC-Net: A Transformer Capsule Network for EEG-based emotion recognition. *Computers in biology and medicine*, *152*, 106463.
119. Cui, F., Wang, R., Ding, W., Chen, Y., & Huang, L. (2022). A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition. *Mathematics*, *10*(4), 582.
120. Algarni, M., Saeed, F., Al-Hadhrami, T., Ghabban, F., & Al-Sarem, M. (2022). Deep learning-based approach for emotion recognition using electroencephalography (EEG) signals using bi-directional long short-term memory (Bi-LSTM). *Sensors*, *22*(8), 2976.
121. Islam, M. R., Islam, M. M., Rahman, M. M., Mondal, C., Singha, S. K., Ahmad, M., ... & Moni, M. A. (2021). EEG channel correlation based model for emotion recognition. *Computers in Biology and Medicine*, *136*, 104757.
122. Peng, Y., Qin, F., Kong, W., Ge, Y., Nie, F., & Cichocki, A. (2021). GFIL: A unified framework for the importance analysis of features, frequency bands, and channels in EEG-based emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, *14*(3), 935-947.
123. Huang, D., Zhang, H., Ang, K., Guan, C., Pan, Y., Wang, C., & Yu, J. (2012, March). Fast emotion detection from EEG using asymmetric spatial filtering. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 589-592). IEEE.
124. Chen, R., Sun, Z., Diao, X., Wang, H., Wang, J., Li, T., & Wang, Y. (2022). Happy or sad? Recognizing emotions with wavelet coefficient energy mean of EEG signals. *Technology and Health Care*, *30*(4), 937-949.
125. Wang, Y., Qiu, S., Ma, X., & He, H. (2021). A prototype-based SPD matrix network for domain adaptation EEG emotion recognition. Pattern Recognition, 110, 107626.
126. Ranasinghe, N., Jain, P., Karwita, S., Tolley, D., & Do, E. Y. L. (2017, May). Ambiotherm: enhancing sense of presence in virtual reality by simulating real-world environmental conditions. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 1731-1742).
127. Zhu, L., Zhu, Z., Zhang, C., Xu, Y., & Kong, X. (2023). Multimodal sentiment analysis based on fusion methods: A survey. *Information Fusion*, *95*, 306-325.
128. Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, *1*(1), 18-37.
129. Wang, Q., Wang, M., Yang, Y., & Zhang, X. (2022). Multi-modal emotion recognition using EEG and speech signals. *Computers in Biology and Medicine*, *149*, 105907.
130. Tsiourti, C., Weiss, A., Wac, K., & Vincze, M. (2019). Multimodal integration of emotional signals from voice, body, and context: Effects of (in) congruence on emotion

recognition and attitudes towards robots. *International Journal of Social Robotics*, *11*, 555-573.

131. Robins, D. L., Hunyadi, E., & Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain and cognition*, *69*(2), 269-278.
132. Saarimäki, H. (2021). Naturalistic stimuli in affective neuroimaging: A review. *Frontiers in human neuroscience*, *15*, 675068.
133. Cimtay, Y., Ekmekcioglu, E., & Caglar-Ozhan, S. (2020). Cross-subject multimodal emotion recognition based on hybrid fusion. *IEEE Access*, *8*, 168865-168878.
134. Li, J., Li, S., Pan, J., & Wang, F. (2021). Cross-subject EEG emotion recognition with self-organized graph neural network. *Frontiers in Neuroscience*, *15*, 611653.
135. Shen, X., Liu, X., Hu, X., Zhang, D., & Song, S. (2022). Contrastive learning of subject-invariant EEG representations for cross-subject emotion recognition. *IEEE Transactions on Affective Computing*.
136. Wang, Y., Qiu, S., Li, D., Du, C., Lu, B. L., & He, H. (2022). Multi-modal domain adaptation variational autoencoder for EEG-based emotion recognition. *IEEE/CAA Journal of Automatica Sinica*, *9*(9), 1612-1626.
137. Guo, W., Xu, G., & Wang, Y. (2023). Multi-source domain adaptation with spatio-temporal feature extractor for EEG emotion recognition. *Biomedical Signal Processing and Control*, *84*, 104998.
138. He, Z., Zhong, Y., & Pan, J. (2022, May). Joint temporal convolutional networks and adversarial discriminative domain adaptation for EEG-based cross-subject emotion recognition. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3214-3218). IEEE.
139. Sartipi, S., & Cetin, M. (2023, April). Adversarial discriminative domain adaptation and transformers for EEG-based cross-subject emotion recognition. In *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 1-4). IEEE.
140. Gao, C.; Uchitomi, H.; Miyake, Y. Influence of multimodal emotional stimulations on brain activity: An electroencephalographic study. Sensors 2023, 23, 4801.
141. Yang, Y. X., Gao, Z. K., Wang, X. M., Li, Y. L., Han, J. W., Marwan, N., & Kurths, J. (2018). A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(8).
142. Du, G., Su, J., Zhang, L., Su, K., Wang, X., Teng, S., & Liu, P. X. (2022). A multi-dimensional graph convolution network for EEG emotion recognition. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-11.
143. Wang, Z. M., Chen, Z. Y., & Zhang, J. (2023). EEG emotion recognition based on PLV-rich-club dynamic brain function network. Applied Intelligence, 53(14), 17327-17345.

144. Tian, C., Ma, Y., Cammon, J., Fang, F., Zhang, Y., & Meng, M. (2023). Dual-encoder VAE-GAN with spatiotemporal features for emotional EEG data augmentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

145. Nasrin, F., & Ahmed, N. I. (2021, August). Predicting the correctness of mental arithmetic task from EEG using deep learning. In *2021 International Conference on Science & Contemporary Technologies (ICSCT)* (pp. 1-5). IEEE.

146. Song, T., Zheng, W., Lu, C., Zong, Y., Zhang, X., & Cui, Z. (2019). MPED: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access*, *7*, 12177-12191.

147. Babiloni, C., Frisoni, G., Steriade, M., Bresciani, L., Binetti, G., Del Percio, C., ... & Rossini, P. M. (2006). Frontal white matter volume and delta EEG sources negatively correlate in awake subjects with mild cognitive impairment and Alzheimer's disease. *Clinical Neurophysiology*, *117*(5), 1113-1129.

148. Babiloni, C., Squitti, R., Del Percio, C., Cassetta, E., Ventriglia, M. C., Ferreri, F., ... & Rossini, P. M. (2007). Free copper and resting temporal EEG rhythms correlate across healthy, mild cognitive impairment, and Alzheimer's disease subjects. *Clinical Neurophysiology*, *118*(6), 1244-1260.

149. Li, F., Xia, Y., Wang, F., Zhang, D., Li, X., & He, F. (2020). Transfer learning algorithm of P300-EEG signal based on XDAWN spatial filter and Riemannian geometry classifier. *Applied Sciences*, *10*(5), 1804.

150. Harry L. Van Trees; Kristine L. Bell. (2007), "Covariance, Subspace, and Intrinsic CramrRao Bounds," in *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking* , IEEE, .430-450

151. Chen, Y., Wiesel, A., Eldar, Y. C., & Hero, A. O. (2010). Shrinkage algorithms for MMSE covariance estimation. *IEEE transactions on signal processing*, *58*(10), 5016-5029.

152. Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2011). Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, *59(4)*, 920-928.

153. Harandi, M. T., Salzmann, M., & Hartley, R. (2014). From manifold to manifold: Geometry-aware dimensionality reduction for SPD matrices. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13* (pp. 17-32). Springer International Publishing.

154. Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, *112*, 172-178.

155. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

156. Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In *International conference on machine learning* (pp. 214-223). PMLR.
157. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. *Advances in neural information processing systems*, *30*.
158. Zhang, Z., Li, M., & Yu, J. (2018). On the convergence and mode collapse of GAN. In SIGGRAPH Asia 2018 Technical Briefs (pp. 1-4).
159. Ghonima, R. (2021, December). Implementation of GANs using federated learning. In *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)* (pp. 142-148). IEEE.
160. Cirillo, M. D., Abramian, D., & Eklund, A. (2021). Vox2Vox: 3D-GAN for brain tumour segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6* (pp. 274-284). Springer International Publishing.
161. Yang, T., Wu, T., Li, L., & Zhu, C. (2020). SUD-GAN: deep convolution generative adversarial network combined with short connection and dense block for retinal vessel segmentation. *Journal of digital imaging*, *33*, 946-957.
162. Hardy, C., Le Merrer, E., & Sericola, B. (2019, May). Md-gan: Multi-discriminator generative adversarial networks for distributed datasets. In *2019 IEEE international parallel and distributed processing symposium (IPDPS)* (pp. 866-877). IEEE.
163. Ganihar, S. A., Joshi, S., Shetty, S., & Mudenagudi, U. (2014). Metric tensor and christoffel symbols based 3d object categorization. In *ACM SIGGRAPH 2014 Posters* (pp. 1-1).
164. Töscher, A., Jahrer, M., & Bell, R. M. (2009). The bigchaos solution to the netflix grand
165. Tang, J., Liang, J., Han, C., Li, Z., & Huang, H. (2019). Crash injury severity analysis using a two-layer Stacking framework. *Accident Analysis & Prevention*, *122*, 226-238.prize. *Netflix prize documentation*, 1-52.
166. She, Q., Chen, T., Fang, F., Zhang, J., Gao, Y., & Zhang, Y. (2023). Improved domain adaptation network based on wasserstein distance for motor imagery EEG classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 1137-1148.
167. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
168. Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(11).

169. Li, M., Xu, H., Liu, X., & Lu, S. (2018). Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technology and health care*, *26*(S1), 509-519.

170. Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering*, *15*(5), 056013.

171. Li, H., Jin, Y. M., Zheng, W. L., & Lu, B. L. (2018). Cross-subject emotion recognition using deep adaptation networks. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V 25* (pp. 403-413). Springer International Publishing.

172. Luo, Y., Zhang, S. Y., & Zheng, W. L. (2018). Bao-Liang Lu. WGAN domain adaptation for eeg-based emotion recognition. In *International Conference on Neural Information Processing* (pp. 275-286).

173. Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2008, June). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)* (pp. 2390-2397). IEEE.

174. Li, X., Zhang, Y., Tiwari, P., Song, D., Hu, B., Yang, M., ... & Marttinen, P. (2022). EEG based emotion recognition: A tutorial and review. *ACM Computing Surveys*, *55*(4), 1-57.

175. Minguillon, J., Lopez-Gordo, M. A., & Pelayo, F. (2017). Trends in EEG-BCI for daily-life: Requirements for artifact removal. *Biomedical Signal Processing and Control*, *31*, 407-418.

176. Mannan, M. M. N., Kamran, M. A., & Jeong, M. Y. (2018). Identification and removal of physiological artifacts from electroencephalogram signals: A review. *Ieee Access*, *6*, 30630-30652.

177. Sun, W., Su, Y., Wu, X., & Wu, X. (2020). A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals. *Neurocomputing*, *404*, 108-121.

178. Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., & Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of neural engineering*, *16*(1), 011001.

179. Hu, Q., Azar, G. A., Fletcher, A., Rangan, S., & Atashzar, S. F. (2024). ViT-MDHGR: Cross-day Reliability and Agility in Dynamic Hand Gesture Prediction via HD-sEMG Signal Decoding. *IEEE Journal of Selected Topics in Signal Processing*.

180. Sheu, L. K., Jennings, J. R., & Gianaros, P. J. (2012). Test–retest reliability of an fMRI paradigm for studies of cardiovascular reactivity. *Psychophysiology*, *49*(7), 873-884.

181. Ye, Y., Luo, T., Huang, W., Sun, Y., & Li, L. (2022, October). ECG-based cross-subject mental stress detection via discriminative clustering enhanced adversarial domain

adaptation. In *2022 16th IEEE International Conference on Signal Processing (ICSP)* (Vol. 1, pp. 495-499). IEEE.

182. Solms, M., & Turnbull, O. (2018). *The brain and the inner world: An introduction to the neuroscience of subjective experience*. Routledge.
183. Aron, E. N., Aron, A., & Jagiellowicz, J. (2012). Sensory processing sensitivity: A review in the light of the evolution of biological responsivity. Personality and Social Psychology Review, 16(3), 262-282.
184. Holland, A. C., & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of life reviews*, *7*(1), 88-131.
185. Panksepp, J. (2003). At the interface of the affective, behavioral, and cognitive neurosciences: Decoding the emotional feelings of the brain. *Brain and cognition*, *52*(1), 4-14.
186. Dolcos, F., Iordan, A. D., & Dolcos, S. (2011). Neural correlates of emotion–cognition interactions: A review of evidence from brain imaging investigations. *Journal of Cognitive Psychology*, *23*(6), 669-694.
187. Romo, R., & de Lafuente, V. (2013). Conversion of sensory signals into perceptual decisions. Progress in neurobiology, 103, 41-75.
188. Bertelson, P., & De Gelder, B. (2004). The psychology of multimodal perception. *Crossmodal space and crossmodal attention*, 141-177.
189. Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P., & Müller, K. R. (2006). Towards adaptive classification for BCI. *Journal of neural engineering*, *3*(1), R13.
190. Gordon, S. M., Jaswa, M., Solon, A. J., & Lawhern, V. J. (2017, March). Real world BCI: cross-domain learning and practical applications. In Proceedings of the 2017 ACM Workshop on An Application-oriented Approach to BCI out of the laboratory (pp. 25-28).

Appendix

Table A1. The corresponding relationship between experimental videos and sources from the film database (Unpleasure (sad: bs), pleasure: yy).

Unpleasure(sad) videos	1	2	3	4	5	6	7	8	9	10
Source from film database	bs1 0:00 ~0:30	bs2 0:00 ~0:30	bs6 0:00 ~0:30	bs11 0:00 ~0:30	bs12 0:00 ~0:30	bs13 0:00 ~0:30	bs15 0:00 ~0:30	bs16 0:00 ~0:30	bs1 0:40 ~1:10	bs2 0:50 ~1:20
Pleasure videos	1	2	3	4	5	6	7	8	9	10
Source from film database	yy1 0:00 ~0:30	yy2 0:00 ~0:30	yy5 0:00 ~0:30	yy11 0:00 ~0:30	yy13 0:00 ~0:30	yy14 0:00 ~0:30	yy15 0:00 ~0:30	yy16 0:00 ~0:30	yy1 0:40 ~1:10	yy2 0:50 ~1:20

A2. Pseudo-random protocol for the experiment

Audio-unpleasure stimuli are named 1-10, visual-unpleasure stimuli are named 11-20, and audio–visual are named 21-30 (corresponding to the order of the Supplementary Table). Audio-pleasure stimuli are named 31-40, visual-pleasure stimuli are named 41-50, and audio-pleasure are named 51-60 (with the order of supplementary table). Additionally, by using the following Python code, the order of 60 stimuli would be determined for each subject with respect to a pseudorandom order.

```
import random
# Generate 60 pseudorandom integers between 1 and 60
for i in range(60):
    rand_num = random.randint(1, 60)
    print(rand_num)
```

A3. Questionnaire before emotion EEG experiment

Questionnaire time: year month day hour minute

Are you left-handed? (Y/N)

Did you sleep well last night? (Y/N)

Did you have mental illnesses recently, history of mental illness or been treated by a psychotherapist? (Y/N, if YES, please talk to the experimenter immediately)

Did you have physical injuries recently? (Y/N, if YES, please talk to the experimenter immediately)

Has there been anything that has made you feel seriously emotionally affected recently? (Y/N, if YES, please talk to the experimenter immediately)

Do you have hearing or visual impairment (Y/N, if YES, please talk to the experimenter immediately)