PAPER

Decoding auditory-evoked response in affective states using wearable around-ear EEG system

Biomedical Physics & Engineering Express

PAPER

Decoding auditory-evoked response in affective states using wearable around-ear EEG system

RECEIVED
10 April 2023

REVISED
4 August 2023

ACCEPTED FOR PUBLICATION
17 August 2023

PUBLISHED
25 August 2023

Jaehoon Choi[1,5], **Netiwit Kaongoen**[1,5], **HyoSeon Choi**[2], **Minuk Kim**[3], **Byung Hyung Kim**[2,4,*] and **Sungho Jo**[1,*]

[1] School of Computing, KAIST, Daejeon, Republic of Korea
[2] Department of Electrical and Computer Engineering, Inha University, Incheon, Republic of Korea
[3] School of Electrical Engineering, KAIST, Daejeon, Republic of Korea
[4] Department of Artificial Intelligence, Inha University, Incheon, Republic of Korea
[5] Both authors contributed equally to this work.
[*] Authors to whom any correspondence should be addressed.

E-mail: bhyung@inha.ac.kr **and** shjo@kaist.ac.kr

Keywords: electroencephalogram, affective computing, emotion recognition, ear-eeg, wearable device

Abstract

Objective. In this paper, an around-ear EEG system is investigated as an alternative methodology to conventional scalp-EEG-based systems in classifying human affective states in the arousal-valence domain evoked in response to auditory stimuli. *Approach.* EEG recorded from around the ears is compared to EEG collected according to the international 10–20 system in terms of efficacy in an affective state classification task. A wearable device with eight dry EEG channels is designed for ear-EEG acquisition in this study. Twenty-one subjects participated in an experiment consisting of six sessions over three days using both ear and scalp-EEG acquisition methods. Experimental tasks consisted of listening to an auditory stimulus and self-reporting the elicited emotion in response to the said stimulus. Various features were used in tandem with asymmetry methods to evaluate binary classification performances of arousal and valence states using ear-EEG signals in comparison to scalp-EEG. *Main results.* We achieve an average accuracy of 67.09% $\pm$ 6.14 for arousal and 66.61% $\pm$ 6.14 for valence after training a multi-layer extreme learning machine with ear-EEG signals in a subject-dependent context in comparison to scalp-EEG approach which achieves an average accuracy of 68.59% $\pm$ 6.26 for arousal and 67.10% $\pm$ 4.99 for valence. In a subject-independent context, the ear-EEG approach achieves 63.74% $\pm$ 3.84 for arousal and 64.32% $\pm$ 6.38 for valence while the scalp-EEG approach achieves 64.67% $\pm$ 6.91 for arousal and 64.86% $\pm$ 5.95 for valence. The best results show no significant differences between ear-EEG and scalp-EEG signals for classifications of affective states. *Significance.* To the best of our knowledge, this paper is the first work to explore the use of around-ear EEG signals in emotion monitoring. Our results demonstrate the potential use of around-ear EEG systems for the development of emotional monitoring setups that are more suitable for use in daily affective life log systems compared to conventional scalp-EEG setups.

1. Introduction

Affective state, or emotion, is one of the most challenging topics in cognitive neuroscience. While it is difficult to define emotion exactly, it is evident that emotions play an important role in humans, including personality [1], decision-making [2], and various social activities [3]. As such, affective computing has attracted much interest in enabling better interactions between humans and computer systems by involving emotions [4]. Emotion recognition is an essential part of creating an emotion-aware system [5].

Given the subjective nature of emotion, it is difficult to quantify emotional values for classification. Previous studies have employed two types of affective space models for this task: discrete and dimensional models [6]. Discrete emotional models divide emotions into several core emotions that form the basis of more complex emotions, such as 'joy', 'sadness', 'surprise', 'fear', 'anger' and 'disgust'. Dimensional

models partition the affective space with two or more fundamental dimensions. One of the most commonly used dimensional models is the valence-arousal circumplex model [7], where the affective states are defined with two measurements: valence and arousal, each forming an axis of the circumplex. Valence provides a measure of how pleasant, or unpleasant, and arousal how intense, or calming, the emotional experience is. The core emotions in discrete emotional models can be superimposed over the dimensional models; for instance, the feeling of 'anger' would be divided into low valence and high arousal, while 'sadness' would be defined by low valence and low arousal.

Emotion recognition can be performed with various types of input, such as facial images or physiological signals. Common signals used include heart rate variability (HRV) [8], electrodermal activity (EDA) [9, 10], respiration rate (RR) [11], skin temperature (SKT) [12, 13], and electromyography (EMG) [14]. Many existing works also suggest electroencephalography (EEG) as a reliable signal to recognize emotion [15]. It has been previously shown that EEG displays distinctive patterns of spectral power in different brain areas in response to specific emotional states. For example, Choppin [16] reported high power in the EEG alpha wave in the frontal area and high power in the EEG beta wave in the right parietal area for affective states with high valence. Ramirez *et al* [17] observed a positive correlation between the EEG beta band in the frontal lobe with arousal. Zheng *et al* [18] reported higher activities in the beta and gamma bands of the temporal area for activities with high valence, while higher activities in the delta band in parietal and occipital areas and higher activity in the gamma band in the prefrontal area were seen for emotions with low valence. Together with the developments of signal processing and deep learning techniques, it is possible to translate human affective state from the EEG data with high precision [19–21].

There are several ways to measure EEG signals effectively. Most commonly used systems involve electrodes positioned over the scalp based on the international 10-20 systems. Several works demonstrated that high-quality EEG signals measured from the scalp can effectively be used for emotion recognition [18, 22–25]. However, conventional scalp-based EEG acquisition methods are uncomfortable to use over a long period of time, which may lead to frustration and evoke negative emotions over the session. To solve this problem, wearable systems were proposed to aid the emotion recognition process in a daily life environment with a user-friendly EEG acquisition device [26, 27]. Another solution to this is using ear-EEG systems instead. Ear-EEG refers to an EEG acquisition method that acquires EEG centered around the user's ears [28] Li *et al* [29] and Athavipach *et al* [30] both demonstrated above-chance level results for binary classification of affective states using electrodes attached to memory foam or flexible earbuds as an in-ear-EEG acquisition sensor. However, no further comparisons have been made between ear-EEG and scalp-EEG devices to properly validate the effectiveness of ear-EEG for use in emotion recognition.

In this work, we present a custom-made wearable around-ear EEG device and demonstrate the efficacy of the said device in comparison to the conventional scalp-EEG device in emotion recognition tasks. We perform six sessions of affective state experiments based on the international affective digital sounds (IADS) [31] for twenty-one subjects and analyze the performance of different feature extraction and classification techniques. To the best of our knowledge, this is the first work to explore a wearable around-ear EEG system in affective state classification tasks.

2. Methods

2.1. Data acquisition

2.1.1. Ear-EEG

EEG signals are acquired from around the ear from eight channels positioned around both ears of the subject as shown in figure 1(a). The channels are labeled L1, L2, L3, L4 for electrodes around the left ear, and R1, R2, R3, R4 for the electrodes around the right ear. The reference (REF) and ground (GND) electrodes are positioned on the middle of the back of the user's neck to avoid imbalances in the neural activities from the left and right channels. Additionally, photoplethysmography (PPG) sensor is also included in the right headphone below the R1 sensor. However, PPG data is excluded from study due to the experimental design.

For the electrodes around the ear, flexible brush-type electrodes made from a conductive elastomer with Ag-AgCl coating (LAXTHA Inc) are used to ensure contact between the electrodes and the skin in presence of hair. For GND and REF, flat electrodes made of the same material are used instead. EEG signals are acquired from these electrodes using an OpenBCI Cyton biosensing board, with a sampling rate of 250 Hz. Throughout the experiment, the electrodes show and maintain impedance below 20k. The device is easily wearable without external help, and generally takes less than a minute to setup for experiment.

As shown in figure 1(b), our device is designed as a wearable headphone, with the biosensing board, a rechargeable 3000mAh lithium-ion battery, and other circuitry encased inside the ear cups. The frame of the headphone was manufactured using PLA filaments on a 3D printer. The device operates wirelessly through a BLE connection with a personal computer. This device is an extended version of the one presented in our previous study [32], which was validated for its

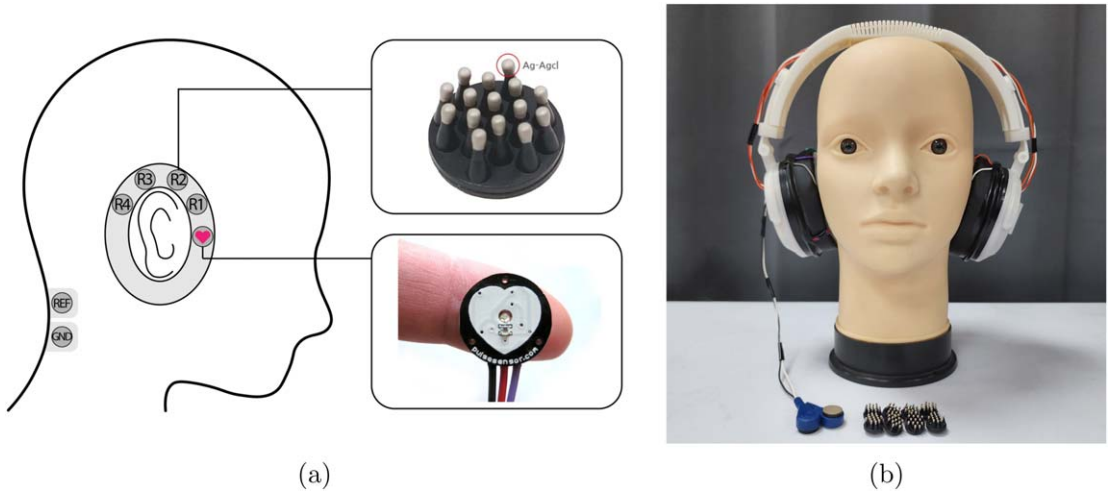

Figure 1. (a) Arrangements of electrodes placed around the ear, with the dry Ag-AgCl electrodes shown in detail. (b) Actual photograph of the wearable around-ear EEG acquisition device.

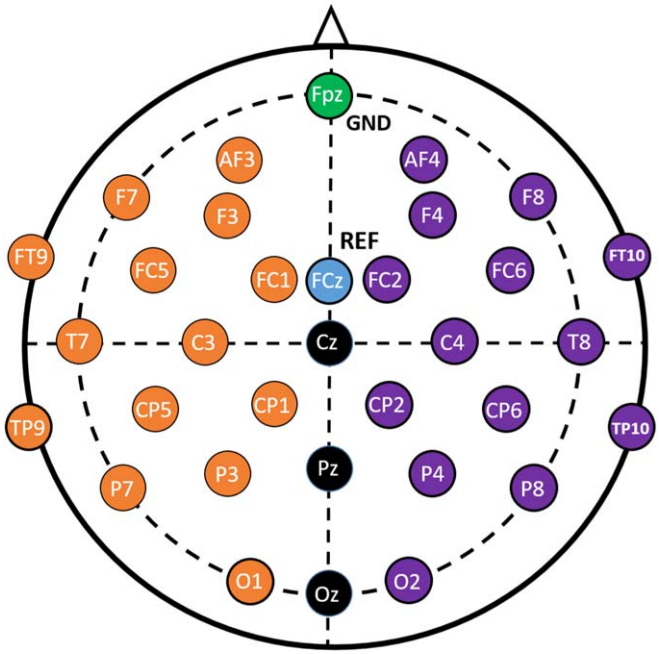

Figure 2. 32 electrodes chosen for the study according to the 10-20 international system. Fpz and FCz electrodes were used as the ground and reference channel respectively.

efficiency in a different BCI paradigm. Importantly, our device was developed at a production cost of less than 1500 USD, making it more cost-effective than some of the other commercial options available.

2.1.2. Scalp-EEG

EEG signals are acquired from the scalp using BrainVision actiCHamp with an EEG cap consisting of 32 Ag/AgCl electrodes placed around the scalp according to the 10–20 international system as shown in figure 2. The FCz channel is used as the ground channel, and the signal is referenced using the FPz channel. Electrolyte gel is inserted between the electrodes and the scalp to ensure that the impedance level of the electrodes remains below 10 kΩ during preparation time, which took around 30 minutes to finish. The signals are collected at a sampling rate of 500 Hz.

2.2. Experiment

2.2.1. Participants

In this study, twenty-one subjects, 20–29 years of age with no previously reported neurological or hearing impairments were recruited. Eighteen of the recruited participants were male. Seven participants had no prior experience with participating in EEG

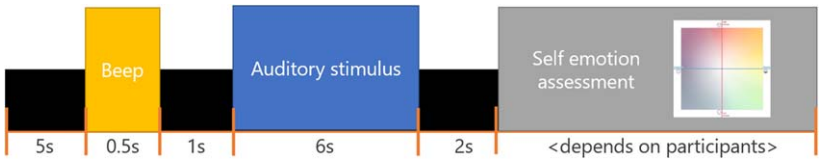

Figure 3. Timeline of a single affective self-reporting task in the experiment. Participants carried out fifty affective tasks per session, with a total of six sessions over three days.

experiments. All subjects gave written informed consent. The KAIST Institutional Review Board approved the experimental protocol of this study (KH2021-172).

2.2.2. International affective digital sounds (IADS

In this study, sounds from the IADS-E dataset [33] were used as stimuli to evoke affective responses. The original IADS is a standardized auditory stimulus database developed by the NIMH Center of Emotion and Attention at the University of Florida. The latest version of IADS, IADS-2, includes 167 digital sounds, each labelled with arousal, valence, and dominance values. IADS-E further builds upon this dataset to provide wider coverage of the affective space, containing a total of 935 sounds rated by 207 participants.

For each session of our experiment, fifty auditory stimuli were randomly sampled from the IADS-E database. The provided valence and arousal labels were considered during sampling to maintain a balanced ratio of high and low affective values of the stimuli.

2.2.3. Experiment protocol

A total of six sessions of emotion recognition experiment were carried out over three consecutive days. On each day, the participants carried out one session using the scalp-EEG device, and another session using the ear-EEG device. The participants were divided into two groups based on a 2×2 Latin square, with one group performing the ear-EEG session first, and the other scalp-EEG session first. The order of the sessions was kept identical for all three days.

Before the first session, each participant was thoroughly informed of the concept of the valence-arousal circumplex model of the affective state. Participants were initially instructed to perform a tutorial task, where they listened to four exemplary auditory stimuli from each quadrant of the circumplex. They were then given an opportunity to practice using the experiment interface; the auditory stimuli used here were excluded from the main experiment. The participants were given enough time to get used to the interface and the task before starting the experiment.

The participants performed fifty affective tasks per session. The experimental protocol is shown in detail in figure 3. In each task, the participants were initially shown a black screen for five seconds. A beep cue that lasted for 0.5 s was then delivered to prepare the participant for the task. One second after the beep, an auditory stimulus was played over six seconds. Two seconds after the auditory stimulus stopped playing, a popup window was displayed on the screen for the participants to report their affective state. They were given a circumplex chart with two axes: valence and arousal, each ranging from –9 to 9. They were instructed to mark a point that best corresponded to their affective state during the auditory stimulus. Once selected, they were given options to confirm, reselect, or skip the reporting process, in case they were unsure of their choice. Overall, one session lasted around 15 to 18 minutes. After six sessions, a total of 300 samples of EEG data and its corresponding self-reported affective state were acquired from every participant, half of which were collected using scalp-EEG and the other half ear-EEG.

2.3. Emotion recognition

2.3.1. Data processing

The scalp-EEG data were first downsampled to 250 Hz to match that of the ear-EEG data. A notch filter of 60 Hz was applied to all EEG data to remove power line noise. A fourth-order Butterworth bandpass filter with 1 to 60 Hz cutoff frequency was further applied to clean the data, after which the signal was segmented into epochs of six seconds from the onset of each auditory stimulus, each with their corresponding arousal and valence values reported by the participants. Samples that the participants skipped were removed, along with EEG epochs with unnaturally high amplitude exceeding $\pm 100\ \mu$V to remove contaminated data samples

2.3.2. Data resampling

We developed two binary classification models to classify data into high or low-value states for valence and arousal, respectively. Although we chose auditory stimuli in a balanced manner, some subjects' self-reported values were imbalanced, which could lead to biased models. To overcome this issue, we used the clustering-based undersampling algorithm [34] to balance the dataset. We generated N clusters for the majority class using the K-Means algorithm, where the smaller class contained N samples. We then selected the resulting N clusters as the new samples for that class. By grouping similar data samples into the same clusters and using cluster centers to represent the

majority class, this method reduces the risk of losing useful data from the majority class, especially when the data is not distributed equally among the classes. Previous studies [34] have shown that this approach outperforms other state-of-the-art data resampling techniques, such as the random undersampling (RUS) method [35] and synthetic minority oversampling (SMOTE) [36], in both small-scale and large-scale datasets. The resampling step was only applied to the training data and was performed separately for the arousal and valence values.

2.3.3. Feature extraction

Two different feature extraction methods commonly used in EEG emotion recognition were selected for comparison: power spectral density (PSD) and differential entropy (DE) to represent the band power and asymmetry for each EEG channel. PSD describes the power of a time series in the frequency domain computed using the Fourier transform [37]. In information theory, DE measures the randomness or complexity of a random variable; it differs from normal entropy in that the random variable can be continuous [38]. It has been shown in previous works that DE can be effectively used as a feature extraction method for the emotion recognition process [39]. Assuming that a time series X obeys the Gaussian distribution $N(\mu, \sigma^2)$, the DE is defined as:

$$h(X) = -\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} \log \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp\frac{(x-\mu)^2}{2\sigma^2} = \frac{1}{2} log\,(2\pi e\sigma^2) \quad (1)$$

Once the different features were acquired, two types of asymmetry features: differential asymmetry (DASM), and rational asymmetry (RASM), were computed for each of the band-power methods for the hemispheric-opposite channel pairs [39]. DASM and RASM are defined as:

$$DSLM_{LR} = B_L - B_R \quad (2)$$

$$RASM_{LR} = B_L / B_R \quad (3)$$

where B represents the band power, L is the EEG channel in the left hemisphere and R is the opposite EEG channel located in the right hemisphere.

The detailed feature extraction steps are as follows: First, PSD and DE of each channel (eight and thirty-one channels for ear-EEG and scalp-EEG, respectively) were calculated in multiple frequency bands ranging from the frequency of 1, 4, 8, 12, 16, . . . , 60 Hz to cover all basic frequency bands of EEG. This process was also performed with a different temporal window length and step size to include the temporal information in the features. Then, in addition to the extracted features, DASM and RASM were further acquired, using four channel pairs in the ear-EEG and fourteen channel pairs in the scalp-EEG. In summary, six different feature extraction methods including two band-power methods: PSD and DE of all EEG channels, two methods that combine PSD of all channels with asymmetry features based on PSD: PSD-DASM and PSD-RASM and two methods that combine DE of all channels with asymmetry features based on DE: DE-DASM and DE-RASM were computed and compared.

2.3.4. Feature selection

Due to the large dimension of the feature vectors, the extra trees classifier was used as the feature selection method. This feature selection method ranks the feature's importance based on the Gini index of each feature obtained from the construction process of the extra trees [40]. The extra trees were constructed based on the training set only, with a number of features set to 256. If the size of the initial feature vector was less than 256, the feature selection method was not applied.

2.3.5. Classification algorithm

Our classification model utilized a multi-layer extreme learning machine (MLELM). MLELM is a type of extreme learning machine (ELM) that has been stacked to create a multi-layer architecture [41]. Unlike traditional deep neural networks, MLELM assigns random weights and biases to each layer, making training much faster. MLELM has been shown to have good generalizability and is effective at handling noisy and non-stationary data, such as EEG signals [42]. These characteristics make MLELM a suitable choice for wearable EEG systems that aim for daily life applications. In previous studies, MLELM outperformed conventional models such as ShallowNet in EEG classification with a similar sample size to the current study [43]. To determine the optimal number of hidden nodes in each layer, MLELMs with three hidden layers were employed, and a grid search method was used to test values ranging from 2 to 256. For a more comprehensive understanding of the MLELM algorithm and its implementation, please refer to [42].

Because of the imbalanced testing dataset, the hyper-parameters were optimized based on the geometric mean (G-mean) defined as:

$$Gmean = \sqrt{sensitivity \times specificity} \quad (4)$$

In this work, evaluations using both subject-dependent and subject-independent strategies were conducted to validate the effectiveness of the ear-EEG acquisition method in the emotion recognition experiment. For the subject-dependent strategy, we performed a 5-fold cross-validation for each participant. In the subject-independent strategy, we used the leave-one-subject-out (LOSO) validation approach, where the data samples from one participant were used as the testing data, and the remaining data from all other participants were used as the training data for the classification model.

The classification accuracies of the ear-EEG and scalp-EEG data in all feature extraction and evaluation strategies were compared. Furthermore, we computed confusion matrices to check the potential bias of our models. This comprehensive approach enabled us to thoroughly validate the effectiveness of the ear-EEG acquisition method in the emotion recognition experiment.

2.4. Visualization

To examine the differences in the neural activity during each affective state, low valence (LV), high valence (HV), low arousal (LA), and high arousal (HA), we computed the PSD values of all channels for different frequency bands and averaged them across all subjects. For each frequency band, the averaged PSD values of all channels from both low and high affective states were then normalized and scaled to values between 0 and 1. To better understand affective activations, we also visualized the time-frequency response (TFR). Time-frequency analysis was conducted using the Morlet wavelet transform [44] using different groups of channels to better analyze the effects of affective states at regions of interest. For ear-EEG, channels were grouped based on the side of the ear where the electrodes were placed, with L1, L2, L3, and L4 channels in one group, and R1, R2, R3, and R4 in the other. For scalp-EEG, channels were chosen based on the activities observed in the PSD plots. FT9, T7 and TP9 channels were chosen from the left hemisphere and FT10, T8 and TP10 were chosen from the right hemisphere. The TFR values of the channels on the right ear are subtracted from the values of the channels on the left ear for ear-EEG, and the TFR values of right hemisphere channels are subtracted from the values of the left hemisphere channels for the scalp-EEG to better display affective lateralizations.

3. Results

3.1. Classification results

3.1.1. Subject-Dependent classification results

Tables 1 and 2 show the subject-dependent classification results using ear-EEG and scalp-EEG respectively. Classifications were carried out using different feature extraction and asymmetry calculation methods, with different window and step sizes.

When using ear-EEG signals, different features performed best for arousal and valence classifications. For arousal, DE-DASM features extracted using a three-second window with a 0.5-second step size showed the best performance at $67.09 \pm 6.14\%$, while for valence, PSD features outperformed the other features with an average accuracy of $66.61 \pm 6.14\%$ when used with three-second window size and one-second step size.

With scalp-EEG, DE-based methods showed better performance for both arousal and valence classifications. DE-DASM features with a three-second window size and one-second step size resulted in the highest classification accuracy at $68.59 \pm 6.26\%$ for arousal. For valence classification, DE features using a three-second window size and 0.5-second step size were the best with an accuracy of $67.10 \pm 4.99\%$. Figure 4 shows the boxplots of best performing features for arousal and valence classification using subject-dependent strategy. Comparing the best features for ear-EEG (DE-DASM for arousal and PSD for valence), differences in performance between ear and scalp-EEG are not significant using a paired t-test at 0.01 significance level.

3.1.2. Subject-Independent classification results

Tables 3 and 4 show the subject-independent classification results using ear-EEG and scalp-EEG respectively. Similarly to subject-dependent classifications, different features and asymmetry methods were used and compared. Subject-independent classifications showed generally inferior performance in comparison to subject-dependent classifications.

In the case of ear-EEG signals, PSD features with no asymmetry calculations applied were the best for both arousal and valence classifications. For arousal, a window size of three seconds and a step size of one second showed the best performance at $63.74 \pm 3.84\%$; similarly, the best classification result for valence was acquired using PSD features with a window and step size of one second at $64.32 \pm 6.38\%$.

Different features performed well for scalp-EEG signals. When classifying arousal, DE-RASM features with a window length of six seconds outperformed other methods with an accuracy of $64.67 \pm 6.91\%$. For valence, PSD-RASM features with a window length of six seconds showed best performance with an accuracy of $64.86 \pm 5.95\%$. Figure 5 shows the boxplots of best performing features for arousal and valence classification using a subject-dependent strategy. Similar to the subject-dependent classification results, the results acquired using the best features for ear-EEG (PSD for both arousal and valence) are not significantly different from that acquired from using scalp-EEG at a confidence level of 0.01.

3.1.3. Confusion matrix

In addition to the classification accuracy, confusion matrices for different classification cases were also plotted. Figures 6(a) and (b) show the confusion matrices of both ear-EEG and scalp-EEG methods in the subject-dependent and subject-independent strategy classifications, respectively, where the row of the confusion matrices represents the true class and the column represents the predicted class. It should be noted that the classification accuracies computed from the confusion matrices are not the same as shown in tables 1 to table 4 due to the difference in the number of class distributions between participants. Results from the confusion matrices display the balance results

Table 1. The average classification results (%) of ear-EEG using subject-dependent strategy with different features and window lengths. Results are shown as a ±b, where a is the average classification accuracy and b the standard deviation. **Bold** numbers represent the best results of arousal and valence models.

Window length (s)	Step size (s)	DE-based methods						PSD-based methods						Average	
		DE		DE-DASM		DE-RASM		PSD		PSD-DASM		PSD-RASM			
		Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
6	—	65.13 ± 4.74	65.35 ± 5.52	65.21 ± 5.14	65.36 ± 5.31	66.26 ± 5.66	64.86 ± 5.66	64.77 ± 3.17	63.64 ± 4.96	65.77 ± 4.69	64.61 ± 5.52	64.34 ± 3.90	64.25 ± 3.89	65.25 ± 4.57	64.68 ± 5.11
3	1	66.04 ± 6.21	64.75 ± 4.73	66.20 ± 5.92	64.05 ± 4.07	66.66 ± 6.02	64.81 ± 4.22	64.48 ± 2.82	**66.61 ± 6.14**	66.85 ± 5.16	65.49 ± 6.12	63.72 ± 4.34	64.19 ± 5.40	65.66 ± 5.24	64.98 ± 5.15
	0.5	66.45 ± 4.73	65.22 ± 4.37	**67.09 ± 6.14**	64.26 ± 4.12	66.68 ± 5.80	64.06 ± 4.79	65.68 ± 3.20	65.34 ± 4.97	65.42 ± 5.27	65.83 ± 5.30	63.51 ± 4.04	64.44 ± 4.63	65.80 ± 5.01	64.86 ± 4.66
1	1	64.85 ± 4.53	65.61 ± 3.39	65.94 ± 4.90	64.90 ± 4.30	65.67 ± 5.24	66.20 ± 6.84	63.99 ± 3.93	65.58 ± 4.47	65.01 ± 4.20	64.49 ± 2.91	63.07 ± 3.66	63.53 ± 4.18	64.75 ± 4.46	65.05 ± 4.51
	0.5	65.88 ± 5.05	65.21 ± 4.74	65.09 ± 5.08	65.39 ± 5.03	64.72 ± 5.11	64.87 ± 4.91	65.37 ± 3.58	65.98 ± 5.01	66.10 ± 4.90	66.01 ± 6.50	62.97 ± 3.71	63.86 ± 3.07	65.02 ± 4.64	65.22 ± 4.93
Average		65.67 ± 5.03	65.23 ± 4.52	65.91 ± 5.40	64.79 ± 4.54	65.60 ± 5.52	64.96 ± 5.30	64.86 ± 3.35	65.43 ± 5.14	65.83 ± 4.80	65.29 ± 5.35	63.52 ± 3.89	64.05 ± 4.23		

Table 2. The average classification results (%) of scalp-EEG using subject-dependent strategy with different features and window lengths. Results are shown as a ± b, where a is the average classification accuracy and b the standard deviation. **Bold** numbers represent the best results of arousal and valence models.

Window length (s)	Step Size (s)	DE-based methods						PSD-based methods						Average	
		DE		DE-DASM		DE-RASM		PSD		PSD-DASM		PSD-RASM			
		Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
6	—	66.65 ± 5.62	66.67 ± 4.98	67.08 ± 5.57	66.27 ± 5.96	66.40 ± 5.90	66.08 ± 5.75	66.23 ± 5.67	66.45 ± 4.38	67.00 ± 6.17	66.23 ± 5.50	66.67 ± 5.38	65.02 ± 4.10	66.67 ± 5.61	66.12 ± 5.08
3	1	66.41 ± 5.90	66.12 ± 4.89	**68.59 ± 6.26**	66.51 ± 4.50	66.52 ± 6.98	6.58 ± 4.16	65.89 ± 6.74	65.50 ± 5.21	66.91 ± 6.21	66.82 ± 5.82	66.50 ± 6.53	66.36 ± 5.89	66.80 ± 6.37	66.19 ± 5.03
	0.5	67.94 ± 5.70	**67.10 ± 4.99**	66.44 ± 6.02	67.04 ± 5.07	66.84 ± 5.83	66.26 ± 5.66	66.63 ± 6.45	66.56 ± 5.28	67.40 ± 6.54	66.35 ± 4.38	66.90 ± 5.85	65.36 ± 4.95	67.02 ± 5.97	66.45 ± 5.00
1	1	66.17 ± 5.92	66.06 ± 5.28	66.86 ± 5.74	65.52 ± 4.94	67.31 ± 6.31	65.91 ± 4.36	66.89 ± 6.00	65.61 ± 3.20	67.32 ± 6.01	65.60 ± 4.96	66.96 ± 5.30	66.37 ± 4.12	66.92 ± 5.78	65.85 ± 4.45
	0.5	66.85 ± 6.51	65.51 ± 4.32	66.93 ± 5.57	66.10 ± 4.37	66.63 ± 6.60	65.73 ± 3.85	67.90 ± 6.63	66.17 ± 4.20	66.33 ± 6.15	66.01 ± 4.78	67.53 ± 5.80	66.92 ± 4.91	67.03 ± 6.12	66.07 ± 4.35
Average		66.80 ± 5.85	66.29 ± 4.84	67.18 ± 5.77	66.29 ± 4.93	66.74 ± 6.23	65.96 ± 4.73	66.71 ± 6.23	66.06 ± 4.45	66.99 ± 6.11	66.20 ± 5.03	66.91 ± 5.69	66.01 ± 4.80		

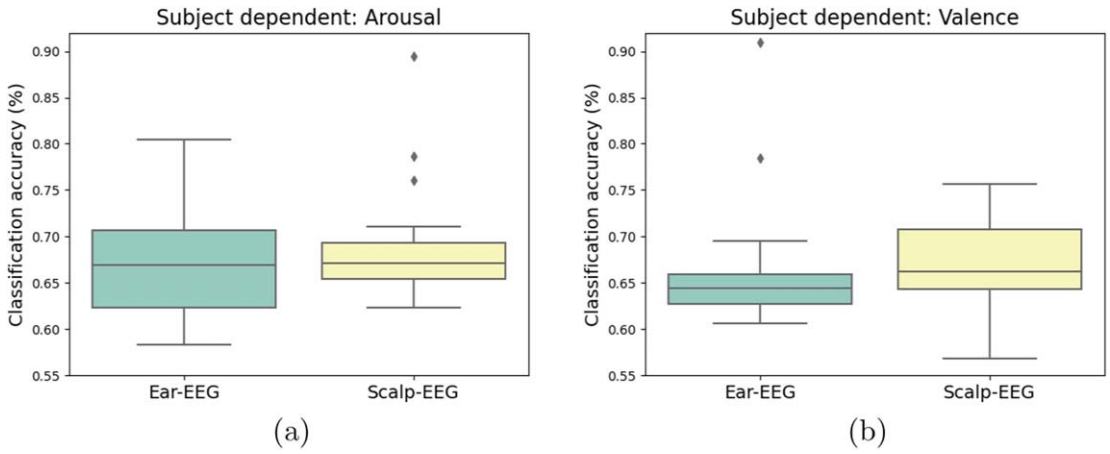

Figure 4. Box plot of best results using subject-dependent strategy for ear and scalp-EEG for (a) arousal and (b) valence classification. The box plot shows the quartiles of the dataset with the whiskers extended to show the rest of the distribution. Outliers are marked by a diamond.

in the true positive rate of all four classes in all evaluation strategies. This indicates that our classification model does not bias toward a specific class and using the geometric mean as the evaluation criteria for the model optimization process can solve the problem of the imbalances in the dataset where both classes are equally important.

3.2. Visualization of neural activities

3.2.1. EEG topographic map

The EEG topographic maps of the LV and HV affective states are compared in figure 7 and the comparison between the neural activity in the LA and HA affective states is shown in figure 8. In the ear-EEG cases, each colored circle represents the grand-averaged PSD value for each ear-EEG channel: from top to bottom, R1, R2, R3, and R4 channels in the right ear, and L1, L2, L3, and L4 in the left ear. For scalp-EEG, each black dot represents an EEG channel according to the conventional 10–20 international system.

As we can see in figure 7(a), the most remarkable difference between the LV and HV states of the ear-EEG data is that the LV state has a higher overall power as compared to the HV state even though they share similar patterns in the PSD values of all eight ear-EEG channels. The difference in the power between the two states is greater in the higher frequency bands including alpha, beta, and gamma while the power difference in the delta and theta bands is notable but insignificant. Comparing the power in the left and right ear-EEG, the average power in the left ear-EEG is higher than the right ear-EEG in all frequency bands in both LV and HV states, with the exception of the beta band, which shows very similar average power between the two sides of ear-EEG. Additionally, we could observe the left/right power lateralization between two affective states by comparing the average left/right power ratio of the ear-EEG data. The HV state shows a higher left/right power ratio in the delta, theta, and gamma bands and the LV state shows a higher left/right power ratio in the alpha band, while the value from the beta band is very similar in both states. Notably, the gamma band shows the highest difference in the left/right power ratio between the LV (approximately 1.4) and HV state (approximately 2.0) which suggests that the HV affective state is associated with the gamma power dominance in the left hemisphere.

Some interesting points can be observed from the scalp-EEG data (figure 7(b)). First, we can see a high beta and gamma power in the area around the T7 and T8 channels of both LV and HV states. These high-power activities likely represent the activities in the auditory cortex (located in the temporal lobe) which were occurred in response to the auditory stimuli. Both LV and HV states share a similar pattern of power among all channels in all frequency bands, with no significant power difference in the scalp-EEG channels positioned around the ear between the low and high valence states.

In the comparison between the LA and HA affective states from the ear-EEG results (figure 8(a)), we can see a higher overall alpha, beta, and gamma power in the LA state; however, the power patterns among all channels in the delta and theta bands are almost identical in both states. Unlike the valence states, the difference in the average left/right power lateralization in all frequency bands between the LA and HA states is very slight, with an exception of the L4 channel which has markedly higher gamma power in the LA state.

For the scalp-EEG data, again, the power concentrated around the area where the auditory cortex is located could be observed (figure 8(b)). Some differences in the power pattern can be seen between the low and high arousal states. First, a lower delta, theta, and alpha power in the parietal and occipital areas can be observed in the HA state. Second, the beta and gamma

Table 3. The average classification results (%) of ear-EEG using a subject-independent strategy with different features and window lengths. Results are shown as a ± b, where a is the average classification accuracy and b the standard deviation. **Bold** numbers represent the best results of arousal and valence models.

Window length (s)	Step Size (s)	DE-based methods						PSD-based methods						Average	
		DE		DE-DASM		DE-RASM		PSD		PSD-DASM		PSD-RASM			
		Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
6	—	60.60 ± 4.66	60.31 ± 5.38	60.81 ± 5.09	62.04 ± 5.55	60.79 ± 4.59	60.68 ± 3.52	61.83 ± 5.99	62.33 ± 4.84	61.07 ± 4.61	63.14 ± 4.72	62.29 ± 5.28	62.15 ± 5.77	61.23 ± 5.00	61.77 ± 5.01
3	1	61.46 ± 5.94	61.52 ± 3.55	62.42 ± 6.28	61.72 ± 5.16	62.88 ± 4.96	62.44 ± 4.79	**63.74 ± 3.84**	62.68 ± 5.28	62.39 ± 3.90	63.06 ± 5.50	63.35 ± 4.52	62.81 ± 5.72	62.70 ± 4.95	62.37 ± 4.98
	0.5	62.59 ± 5.02	61.47 ± 3.55	61.24 ± 4.20	62.64 ± 3.05	62.65 ± 4.68	62.18 ± 4.45	63.34 ± 5.43	62.52 ± 4.78	63.37 ± 6.21	62.22 ± 3.16	62.78 ± 5.44	61.59 ± 5.02	62.66 ± 5.15	62.10 ± 4.12
1	1	62.41 ± 5.94	61.79 ± 5.90	62.76 ± 5.87	61.72 ± 5.16	63.22 ± 3.91	62.86 ± 3.46	63.18 ± 4.08	**64.32 ± 6.38**	62.90 ± 3.80	63.07 ± 5.26	60.61 ± 3.10	63.97 ± 7.46	62.51 ± 4.57	62.95 ± 5.51
	0.5	61.64 ± 3.43	63.02 ± 5.10	60.56 ± 4.11	62.82 ± 4.49	63.35 ± 6.33	62.39 ± 3.72	63.00 ± 4.83	62.78 ± 4.97	62.61 ± 5.46	63.82 ± 6.21	62.90 ± 5.54	63.24 ± 6.26	62.34 ± 5.03	63.01 ± 5.12
Average		61.74 ± 5.03	61.62 ± 4.88	61.56 ± 5.16	62.19 ± 4.40	62.58 ± 4.95	62.11 ± 4.02	63.02 ± 4.85	62.93 ± 5.23	62.47 ± 4.85	63.06 ± 5.00	62.39 ± 4.86	62.75 ± 6.04		

Table 4. The average classification results (%) of scalp-EEG using subject-independent strategy with different features and window lengths. Results are shown as a ± b, where a is the average classification accuracy and b the standard deviation. **Bold** numbers represent the best results of arousal and valence models.

Window length (s)	Step Size (s)	DE-based methods						PSD-based methods						Average	
		DE		DE-DASM		DE-RASM		PSD		PSD-DASM		PSD-RASM			
		Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
6	—	64.57 ± 7.06	62.35 ± 3.91	64.13 ± 6.76	63.51 ± 3.87	**64.67 ± 6.91**	62.40 ± 3.74	63.42 ± 7.27	61.86 ± 2.27	62.89 ± 6.16	**64.86 ± 5.95**	63.89 ± 7.24	62.71 ± 2.85	63.93 ± 6.80	62.95 ± 3.98
3	1	62.63 ± 5.89	62.15 ± 4.39	63.89 ± 6.75	61.07 ± 3.48	63.00 ± 7.67	62.38 ± 2.86	64.38 ± 7.10	61.97 ± 4.51	63.47 ± 7.49	62.81 ± 3.79	63.49 ± 5.63	62.09 ± 3.92	63.48 ± 6.68	62.08 ± 3.82
	0.5	63.40 ± 8.23	62.95 ± 4.17	63.87 ± 6.75	62.16 ± 4.91	64.61 ± 7.53	62.01 ± 3.96	63.90 ± 8.16	62.82 ± 4.48	63.64 ± 6.57	61.79 ± 3.92	63.25 ± 7.91	61.24 ± 4.07	63.78 ± 7.42	62.16 ± 4.22
1	1	64.50 ± 6.23	62.60 ± 4.09	63.64 ± 6.97	63.19 ± 4.02	63.15 ± 6.80	62.60 ± 3.56	63.68 ± 6.43	62.09 ± 3.39	63.23 ± 6.47	62.52 ± 3.34	64.23 ± 6.25	62.02 ± 3.72	63.74 ± 6.42	62.50 ± 3.64
	0.5	63.17 ± 7.18	62.92 ± 5.28	62.41 ± 7.93	62.72 ± 4.42	63.16 ± 7.38	62.00 ± 4.74	62.91 ± 6.73	62.22 ± 4.30	63.79 ± 8.01	62.66 ± 4.21	63.16 ± 6.89	61.94 ± 4.42	63.10 ± 7.23	62.41 ± 4.50
Average		63.65 ± 6.87	62.60 ± 4.32	63.59 ± 6.94	62.53 ± 4.18	63.72 ± 7.17	62.28 ± 3.76	63.66 ± 7.04	62.19 ± 3.83	63.40 ± 6.85	62.93 ± 4.38	63.61 ± 6.71	61.70 ± 3.78		

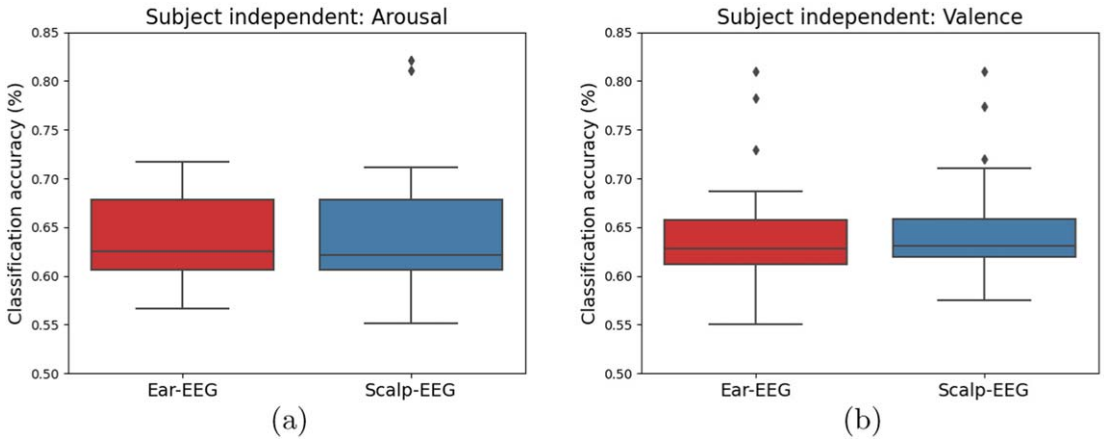

Figure 5. Box plot of best results using a subject-independent strategy for ear and scalp-EEG for (a) arousal and (b) valence classification. The box plot shows the quartiles of the dataset with the whiskers extended to show the rest of the distribution. Outliers are marked by a diamond.

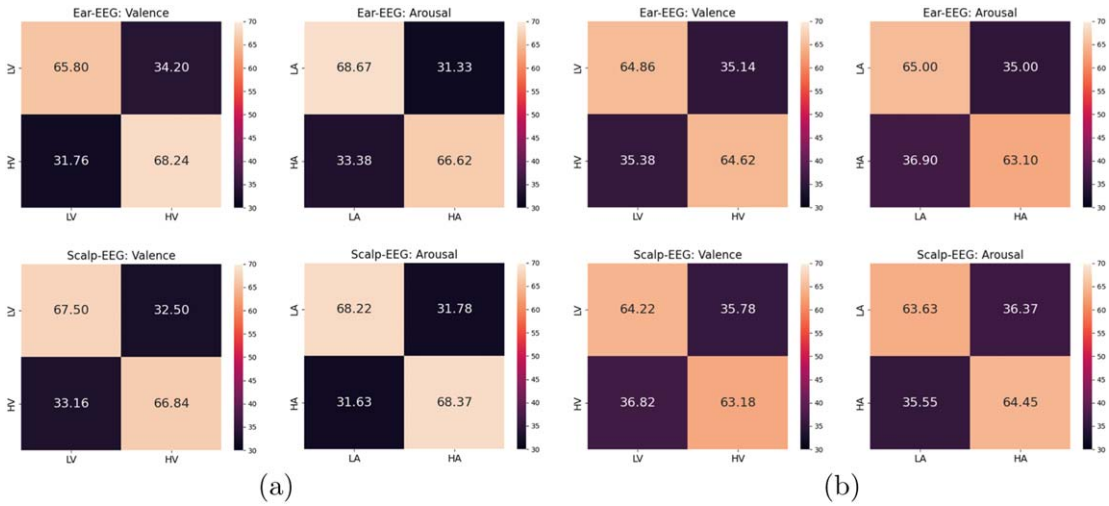

Figure 6. Confusion matrix of classification results using (a) subject-dependent and (b) subject-independent.

powers in the area around the T7 and T8 channels are higher in the HA state. Interestingly, unlike the results from the ear-EEG data, we could observe a clear difference in the beta and gamma left/right power lateralization in the auditory cortex area between the LA and HA states. In the beta band, the left power is slightly higher than the right power in the LA state and the right power is slightly higher than the left power in the HA state. Also, the left/right power lateralization in the gamma band is much greater in the LA state. This again supports the theory that EEG in the gamma band plays a very important part in the emotional process of the brain.

3.2.2. Time-frequency analysis

The time-frequency plots show more detailed insight into changes in spectral and temporal domains during the emotional process. With regards to the HV state in ear-EEG as shown in figure 9(a), while the right ear briefly shows more activities at the one-second mark from 10 to 50 Hz, the left ear shows predominantly more activities when compared to the right ear between 10 to 35 Hz overall. In frequencies above 40 Hz, both sides show similar activities, with the exception of three-second and five-second points, where activities in the left ear flare up. For the LV state, the right ear shows more activities in the frequency domain above 20 Hz after 0.5 s from the onset of the stimulus, with some higher activities in the left ear seen for frequency below 20 Hz sparingly.

Scalp-EEG TFR plots in figure 9(b) show slightly different responses. For the HV state, until 2 s after the onset of the stimulus, the right temporal channels show higher activities between 10 to 35 Hz, with higher activities in the left temporal channels for the frequency range above that. After two seconds, higher activities are seen in the left temporal channels between 10 to 35 Hz until the end of the task. At the three and five-second marks, left temporal channels again briefly show higher activities at higher frequency

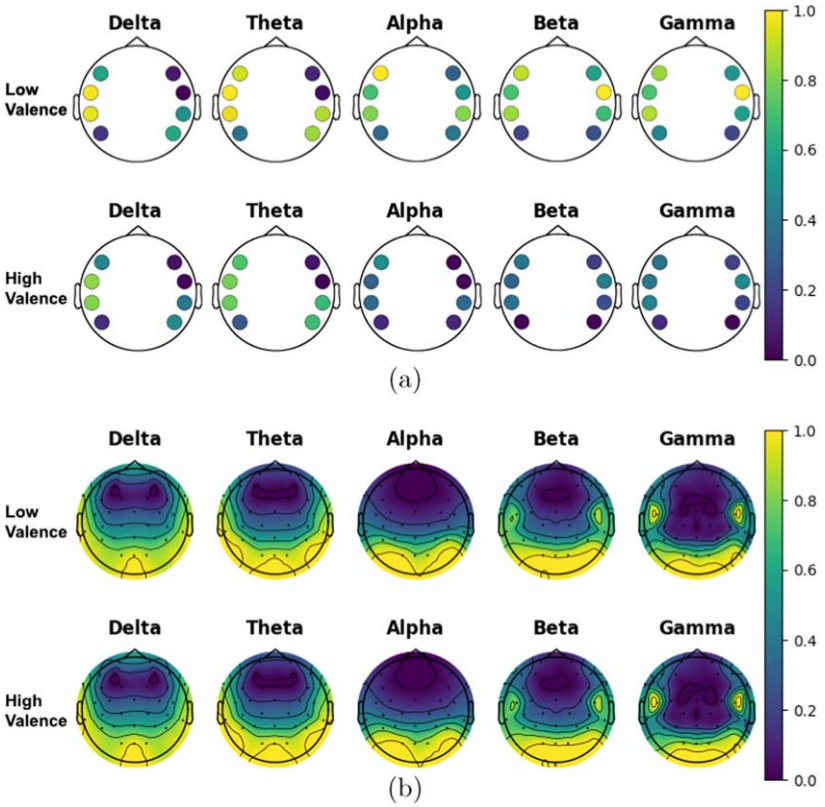

Figure 7. Topoplots of (a) ear and (b) scalp-EEG for low and high valence states.

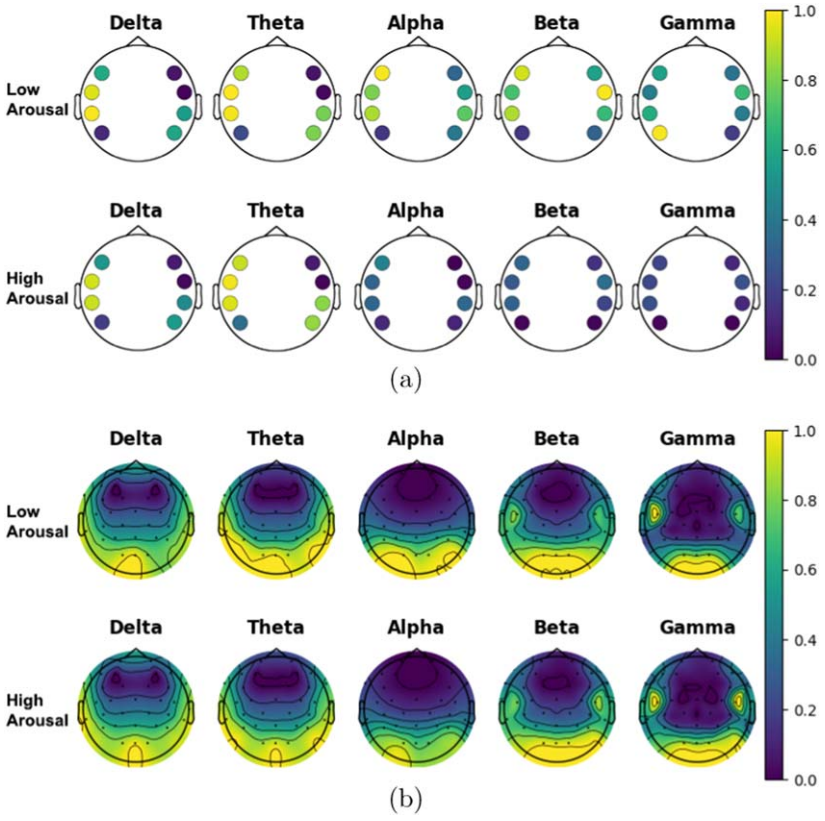

Figure 8. Topoplots of (a) ear and (b) scalp-EEG for low and high arousal states.

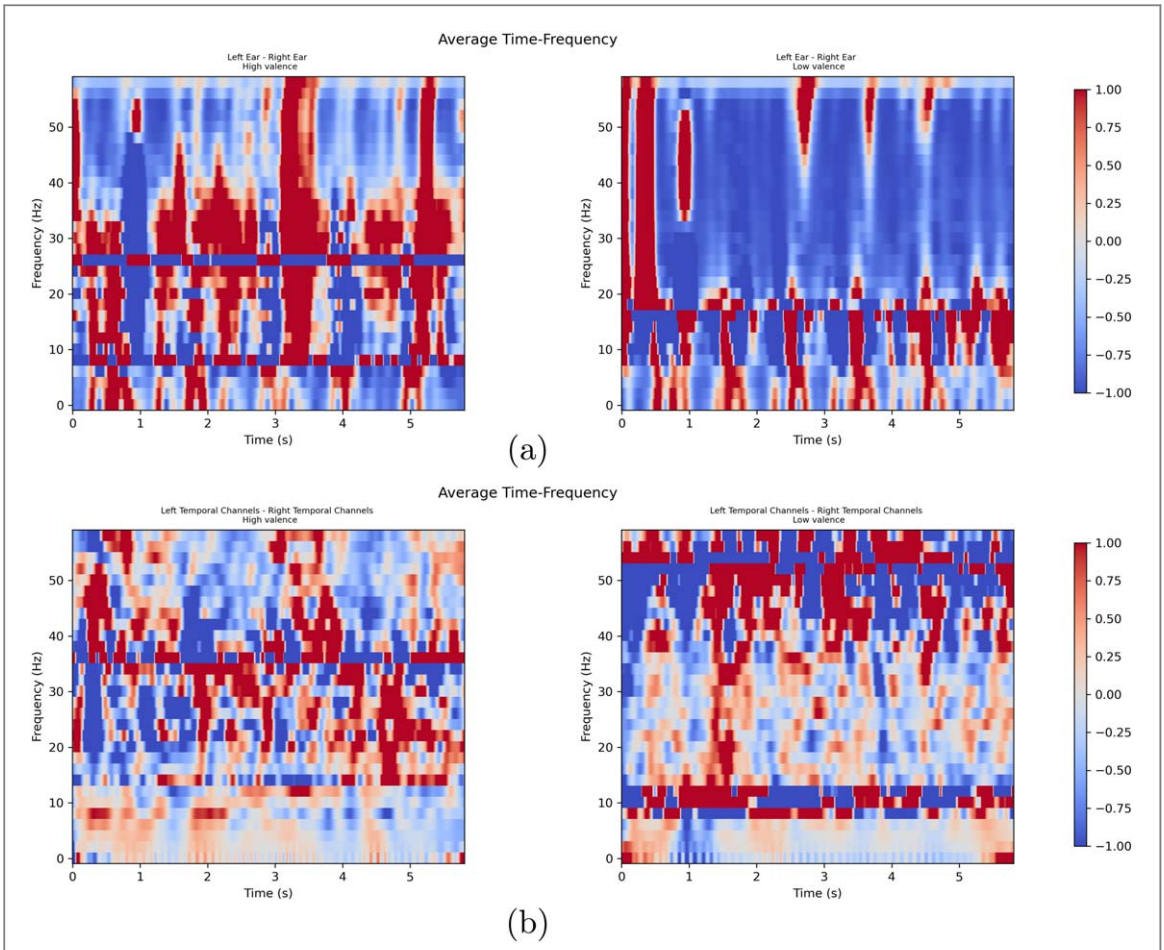

Figure 9. Time-Frequency Response of (a) ear and (b) scalp-EEG for low and high valence states. For ear-EEG, the difference between left ear and right ear channels are shown, while for scalp-EEG, the difference between left temporal channels (FT9, T7, TP9) and right temporal channels (FT10, T8, TP10) are shown.

ranges. For the LV state, both left and right temporal channels show similar levels of activities, with a brief spike in the left temporal channels at the 1.5-second mark. At frequencies above 40 Hz, right temporal channels show more activities until 1.5 s, after which the left temporal channels show comparatively stronger signals. After this, at three, four, and five-second marks, the right temporal channels show some spike of higher activities.

In figure 10(a), the TFR for ear-EEG during HA and LA states can be seen. While both HA and LA states show higher activations in the right ear at frequencies above 10 Hz, the LA TFR shows higher differences between the left and right ear. Both HA and LA show higher left ear activity one second after the onset of the stimulus, but at different frequency ranges; HA states between 30 to 55 Hz and LA between 15 to 50 Hz. LA TFR plot also shows more frequent spikes of higher activities in the left ear at frequency ranges higher than 35 Hz comparatively.

For the scalp-EEG (figure 10(b)), HA state TFR shows similar responses between left and right temporal channels at frequency ranges between 10 to 35 Hz, with the left temporal channels briefly showing higher activities 1.5 s after the start of the stimulus. At frequencies higher than 40 Hz, left temporal channels generally show higher levels of activities at all time periods with the exception of the 1–2 s range. In the LA state TFR, left and right temporal channels show similar values above 40Hz; below this, left temporal channels tend to show higher activities.

Overall, for both scalp and ear-EEG, TFR patterns generally stabilized two seconds after the start of the task. Such patterns may be due to the affective stimulus used; users may require some time for the auditory stimulus to play sufficiently before being cognizant of their affective response. The TFR plots suggest that at least two seconds of audio needs to be played before a proper affective response can be seen.

4. Discussion

4.1. Result discussion

4.1.1. Effectiveness of different features

DE-based features outperformed PSD-based features for subject-dependent classifications. PSD did perform the best for ear-EEG classification of valence at an accuracy of 66.61%, but DE-based features still showed comparable performance, with DE-RASM at 66.20%. For subject-independent classifications, PSD

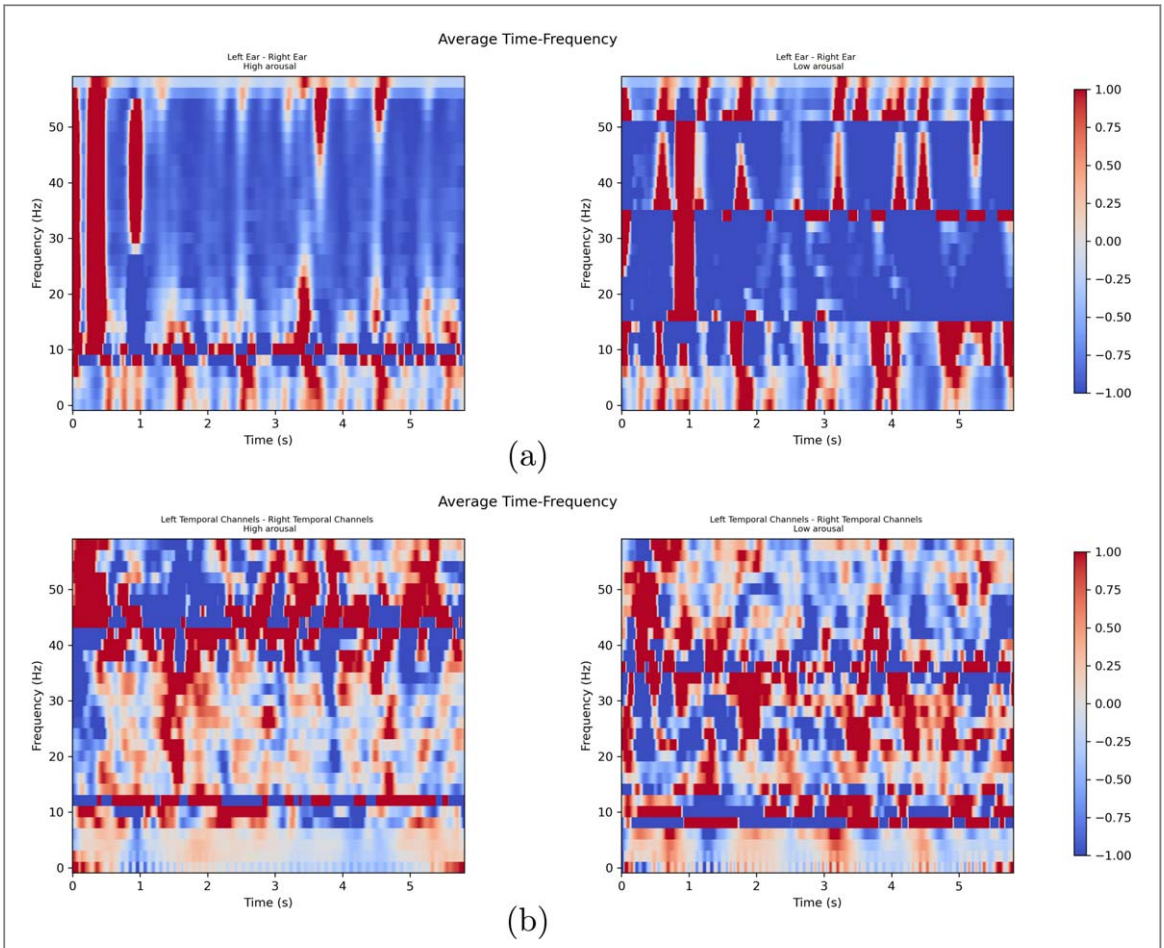

Figure 10. Time-Frequency Response of ear and scalp-EEG for low and high arousal states. For ear-EEG, the difference between left ear and right ear channels are shown, while for scalp-EEG, the difference between left temporal channels (FT9, T7, TP9) and right temporal channels (FT10, T8, TP10) are shown.

features showed higher accuracy, with the exception of arousal classification using scalp-EEG, where using DE-RASM features worked the best. Likewise, PSD features still showed comparable performance to DE-RASM features for this classification task. As for the time window used, the three-second window size showed the best performance overall for subject dependent strategy. This suggests that emotional response could be observed within three seconds. Coupled with observations from the TFR plot, emotional response may be mainly located between two to five seconds during the six seconds of the auditory stimulus. For subject-independent classifications using scalp-EEG, a six-second time window perform the best, while for ear-EEG, three and one-second windows show the best results for arousal and valence respectively. The differences in subject-independent classifications from subject-dependent classifications may result from individual differences, with different subjects having different delays in responses to auditory stimuli.

4.1.2. Around-ear EEG performance on affective decoding

In this study, our around-ear EEG method achieved a classification accuracy of 67.09% and 66.61% for the arousal and valence models, respectively, using a subject-dependent strategy. However, our method showed inferior performance than the in-ear EEG study in [30], which reported an accuracy of 72.89% and 71.07% for the arousal and valence models, respectively. One reason for their superior results may be that their experiments involved a longer stimulation period of 30 s (pictures from IAPS database), compared to our study, which used EEG while the stimuli were presented for only 6 s. Additionally, they used a 10-fold cross-validation method with forty samples obtained from a single session to validate their results, while our study performed cross-validation with samples mixed from three different sessions conducted on three different days. Considering the non-stationary nature of the EEG signal, this may have contributed to the lower performance in our study. For the subject-independent strategy, the study in [29] achieved a significantly higher classification accuracy of 94.1% for negative-vs-excited emotion binary classification, compared to our classification accuracy of approximately 64% using the same leave-one-subject-out validation method. One possible explanation for their exceptional accuracy, despite using a common feature extraction and classification method, is that they used video clips as visual-auditory stimuli

to induce emotion. The videos were four minutes in length and had a story, which might stimulate emotion better than the pure auditory stimuli used in our study. However, a limitation of the study in [29] is that each participant only performed one trial of the experiment, resulting in only one EEG sample per type of emotion. This limited the generalizability of their results. Nevertheless, our results do not imply that around-ear EEG methods are inherently inferior to in-ear EEG methods for affective BCI systems. While there are clear advantages and disadvantages to each method in terms of wearability and design, further studies on affective decoding using both ear-EEG methods, particularly ones that directly compare their performance, are necessary to fully evaluate their capabilities. One interesting observation in our results is the differences in performance when changing from subject-dependent to subject-independent classifications. While scalp-EEG showed slightly higher accuracies for both subject-dependent and independent classifications of arousal and valence in comparison to ear-EEG, ear-EEG showed significantly smaller changes in performance between subject-dependent and independent classifications, with some feature methods outperforming scalp-EEG. Scalp-EEG showed an average decrease of 3.28 in arousal and 3.72 in valence classification when changed to subject-independent strategy, in contrast to 3.01% and 2.51% decrease in ear-EEG performances. This may indicate that ear-EEG features are more robust and better suited for creating a subject-independent classification system. These results could be due to the smaller number of electrodes used in ear-EEG, resulting in the classification model not learning features more specific to individual participants. Scalp-EEG setups also had electrodes in regions where we did not see activities in response to affective states, which may have caused the classification model to learn features that are less relevant to the emotional processes of the human brain.

4.2. Neural activities during affective responses

In the topographic map of ear-EEG of valence states in figure 7a, we observed that the LV state has dominant overall power in the alpha, beta, and gamma bands, and that the left/right power lateralization, especially in the gamma band, might be used as a factor to discriminate the LV and HV state using the ear-EEG data. This corresponds with the valence model hypothesis of the brain lateralization of the emotional process which states that the left hemisphere is specialized for positive emotion and the right hemisphere is specialized for negative emotion [45]. This corresponds to the founding of the previous studies [18, 23], which state that the higher frequency band including the gamma band in the prefrontal, temporal, and parietal areas which are close to ear-EEG locations, highly correlates with the valence affective state. In contrast, for arousal states, our results, as shown in figure 8a, demonstrated smaller differences between the left and right channels in the lower frequency bands. These results suggest that the LA state is associated with a higher power strictly in the higher EEG frequency bands including alpha, beta, and gamma, and the left/right power lateralization might play a smaller part in differentiating the LA and HA state in contrast to the valence state. The results of TFR plots (figure 10) normally showed clear signs of emotional lateralizations at frequency ranges above 15Hz, with some affective states showing different responses above 40 Hz, also corresponding with the earlier observations made with topographic maps. Furthermore, our scalp-EEG topographic maps (figures 7(b) and 8(b) indicate an absence of activity in the frontal and central regions of the brain, which is inconsistent with previous findings. One possible explanation for this discrepancy is our use of Fpz as the common reference point in our scalp-EEG acquisition setup, which may have caused reduced activity in the areas surrounding Fpz, including the frontal and central regions, compared to other regions.

4.3. Future works

This work serves as a preliminary study that examines the performance of a wearable ear-EEG acquisition device in the emotion recognition process in comparison to the conventional EEG acquisition method. In this work, two binary classification models were used to recognize the user's emotional state from the data distinctively between the valence and arousal values, allowing the categorization of four emotional states, including LV-LA, LV-HA, LV-HA, and HV-HA, when combined. Developing the work further with more detailed labeling of valence and arousal values, such as the 'neutral' class, would provide further insight into emotional recognition using ear-EEG. Training a regression model based on continuous values of the affective states may also be an alternative to consider. We found dominant neural activities in the temporal areas of the brain during the stimulus observation period, suggesting that the emotional process of the human brain is dominant in the temporal areas. As discussed in the previous section, this founding corresponds to the results from the previous studies, however, there is also a possibility that the dominant neural activities in the temporal area that we discovered came purely from the auditory information processing of the auditory cortex which is also located in the temporal area as well. Further studies are required to validate the effectiveness of using ear-EEG in emotional recognition tasks, especially with different methods of eliciting affective responses. For example, visual stimuli such as the IAPS database might be used as an additional set of stimuli to compare the EEG in response to the stimuli to confirm the neural activity of the emotional process in the

temporal area in the absence of auditory stimuli, or the experimental protocol might be redesigned to lengthen the period after the offset of the stimulus in order to observe and compare the EEG data between the period where the stimulus is presenting and after the offset of the stimulus. Using video clips from films that provide a longer duration of visual and auditory stimuli with meaningful narratives could potentially elicit a more robust emotional response in the participants and thus provide a more rigorous test of the proposed method. Techniques, such as data augmentation [46], may also be applied to reduce or maintain similar experimental time while collecting sufficient samples. It would also be interesting to use and observe the ear-EEG in a self-induced emotion experiment that does not require any external stimuli in the experimental process as well. These studies will make sure that what we acquire from the EEG data is the neural activity of the emotional process and not purely from the cortical response to the given stimulus. Lastly, while our study offers a valuable comparison between conventional scalp-EEG and the proposed around-ear EEG approach, the classification performance of our method was comparatively subpar to state-of-the-art algorithms. Therefore, future investigations should explore more advanced machine learning techniques, such as those presented in [47] and [48], to improve the robustness and accuracy of our system. Despite this limitation, we maintain that our findings remain credible and significant, as they are consistent with the performance of other typical algorithms utilized in similar experimental setups [15, 49, 50]. Nevertheless, we acknowledge the potential for enhancement and will consider integrating these advanced techniques in our future research to enhance the overall performance of our system.

5. Conclusion

This work proposed ear-EEG as an alternative acquisition method to scalp-EEG for emotion recognition. The study conducted experiments to validate the effectiveness of ear-EEG acquired by the proposed wearable ear-EEG headphone by comparing its performance with the conventional 32-channel scalp-EEG. Using PSD and DE, and their brain asymmetry features as the feature extraction methods and MLELM as the classification model, the results show that ear-EEG produced results in classification accuracy comparable to the scalp-EEG method. Statistical tests were also performed to confirm that the differences between the results of these two EEG acquisition methods were not significant. Our results show that the DE-based feature extraction method gave the best classification accuracy compared to the other two methods, and incorporating the brain-asymmetry features can improve the results. Our results also suggest that the around-ear EEG may be better suited for subject-independent analysis when compared to scalp-EEG. In the data analysis, the topographic maps of both ear-EEG and scalp-EEG data show some distinct patterns in neural activity between each affective state and these distinctive characteristics are dominant in the Gamma frequency band, with time-frequency responses showing most neural activities in between two and five seconds after the start of the auditory stimulus. In conclusion, the results of this study indicate that around-ear EEG can be used as an alternative EEG acquisition method to the conventional scalp-EEG, advancing the development of wearable EEG systems for applications that require a long period of neural activity monitoring.

Acknowledgments

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant (2017-0-00432), and in part by INHA UNIVERSITY Research Grant.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Jaehoon Choi https://orcid.org/0000-0002-6074-711X
Netiwit Kaongoen https://orcid.org/0000-0002-5935-9662
HyoSeon Choi https://orcid.org/0000-0003-1517-8381
Byung Hyung Kim https://orcid.org/0000-0003-3374-0876
Sungho Jo https://orcid.org/0000-0002-7618-362X

References

[1] Reisenzein R, Hildebrandt A and Weber H 2020 Personality and emotion *The Cambridge handbook of personality psychology* (Cambridge: Cambridge University Press) pp 81–99
[2] Seymour B and Dolan R 2008 Emotion, decision making, and the amygdala *Neuron* **58** 662–71
[3] Livingstone A G, Spears R, Manstead A S, Bruder M and Shepherd L 2011 We feel, therefore we are: emotion as a basis for self-categorization and social action *Emotion* **11** 754
[4] Picard R W 2000 *Affective Computing* (Cambridge, MA: MIT Press) (https://doi.org/10.7551/mitpress/1140.001.0001)
[5] Cowie R *et al* 2001 Emotion recognition in human-computer interaction *IEEE Signal Process Mag.* **18** 32–80
[6] Harmon-Jones E, Harmon-Jones C and Summerell E 2017 On the importance of both dimensional and discrete models of emotion *Behavioral Sciences* **7** 66
[7] Russell J A 1980 A circumplex model of affect *Journal of Personality and Social Psychology* **39** 1161
[8] Zhu J, Ji L and Liu C 2019 Heart rate variability monitoring for emotion and disorders of emotion *Physiol. Meas.* **40** 064004

[9] Villarejo M V, Zapirain B G and Zorrilla A M 2012 A stress sensor based on galvanic skin response (gsr) controlled by zigbee *Sensors* **12** 6075–101
[10] Sequeira H, Hot P, Silvert L and Delplanque S 2009 Electrical autonomic correlates of emotion *International Journal of Psychophysiology* **71** 50–6
[11] Boiten F A, Frijda N H and Wientjes C J 1994 Emotions and respiratory patterns: review and critical analysis *International Journal of Psychophysiology* **17** 103–28
[12] Salazar-López E *et al* 2015 The mental and subjective skin: Emotion, empathy, feelings and thermography *Consciousness and Cognition* **34** 149–62
[13] Gouizi K, Bereksi Reguig F and Maaoui C 2011 Emotion recognition from physiological signals *J. Med. Eng. Technol.* **35** 300–7
[14] Dimberg U 1988 Facial electromyography and the experience of emotion *Journal of Psychophysiology* **2** 277–82
[15] Bos D O *et al* 2006 Eeg-based emotion recognition *The Influence of Visual and Auditory Stimuli* **56** 1–17
[16] Choppin A 2000 Eeg-based human interface for disabled individuals: emotion expression with neural networks *Unpublished Master's Thesis* Tokyo Institute of Technology
[17] Ramirez R, Palencia-Lefler M, Giraldo S and Vamvakousis Z 2015 Musical neurofeedback for treating depression in elderly people *Frontiers in Neuroscience* **9** 354
[18] Zheng W-L, Zhu J-Y and Lu B-L 2017 Identifying stable patterns over time for emotion recognition from eeg *IEEE Trans. Affective Comput.* **10** 417–29
[19] Craik A, He Y and Contreras-Vidal J L 2019 Deep learning for electroencephalogram (eeg) classification tasks: a review *J. Neural Eng.* **16** 031001
[20] Kim B H and Jo S 2020 Deep physiological affect network for the recognition of human emotions *IEEE Trans. Affective Comput.* **11** 230–43
[21] Suh Y-J and Kim B H 2021 Riemannian embedding banks for common spatial patterns with EEG-based SPD neural networks *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)* **35** 854–62
[22] Li M and Lu B-L 2009 Emotion classification based on gamma-band eeg *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE* pp 1223–6
[23] Yang K, Tong L, Shu J, Zhuang N, Yan B and Zeng Y 2020 High gamma band eeg closely related to emotion: evidence from functional network *Frontiers in Human Neuroscience* **14** 89
[24] Wu X, Zheng W-L, Li Z and Lu B-L 2022 Investigating eeg-based functional connectivity patterns for multimodal emotion recognition *J. Neural Eng.* **19** 016012
[25] Li G, Chen N and Jin J 2022 Semi-supervised eeg emotion recognition model based on enhanced graph fusion and gcn *J. Neural Eng.* **19** 026039
[26] Schmidt P, Reiss A, D¨urichen R and Van Laerhoven K 2019 Wearable-based affect recognitiona review *Sensors* **19** 4079
[27] Kim B H, Jo S and Choi S 2022 ALIS: learning affective causality behind daily activities from a wearable life-log system *IEEE Transactions on Cybernetics* **52** 13212–24
[28] Kaongoen N, Choi J and Jo S 2021 Speech-imagery-based brain-computer interface system using ear-eeg *J. Neural Eng.* **18** 016023
[29] Li G, Zhang Z and Wang G 2017 Emotion recognition based on low-cost in-ear eeg *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS). IEEE* pp 1–4
[30] Athavipach C, Pan-Ngum S and Israsena P 2019 A wearable in-ear eeg device for emotion monitoring *Sensors* **19** 4014
[31] Redondo J, Fraga I, Padrón I and Piñeiro A 2008 Affective ratings of sound stimuli *Behavior Research Methods* **40** 784–90
[32] Kaongoen N, Choi J and Jo S 2022 A novel online bci system using speech imagery and ear-eeg for home appliances control *Comput. Methods Programs Biomed.* **224** 107022
[33] Yang W *et al* 2018 Affective auditory stimulus database: an expanded version of the international affective digitized sounds (iads-e) *Behavior Research Methods* **50** 1415–29
[34] Lin W-C, Tsai C-F, Hu Y-H and Jhang J-S 2017 Clustering-based undersampling in class-imbalanced data *Inf. Sci.* **409** 17–26
[35] Liu B and Tsoumakas G 2020 Dealing with class imbalance in classifier chains via random undersampling *Knowl.-Based Syst.* **192** 105292
[36] Chawla N V, Bowyer K W, Hall L O and Kegelmeyer W P 2002 Smote: synthetic minority over-sampling technique *J. Artif. Intell. Res.* **16** 321–57
[37] Ameera A, Saidatul A and Ibrahim Z 2019 Analysis of eeg spectrum bands using power spectral density for pleasure and displeasure state *IOP Conference Series: Materials Science and Engineering* vol 557 (Bristol: IOP Publishing) 012030 (https://doi.org/10.1088/1757-899X/557/1/012030)
[38] Michalowicz J V, Nichols J M and Bucholtz F 2013 *Handbook of Differential Entropy* (Boca Raton, FL: CRC Press) (https://doi.org/10.1201/b15991)
[39] Duan R-N, Zhu J-Y and Lu B-L 2013 Differential entropy feature for eeg-based emotion classification *2013 VI International IEEE/EMBS Conference on Neural Engineering (NER). IEEE* pp 81–4
[40] Baby D *et al* 2021 Leukocyte classification based on feature selection using extra trees classifier: atransfer learning approach. Turkish *Journal of Electrical Engineering and Computer Sciences* **29** 2742–57
[41] Huang G-B, Zhu Q-Y and Siew C-K 2006 Extreme learning machine: theory and applications *Neurocomputing* **70** 489–501
[42] Ahmad M, Khan A M, Mazzara M and Distefano S 2019 Multi-layer extreme learning machine-based autoencoder for hyperspectral image classification *VISIGRAPP (4: VISAPP)* pp 75–82
[43] Kaongoen N, Choi J and Jo S 2021 Speech-imagery-based brain-computer interface system using ear-eeg *J. Neural Eng.* **18** 016023
[44] Büssow R 2007 An algorithm for the continuous morlet wavelet transform *Mech. Syst. Sig. Process.* **21** 2970–9
[45] Demaree H A, Everhart D E, Youngstrom E A and Harrison D W 2005 Brain lateralization of emotional processing: historical roots and a future incorporating dominance *Behavioral and Cognitive Neuroscience Reviews* **4** 3–20
[46] Luo Y, Zhu L-Z, Wan Z-Y and Lu B-L 2020 Data augmentation for enhancing eeg-based emotion recognition with deep generative models *J. Neural Eng.* **17** 056021
[47] Du X *et al* 2020 An efficient lstm network for emotion recognition from multichannel eeg signals *IEEE Trans. Affective Comput.* **13** 1528–40
[48] Zhang G, Yu M, Liu Y-J, Zhao G, Zhang D and Zheng W 2021 Sparsedgcnn: recognizing emotion from multichannel eeg signals *IEEE Trans. Affective Comput.* **14** 537–5
[49] Lan Z, Sourina O, Wang L and Liu Y 2016 Real-time eeg-based emotion monitoring using stable features *The Visual Computer* **32** 347–58
[50] Hettich D T, Bolinger E, Matuz T, Birbaumer N, Rosenstiel W and Spüler M 2016 Eeg responses to auditory stimuli for automatic affect recognition *Frontiers in Neuroscience* **10** 244