Real-time emotion recognition system with multiple physiological signals

Jyun-Rong ZHUANG*, Ya-Jing GUAN*, Hayato NAGAYOSHI*, Keiichi MURAMATSU**, Keiichi WATANUKI** and Eiichiro TANAKA*
*Graduate School of Information, Production and Systems, Waseda University
2-7 Hibikino, Wakamatsu-ku, Kita-Kyushu, Fukuoka 808-0135, Japan
E-mail: tanakae@waseda.jp
** Graduate School of Science and Engineering, Saitama University
255 Shimo-Okubo, Sakura-ku, Saitama City, Saitama 338-8570, Japan

Received: 11 April 2019; Revised: 18 September 2019; Accepted: 9 October 2019

Abstract
Emotion is an internal and subjective experience that plays a significant role in human life. There are several methods of recognizing emotions in people, the most authentic of which is using physiological signals, as they are beyond one’s control and strongly correlated with human emotions. This study aims to develop an emotion recognition system based on three physiological signals, namely, brainwave, heartbeat, and facial muscular activity. It utilizes deep neural network (DNN) and the T method of Mahalanobis-Taguchi system (MTS) to process the multiple physiological signals and further recognize the states of human emotion. As such, nine emotions are effectively recognized on a two-dimensional model through the DNN, then compared against several other algorithms, such as MTS, SVM, Naive Bayes, and K-means, where its superior accuracy is validated. Moreover, although the T method only improves the classification accuracy on the valence state, it rather obtains the intensity of emotion in different states. Furthermore, in this study, the proposed DNN is implemented into a wide range of applications for an accurate understanding of the human emotional states, whereas the T method is utilized to respond to the emotional intensity in different states. Finally, a real-time emotion recognition system is developed with DNN as the classifier; this system can directly monitor the variation of the human emotion through reliable and objective emotion analysis results from the physiological signals. Thus, the method can provide useful treatment effect information for robots or assistive apparatus serving activities of daily living.

Keywords: Emotional evaluation, Multiple physiological signals, Two-dimensional emotion model, Deep neural networks, Real-time emotion recognition system

1. Introduction

Emotion recognition is a key technology of human-computer interaction. Recent studies on recognizing emotions have extensively focused on facial expressions, speech, posture, and physiological signals (Adolphs et al., 1996), but since understanding and expression of emotions promote interpersonal communication, human–computer interaction methods that pay attention to the user’s emotional information are more suitable. Emotion recognition based on physiological signals has significantly different characteristics compared with that utilizing speech-and-image elements; thus, the former method becomes an important direction in the field of affective computing (Picard et al., 2001). Psychology and physiology provide evidences of strong correlation existing between physiological responses and human emotional states. For instance, the physiological signal is directly controlled by the autonomic nervous system and is not subjectively influenced by the subject; therefore, the recognition result becomes more realistic and objective (Andreassi, 2000). Through measuring electrocardiogram (ECG), the arousal emotion could be assessed by the ratio of sympathetic
value and parasympathetic value (Tanaka et al., 2017). Agrafioti et al. (2012) devoted to analyzing the ECG pattern to determine the arousal state for emotion recognition. In addition, facial electromyography (EMG) is a technique for monitoring the facial muscle activities. For emotion recognition, zygomatic and corrugator muscles’ activities are usually used to identify valence emotion, such as smile and frown. Former study had studied individual differences in emotion perception according to facial EMG signals (Künecke et al., 2014). Furthermore, electroencephalogram (EEG) analysis is one of the most effective approaches to identify emotions. An emotion recognition method based on EEG features were proposed (Mehmood et al., 2017). Lin et al. (2010) proposed a framework to process EEG-based emotion recognition. Studying EEG is valid for comprehending the human emotion. Generally, previous studies used a single physiological signal for judging emotion and ignored the fact that different emotional states may trigger multiple physiological signals, through which a fusion analysis may more accurately identify specific emotional states. Emotion recognition technique is useful to various fields. Especially, during the rehabilitation, it is necessary to understand patients’ mental state, which may provide medical staff to know their current condition further give mental assistance to patients. In our previous research (Tanaka et al., 2016a), we focused on developing an assistive walking device for hemiplegic patients and the elderly due to muscle weakness. The device can drive the joint in the ankle to trigger stretch reflex mechanism and further raise the user’s leg via a muscle linkage effect to prevent them from stumbling. Nevertheless, from the viewpoint of rehabilitation, the patients’ physical and mental aspects should be considered. Rehabilitation is a long-term and difficult work, and the patients are required to maintain their mental states for completion of the tasks involved. In other words, the mental states of the users are taken into account prior to using the assistive device for rehabilitation (Tanaka et al., 2018). Moreover, we applied the beat sounds to investigate the emotional variation of human in walking (Tanaka et al., 2016b). Afterward, we used the clustering algorithm to analyze the physiological signals mapping on the two-dimensional (2D) emotion map (Zhang and Tanaka, 2017) and attempted to control the assistive walking device using the users’ heartbeat signals (Tanaka et al., 2017). To recognize emotion precisely, we introduced DNNs to classify the multiple physiological signals, along with an answered questionnaire, and then identified nine emotional states on a 2D emotion map (Zhuang et al., 2018).

Herein, we aim to develop such emotion recognition system that is based on the analysis of multiple physiological signals, extracted via emotional stimuli experiments, and apply the proposed classification algorithms to accurately realize the states of human emotion on an arousal–valence plane. Our final goal is to employ this emotion recognition system to real life for offering beneficial in various fields.

2. Related works

2.1 Emotion model

Figure 1 (Posner et al., 2005) describes a 2D emotion model using two indicators, valence and arousal, of emotional state measurement. Valence is represented by the horizontal axis, which is divided into two types of emotions: positive and negative. Positive emotions refer to feelings that are pleasant, i.e., happy and contented, whereas negative emotions include unpleasant feelings, i.e., sad, upset, etc. On the other hand, arousal is indicated by the vertical axis and reflects the intensity of the emotions.

![Fig. 1 A two-dimensional emotion model (Posner et al., 2005).]
2.2 Questionnaire Survey–SAM (Self-Assessment Manikin)

Imprecise understanding of words in conventional studies has often led to wrong description of emotions. Bradley and Lang (1994) attempted to solve this problem by designing a picture tool called Self-Assessment Manikin (SAM) to directly evaluate pleasure and arousal. SAM is mainly characterized by a nonverbal pictorial assessment, that is, it uses simple figures to describe the dimensions of the emotion model. Moreover, SAM representative expressions provide a range of dimensions from smiling to frowning (pleasure) and from excited to relaxed (arousal). In our experiment, we presented a level 1–7 scale for both dimensions, as shown in Fig. 2.

2.3 Physiological signals

Methods of emotion recognition based on physiological signals analyze the autonomic and central nervous systems. Recognition based on the autonomic nervous system identifies the corresponding emotional states through measurement of physiological signals, such as heart rate, skin impedance, electromyography, and respiration (McCraty et al., 2015). On the other hand, recognition based on the central nervous system analyzes signals generated by the brain under different emotional states (Jenke et al., 2014; Javaid et al., 2015; Ismail et al., 2016). Emotion is recognized through changes in the physiological signals using both methods and cannot be hidden; thus, we can get objective results. This paper considered the measurement of three physiological signals, namely, heartbeat, brainwave, and facial muscle activity.

3. Emotion elicitation experiment
3.1 Normative–affective stimuli

Samson et al. (2015) proposed a film library storing 199 film clips (51 positive, 39 negative, 59 mixed, and 50 neutral), which were selected from 300 clips. These film databases are available for eliciting positive, negative, and mixed emotional states. The emotions of 411 participants stimulated by watching the clips were recorded and analyzed. Based on the participants’ ratings, the researchers could give the films an accurate judgment of the emotion they elicit.

3.2 Experiment devices

Figure 3 shows the devices EEG (electroencephalogram), EMG (electromyography), and ECG (electrocardiogram) acquired using EMOTIV EPOC+, Personal EMG, and myBeat, respectively.

1. EEG: We used EMOTIV EPOC+ (Emotiv Inc., 2018) with 14 channel electrode caps to collect the brain signals. The 14 channels were AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. We used two references (CMS/DRL) at P3/P4. The sampling rates were configured at 128 Hz. Built-in digital 5th order sinc filter (at 50–60 Hz) was applied to EEG signals at a bandwidth of 0.16–43Hz.

2. EMG: We used Personal EMG (Oisaka Electronic Device Ltd., 2018) as our surface muscular activity acquisition device. The EMG and iEMG (integrated electromyography) were sampled at 3000 Hz using a 12-bit A/D converter. EMG sensors were placed on zygomatic and corrugator muscles for measurement of “smile” and “frown,” respectively.

3. ECG: We used myBeat (Union Tool, Co., 2018) to obtain the heartbeat signals. Sampling rates were 1000 Hz. From
a matched visualized software, we were able to obtain real-time records of people’s heart rate. We determined LF/HF (LF: low frequency (0.04–0.15 Hz)/HF: high frequency (0.15–0.4 Hz)) ratio to understand the balance of a human autonomic nervous activity.

Consideration of the practical application, it is necessary to prepare the simple wearable devices without losing the accuracy of recognition emotion. In this study, we used the portable brainwave detector, wearable heartbeat detector, and muscle activities detector to build the emotion recognition system. Muscle activities detector could be replaced to a wireless type of muscle activities detector in the future. Through consideration of the practical application and the effective physiological signal selection, the authors prepare these three physiological signals.

3.3 Emotion elicitation experiment protocol

Before the experiment, the authors carefully explained the experiment condition and attained the consent of each subject. We did not force the subjects to conduct this experiment. Twenty healthy students (16 males and 4 females; 21–27 years old) participated in the emotion elicitation experiments, presented in a rigorous arrangement to ensure the quality of the collected physiological signals. Figure 4 depicts the actual experimental scenario. Elicitation materials consist of 30 films selected from the film library. At the experiment onset, the MVC (maximum voluntary contraction) of each participant was measured for preliminary assessment of the ultimate ability of their muscle activity. The setup of the emotion elicitation protocol was as follows:

1. Rest period: The subjects were asked to calm themselves and relax for 1 min. The generated physiological signals were recorded as their baseline.
2. Emotion elicitation period: The subjects were asked to watch the emotional stimuli clips.
3. Self-assessments period: After watching each clip, the subjects were instructed to evaluate their emotion by completing a 30-s survey questionnaire.
4. Washout period: Before going to the next clip, the subjects watched a 30-s washout video, mainly of a landscape accompanied by light music, to eliminate the emotional influence of the previous clip.

Figure 5 describes the entire emotion elicitation process. Compared with pure visual or auditory stimulation, watching a video film can simultaneously stimulate people’s visual and auditory senses and, thus, can give them a stronger sense of substitution and better emotional evocation. Throughout the experiments, the subjects were instructed to wear three physiological detectors. In the experiment, the number of subjects exhibited gender imbalance. However, to prevent the difference caused by gender, we had eliminated the individual difference for each subject through processing the physiological baseline signal.

![Fig. 3 Three devices, EMOTIV, Personal EMG, and myBeat, used in the elicitation experiments.](image)

![Fig. 4 Actual scenario of the emotion elicitation experiment.](image)
4. Data processing

4.1 Definition of the 2D emotion model

Three levels of valence and arousal states were mapped as follows. For the valence scales: 1–3 (un-pleasure), 4 (neutral), and 5–7 (pleasure). For the arousal scales: 1–3 (un-excite), 4 (neutral), and 5–7 (excite). Figure 6 enumerates the obtained nine emotional states: happy, pleased, relaxed, excited, neutral, calm, stressed, sad, and depressed.

Table 1 Extracted physiological features.

| Physiological data | Raw features | Extracted features |
|--------------------|--------------|--------------------|
| Facial EMG         | iEMG         | 1. Maximum iEMG divided by MVC of the corrugator muscle (C.M.) |
|                    | MVC          | 2. Maximum iEMG of the corrugator muscle (C.M.) |
|                    |              | 3. Maximum iEMG divided by MVC of zygomatic muscle (Z.M.) |
|                    |              | 4. Maximum iEMG of zygomatic muscle (Z.M.) |
| ECG                | Heart rate   | 5. Maximum LF/HF |
|                    | LF/ HF       | 6. Maximum variance of LF/HF |
|                    |              | 7. Maximum heart rate |
|                    |              | 8. Minimum variance of LF/HF |
|                    |              | 9. Mean value of LF/HF |
|                    |              | 10. Mean variance value of LF/HF |
|                    |              | 11. Standardized (Std.) LF/HF |
|                    |              | 12. Standardized (Std.) heart rate |
| EEG                | Theta (θ) wave | 13. Mean theta (θ) wave’s power spectrum (P.S.) |
|                    | Alpha (α) wave | 14. Mean Alpha (α) wave’s power spectrum (P.S.) |
|                    | Beta (β) wave | 15. Mean low beta (β) wave’s power spectrum (P.S.) |
|                    | Gamma (γ) wave | 16. Mean high beta (β) wave’s power spectrum (P.S.) |
|                    |              | 17. Mean Gamma (γ) wave’s power spectrum (P.S.) |
4.2 Feature extraction

We employed the mean and maximum values of the generated physiological signals as the features represent the physiological level of the emotion data from the experiment, as we proved these parameters to be more sensitive than the other values in our previous experiences. All the features were calculated from the raw physiological data, which were collected while the subjects watched the specific film clips. Table 1 presents the 17 extracted features, processed by the raw data of three physiological signals (EMG, ECG, and EEG). Previous researches have shown that the levels of biological signals are significantly different in subjects, as each has a unique physiological baseline. Therefore, the physiological baseline signal measured at the initial recording period was utilized to eliminate the effect of individual differences on the experimental results. Thus, the 17 extracted features were calculated by dividing the raw physiological signals with the baseline signal to obtain sample data for individual differences.

4.3 Evaluate dataset

We collected the experimental data to determine the correlation between emotion and physiological signals, rather than to obtain the precise emotion elicited in response to a specific stimulus. Hence, it did not matter whether the response emotion did not match the expected emotion. In general, human emotions are dynamic, due to different environmental factors, including culture, nationality, and memory. The same stimulus may elicit different emotions in different people (or even to the same person) or situations. We compared the actual emotional states (based on the questionnaire results) with those from the database, as shown in Fig. 7. In the right side of the figure, the actual emotions were generally close to the selected emotion stimuli area; nevertheless, with respect to personal independence, the same emotions could be felt in different stimuli. For example, after the subjects watched the films eliciting excited and un-pleasure (solid red circle) emotions, some subjects had responses close to those on the database, whereas others gave different feedback. As the factual basis of this study, the actual emotional responses of the subjects were used. After the emotion elicitation experiment, all actual emotional responses would follow our emotion definition (in Section 4.1) to as emotion label. After that, we used the emotional label to correspond to physiological data (17 extracted features indicated in Section 4.2). From the mentioned steps, we created the 20 subjects’ physiological signal-emotion dataset.
Table 2 Selected physiological features on two emotional states.

| Rank attributes | Extracted features | USE | Rank attributes | Extracted features | USE |
|-----------------|--------------------|-----|-----------------|--------------------|-----|
| 0.3292          | 15. Mean low β P.S.| ○   | 0.3562          | 10. Mean variance value of LF/HF | ○   |
| 0.2452          | 14. Mean α P.S.    | ○   | 0.2629          | 6. Max. variance of LF/HF   | ○   |
| 0.2105          | 2. Max. iEMG of the C.M. | ○ | 0.2357          | 5. Max. value of LF/HF    | ○   |
| 0.2034          | 16. Mean high β P.S. | ○ | 0.2319          | 13. Mean θ P.S.        | ○   |
| 0.1696          | 13. Mean θ P.S.    | ○   | 0.1602          | 11. Std. LF/HF        | ○   |
| 0.1361          | 1. Max. iEMG/MVC of the C.M. | ○ | 0.1531          | 15. Mean low β P.S.   | ○   |
| 0.1315          | 4. Max. iEMG of the Z.M. | ○ | 0.1316          | 14. Mean α P.S.       | ○   |
| 0.1283          | 17. Mean γ P.S.    | ○   | 0.0923          | 16. Mean high β P.S.  | ○   |
| 0.0999          | 3. Max. iEMG/MVC of the Z.M. | ○ | 0.0651          | 7. Max. heart rate    | ○   |
| 0               | 12. Std. heart rate | ×   | 0.0614          | 17. Mean γ P.S.       | ○   |
| 0               | 11. Std. LF/HF     | ×   | 0.0554          | 9. Mean value of LF/HF | ○   |
| 0               | 10. Mean variance value of LF/HF | × | 0            | 2. Max. iEMG of the C.M. | ×   |
| 0               | 9. Mean value of LF/HF | × | 0            | 12. Std. heart rate   | ×   |
| 0               | 5. Max. value of LF/HF | × | 0            | 3. Max. iEMG/MVC of the Z.M. | ×   |
| 0               | 6. Max. variance of LF/HF | × | 0            | 4. Max. iEMG of the Z.M. | ×   |
| 0               | 7. Max. heart rate | × | 0            | 1. Max. iEMG/MVC of the C.M. | ×   |
| 0               | 8. Min. variance of LF/HF | × | 0            | 8. Min. variance of LF/HF | ×   |

5. Methodology

5.1 Feature selection

We selected a valid amount of physiological data from the experimental results and removed redundant information to prevent data noise from greatly reducing the accuracy of our data classification. In the process, we proposed feature selection to improve both the recognition result and the efficiency of the classifier and obtain only the valid features. In principle, feature selection extracts only the most effective of the features available as different types carry different signal information. For example, some features may express one’s emotions more clearly, whereas others may contain invalid information. After selection, we individually analyzed the valence and arousal emotions; thereby, we had two classifiers.

Moreover, we utilized the information gain ratio, a statistical method that employs entropy to evaluate the worth of an attribute by the measured gain ratio (Karegowda et al., 2010) for this study. Table 2 breaks down the selected physiological features of the two emotional states. We selected the valid features according to the rank attributes calculated via the gain ratio; hence, we were able to trim down the number of extracted features for both states. The extracted features for the valence axis went down from 17 to 9 and from 17 to 11 for the arousal axis.

5.2 Deep neural networks

After processing the physiological signal data, we applied the classification methods for emotion recognition. In particular, we used DNN, a digital simulation of the biological neural networks (LeCun et al., 2015). Corresponding to the organism’s nervous system, the artificial neural network is also composed of a large number of neurons as basic functional units, as in the many nerve cells of a biological nervous system (Sutskever et al., 2014), constituting neural modules with different functions. Each module performs its functions and cooperates to complete varied tasks.

DNN is a supervised learning method, described by hidden layers and a softmax classifier; each layer undergoes a layer-by-layer training initially before the whole neural network is trained. Furthermore, DNN training minimizes the difference between an output and an input. In this paper, we necessarily had to label the corresponding physiological data before learning. The emotional label is the emotional evaluation data for the stimuli tested during the experiment.
We specifically used the activation function RELU and gradient descent algorithm for our proposed DNN for processing of a highly nonlinear problem; moreover, we employed softmax layer for classifying the learned features of the deep network structure. The weight and bias parameters of the softmax layer are trained through the supervised learning methods. After the network has completed learning of the parameters in the softmax classifier, the algorithm must fine-tune them in the entire network simultaneously, which improves all weights for all layers through backpropagation techniques with supervised methods. Additionally, backpropagation processes are used to learn network weights and biases based on marked training examples, for minimizing classification errors.

Herein, by collecting all physiological data and questionnaire answers from 20 subjects, we built the training set to create the DNN models. We constructed two classifiers, one with 9 neuron nodes of valence emotional states and the other with 11 neuron nodes of arousal states. The three hidden layer neurons had 60, 30, and 20 nodes, respectively, giving three output classifications, as shown in Fig. 8 (a). Moreover, the number of neurons in the input layer was the same as the number of dimensions of the feature vector, and the number of neurons in the hidden layer was determined through repeated experimental training. If there are too many nodes in the hidden layer, then over-fitting will occur; otherwise, under-fitting is expected. The softmax classifier outputs the probability of categorizing the emotional states, where the final discriminant results are the most probable. Figure 8 (b) reveals the accuracy and loss function of the valence and arousal classifiers (learning rate: 0.01; number of epoch: 500). Results of the test sets indicate that our constructed DNN was able to fit the training sets and then achieve high accuracy with low loss.
5.3 T method of Mahalanobis–Taguchi system

Subsequently, we divided the emotions on both axes into three classes. However, based on the results, we could know only the classes of emotions, but not their intensity; as such, we used the T method of the Mahalanobis–Taguchi system (MTS) for prediction.

The T method technique was developed by Genichi Taguchi for overall estimation of calculated values based on the signal-to-noise (S/N) ratio (Taguchi, 2005). Compared with MTS, this method does not need to use either the Mahalanobis distance, or the Gram–Schmidt orthogonalization, to remove variables. In general, the T method follows a three-step procedure.

First, a unit space is chosen from a sample dataset (for an α number of sample datasets), including the output values. Thus, the unit space (x_{ij}; i = 1, 2, …n; j = 1,2, …k) is the dataset that presents the relation between the input and the output values (y_{ij}; i = 1, 2, …n; j = 1 and 2) (Taguchi, 2005; Kawada et al., 2015). The unit space scale is shown in Table 3, for n number of sample data. The input and values represented the extracted features and the questionnaire results, respectively.

Table 3 Unit space scale.

| Unit space | Feature items (input value) | Output value |
|------------|----------------------------|--------------|
|            | x_1 | x_2 | x_3 | ⋯ | x_k | Valence | Arousal |
| Sample 1   | x_{11} | x_{12} | x_{13} | ⋯ | x_{1k} | y_{11} | y_{12} |
| Sample 2   | x_{21} | x_{22} | x_{23} | ⋯ | x_{2k} | y_{21} | y_{22} |
| ⋮          | ⋮   | ⋮   | ⋮   | ⋯ | ⋮     | ⋮     | ⋮     |
| Sample n   | x_{n1} | x_{n2} | x_{n3} | ⋯ | x_{nk} | y_{n1} | y_{n2} |
| Average value | ⠅_1  | ⠅_2  | ⠅_3  | ⋯ | ⠅_k  | ⠅_1  | ⠅_2  |

Second, the signal space is constructed with the rest of the sample data after selection of the unit space. The data of the signal space would be normalized by subtracting its average values (X_{ij} = x_{ij} - \bar{x}_k; M_{ij} = y_{ij} - \bar{y}) (Taguchi, 2005; Kawada et al., 2015). Normalization of the signal space (input value: X_{ij}; i = 1, 2, …l; j = 1,2, …k; output value: M_{ij}; i = 1, 2, …l; j = 1 and 2) measures the data offset from a given unit space. The number of the signal space datasets is l = a-n. Table 4 illustrates a normalized signal space scale. Furthermore, the unit and signal spaces are used for construction and validation of the predictive model.

Table 4 Normalization of the signal space scale.

| Signal space | Feature items (input value) | Output value |
|--------------|----------------------------|--------------|
|              | x_1 | x_2 | x_3 | ⋯ | x_k | Valence | Arousal |
| Sample 1     | X_{11} | X_{12} | X_{13} | ⋯ | X_{1k} | M_{11} | M_{12} |
| Sample 2     | X_{21} | X_{22} | X_{23} | ⋯ | X_{2k} | M_{21} | M_{22} |
| ⋮            | ⋮   | ⋮   | ⋮   | ⋯ | ⋮     | ⋮     | ⋮     |
| Sample l     | X_{l1} | X_{l2} | X_{l3} | ⋯ | X_{lk} | M_{l1} | M_{l2} |
Third, sensitivity ($\beta$) (the sensitivity of the output concerning the input data) and S/N ratio ($\eta$) are calculated using Eqs. (1) and (2) (Taguchi, 2005; Kawada et al., 2015), respectively. These data give an overall estimation of the actual output value. Table 5 shows the sensitivity ($\beta_{j,(1,2)}$) and S/N ratio ($\eta_{j,(1,2)}$) of each $x_j$.

From the mentioned steps, the predictive model can be established, and the overall predicted output ($Y$), which is expressed in Eq. (7), can be obtained by a weighted integration with the corresponding S/N ratio.

$$
\beta_{j,(1,2)} = \frac{\sum_{i=1}^{l} M_{i,(1,2)} X_{ij}}{r} \quad j = 1,2,...k
$$

(1)

$$
\eta_{j,(1,2)} = \begin{cases} 
\frac{1}{r} \left( S_{\beta_{j,(1,2)}} - V_{\beta_{j,(1,2)}} \right) & (S_{\beta_{j,(1,2)}} > V_{\beta_{j,(1,2)}}) \\
0 & (S_{\beta_{j,(1,2)}} \leq V_{\beta_{j,(1,2)}})
\end{cases} \quad j = 1,2,...k
$$

(2)

where parameters $r$, $S_{\beta_{j,(1,2)}}$, $V_{\beta_{j,(1,2)}}$, and $S_{T_{j}}$ are calculated as follows:

$$
r = \sum_{i=1}^{l} M_{i}^{2}
$$

(3)

$$
S_{\beta_{j,(1,2)}} = \left( \frac{\sum_{i=1}^{l} M_{i,(1,2)} X_{ij}}{r} \right)^2 \quad j = 1,2,...k
$$

(4)

$$
V_{\beta_{j,(1,2)}} = \frac{S_{T_{j}} - S_{\beta_{j,(1,2)}}}{r} \quad j = 1,2,...k
$$

(5)

$$
S_{T_{j}} = \sum_{i=1}^{l} X_{ij}^2 \quad j = 1,2,...k
$$

(6)

$$
Y_{i,(1,2)} = \frac{\eta_{i,(1,2)} X_{i1} + \eta_{2,(1,2)} X_{i2} + \cdots + \eta_{k,(1,2)} X_{ik}}{\sum_{j=1}^{k} \eta_{j,(1,2)}} \quad i = 1,2,...l
$$

(7)

Table 5 Sensitivity and S/N ratio of each feature.

| Feature items (input value) | $x_1$ | $x_2$ | $x_3$ | $x_k$ |
|-----------------------------|------|------|------|------|
| Valence                    | $\beta_{11}$ | $\beta_{12}$ | $\beta_{21}$ | $\beta_{22}$ | $\beta_{31}$ | $\beta_{32}$ | $\cdots$ | $\beta_{k1}$ | $\beta_{k2}$ |
| Arousal                    | $\eta_{11}$ | $\eta_{12}$ | $\eta_{21}$ | $\eta_{22}$ | $\eta_{31}$ | $\eta_{32}$ | $\cdots$ | $\eta_{k1}$ | $\eta_{k2}$ |
6. Results and discussion

6.1. DNN results

As mentioned earlier, we normalized the experimental data of physiological signals for various emotional states. We selected a total of 460 emotional samples as the training set for the classifier and used another 50 samples as the test set. We can separately divide each valence and arousal state into three groups using our proposed DNN and achieve a classification accuracy of up to 79.2%, for discriminating the valence state (un-pleasure, neutral, and pleasure), and 81.1%, for discriminating the arousal states (un-excited, neutral, and excited). To validate this accuracy, we employed other algorithms (SVM, Naive Bayes, and K-means) and compared their accuracy with that achieved by DNN. For the comparison, we conducted several classification accuracy measurements and determined the average accuracy, as shown in Fig. 9. Results confirmed that our method could attain the highest accuracy following the same data processing procedure. As classification accuracy is sometimes influenced by the initial weight and bias of the data, we moderately adjusted these parameters in each layer of the whole neural networks. The layered structure of the neural network maps the sample in the original space into a new feature space through a layer-by-layer feature transformation, thus making the classification easier. Moreover, by feature selection, it can effectively improve the classification accuracy by extraction of the most important input features to the DNN while eliminating the worse effect from irrelevant features. We can dig a deep relationship between the features through the deep structure of DNN; therefore, our DNN classifier is capable of recognizing human emotions in the next step.
Through individually discriminating the valence and arousal states into three groups, we were able to obtain a combination of the nine emotional states. Figure 10 gives the recognition results of DNN. The graph on the left shows the results of combining the two classifiers on a 2D plane, where the nine emotional states are displayed; each color marker represents the data recognizable in the specific emotion class, such as “excited but un-pleasure.” Moreover, the cross symbols represent unrecognized data, which contain misclassified emotions. On the other hand, the graph on the right presents grids that correspond to the emotional states of the left-side graph and their respective classification accuracy. Here, 27/11 would mean that for 38 data, 27 are recognized and 11 are unrecognized, thereby giving an accuracy of $\frac{27}{38} \times 100\% = 71.1\%$. From the point-of-view of the database results, we believe that good emotion prediction is achieved when the user is stimulated by a specific film, which is depicted by our recognition results that were located at the region similar with that of the database. Moreover, we think that better emotion prediction is achieved when a specific film stimulates the user, which is depicted by our recognition results and the user response (questionnaire) being located in the same region. Nevertheless, based on our experiences with different people, our recognition method can achieve approximately 70% of accuracy, indicating its feasibility, and further, its ability in general, to predict real emotion.

6.2. T method results

Following the procedure in Section 5.3, we utilized the specified unit and signal spaces to individually predict the output on the two axes (valence and arousal states). Furthermore, based on the two distributions (positive and negative states on a single axis) of the predicted output, we could respectively classify the “unpleasant or pleasant” and “deactivation or activation” on the valence and arousal states.
With this method, selection of the unit space would create substantial influence on the results. At the first trial, we selected the datasets of score 4 as unit space for the two axes; however, the results yielded a figure of the two distributions overlapping. Thus, we could not use this dataset as unit space to classify the emotional state. Next, we applied the datasets of other scores. Finally, for the valence state, we found the datasets of score 7 as unit space giving the best classification results; the overall distribution results displayed that “happy” and “unhappy” can be discernibly classified on the valence state. Figure 11(a) shows the distribution of scores 1-6 against the overall distribution: Scores 1, 2, and 3 are locally distributed in the right side from 0 while scores 5 and 6 in the left side from 0. Hence, for the valence state, this situation illustrated overall distribution results having a one-directional tendency (from negative to positive emotion), which is consistent with the direction of the emotional score axis. Correspondingly, on the arousal state, the datasets of score 2 as unit space yielded better classification results (“un-excited” or “excited”) than the others. Figure 11(b) shows the individual distribution of scores 1 and 3-7 against the overall distribution: Scores 4 and 5 are locally distributed in the right side from 0 while scores 1 and 3 in the left side from 0. Although some scores showed irregular phenomena, the overall distribution results had a one-directional tendency.

Furthermore, these results indicate that we can observe the emotional intensity in different states through the individually distributed score. To verify the accuracy of the T method, we examined 30 test samples 10000 times. The resulting classification accuracy was 77% and 47% for the valence and arousal states, respectively. The classification accuracy of arousal states showed the low value. We thought that the score distribution mainly concentrated at score 3, 4, and 5, which may lead to difficult to classify the data by using the T method for judging unexcited and excited. We will add more subjects to participate in the emotion elicitation experiments to further acquire the more valid data, then increase the classification accuracy.

6.3 Discussion

We devoted our time on recognizing states of human emotion through classification algorithms. Two methods demonstrated advantages on this respect. Firstly, through DNN, we were able to accurately classify nine states of emotion on a 2D model. During validation, DNN outperformed the classification accuracy of conventional algorithms. In addition, it reduced the calculation time, which makes it very suitable for real-time emotion recognition. Secondly, although the different human emotions could be easily recognized, their intensity was needed for a more precise evaluation of the mental states; thus, through the T method, we were able to individually distribute the scores of each emotion and determine their intensity. Should assistive robots and other advanced technology be allowed to use the technique we employed in this study, it is inevitable that they could accurately predict or recognize the emotional states and, further, mental states of people.

6.4 Real-time emotion recognition system

Finally, we developed a real-time emotion recognition system using the proposed DNN as the classifier. The system is presumed to quickly process many physiological signals and accurately recognize emotion. Initially, we collected real-time physiological data from users. Next, the data was analyzed for feature creation, feature selection, and data classification via the Python Software. Before using the platform, we first had to establish the suitable DNN model. A trained DNN model can operate on the physiological signal features input to the emotion recognition platform to output the emotional states we need to identify. Finally, we mapped the classification results on our developed emotion recognition interface. During practical applications, we found that users were uncomfortable wearing the facial sensors, i.e., makeup factor in women. Thus, we tried applying the system without the sensors and observed good results. Figure 12 illustrates the respective flow diagram. As a whole, the recognition system is composed of data input/output, feature extraction, DNN, and emotion recognition modules.

To evaluate the efficiency of this system, we conducted the same emotion elicitation experiment as in Section 3. We asked subjects to watch four specific stimuli videos (on the 2D plane; for “un-pleasure and excited,” “pleasure and excited,” “un-pleasure and neutral,” and “pleasure and neutral” states) and then compared the output of the system with the actual responses (questionnaire results) of the subjects. Figure 13 presents the results. We observed that these recognition results have a tendency similar to the specific stimuli. However, the actual emotions of the subjects are the most important. We further compared the actual emotions of the subjects with recognition results. From the actual
response results, for the valence recognition, we found that the recognition results could achieve 83.3% recognition accuracy to actual response (here, for 12 results, 10 results are same as actual responses, thereby giving an accuracy of 10/12 × 100% = 83.3%). For arousal recognition, the recognition results could achieve 50% recognition accuracy to actual responses (here, for 12 results, 6 results are same as actual responses, thus, an accuracy of 6/12 × 100% = 50%). We believe that when the participants filled out the questionnaire, the answer of the arousal evaluation was too neutral (subjects usually answered score 3, 4 and 5), which may lead to system misjudgment, thus it leads to low accuracy. In the future, we will focus more on the analysis of neutral answers. We believe that increasing the number of subjects and thus improving the deep learning model will help improve accuracy. Although the recognition accuracy was not very high, we still can prove our system that exhibited a possibility to as the practical application in recognizing human emotion. In the current article, our system is applicable to younger people (21-27 years old) and has not yet extended to other age groups, such as the elderly. Due to the different emotional characteristics of different age group people, in the future, we will not only increase the number of subjects but invite different age group people to conduct more experiments.

![Flow diagram of the real-time emotion recognition system.](image)

Fig. 12 Flow diagram of the real-time emotion recognition system.

![Evaluation of the real-time emotion recognition system.](image)

Fig. 13 Evaluation of the real-time emotion recognition system.
7. Conclusion

In the research field, emotion recognition has always been met with complicated and comprehensive details. In this study, we proposed algorithms of recognizing human emotions using multiple physiological signal inputs, which may prove beneficial in real-life applications.

To resolve the problem of high feature dimension and large computational complexity in emotion recognition, we employed DNN and the T method. Our proposed DNN was trained by datasets of 20 people’s physiological signals with questionnaire results and came out to successfully identify nine states of human emotion. We validated its superior accuracy over conventional algorithms through a comparison of their respective classification results. Furthermore, to deeply comprehend the mental state of an individual under an elicited emotion, we employed the T method and assessed the intensity of such emotion. T method achieved a high accuracy rate on the valence state, although it had some difficulty on the arousal state. In our future studies, we will optimize the existing algorithms to address this issue. Hence, with these two techniques, we could successfully predict and further recognize the emotion of a certain individual. At a later part, we presented a real-time emotion recognition system for practical applications in different fields; the algorithm includes a fast and accurate identification of the user’s human emotion. We verified its efficiency by assessing human subjects through an experiment identifiable to an earlier test, which yielded affirmative results. We believe this system can offer advantages for human use; for instance, assistive apparatus or devices adopting it can recognize the mental state of rehabilitation patients, which further reflects the pacing of their progress. Our future work direction includes, aside from an already mentioned objective, developing techniques to improve the classification accuracy of our human recognition methods and system (incorporating an emotional intensity function to the system) and applying them into a practical device to improve the quality of human life.

References

Adolphs, R., Damasio, H., Tranel, D., Damasio, A. R., Cortical systems for the recognition of emotion in facial expressions, Journal of neuroscience, Vol. 16, Issue 23 (1996), pp. 7678-7687.
Agrafioti, F., Hatzinakos, D., Anderson, A. K., ECG pattern analysis for emotion detection. IEEE Transactions on Affective Computing, Vol. 3, Issue 1 (2012), pp. 102–115.
Andreassi, J. L., Psychophysiology: human behavior and physiological response, Lawrence Erlbaum Associates, 4th Edition (2000), New Jersey.
Bradley, M. M., Lang, P. J., Measuring emotion: the self-assessment manikin and the semantic differential, Journal of Behavior Therapy and Experimental Psychiatry, Vol. 25, Issue 1 (1994), pp. 49-59.
Emotiv Inc., EMOTIV EPOC+ 14 channel mobile EEG (online), available from <https://www.emotiv.com/>, (accessed on 31 August 2018).
Ismail, W. W., Hanif, M., Mohamed, S. B., Hamzah, N., Rizman, Z. I., Human emotion detection via brain waves study by using electroencephalogram (EEG), International Journal on Advanced Science, Engineering and Information Technology, Vol. 6, No. 6 (2016), pp. 1005-1011.
Javaid, M. M., Yousaf, M. A., Sheikh, Q. Z., Awais, M. M., Saleem, S., Khalid, M., Real-Time EEG-based human emotion recognition, International Conference on Neural Information Processing, (2015), pp. 182-190.
Jenke, R., Peer, A., Buss, M., Feature extraction and selection for emotion recognition from EEG, IEEE Transactions on Affective Computing, Vol. 53, No.3 (2014), pp. 327-339.
Picard, R. W., Vyzas, E., Healey, J., Toward machine emotional intelligence: analysis of affective physiological state,” IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 23, Issue 10 (2001), pp.1175-1191.
Karegowda, A. G., Manjunath, A. S., Jayaram, M.A., Comparative study of attribute selection using gain ratio and correlation based feature selection, International Journal of Information Technology and Knowledge Management, Vol. 2, No.2 (2010), pp. 271-277.
Kawada, H., Nagata, Y., An application of a generalized inverse regression estimator to Taguchi’s T-Method, Total Quality Science, Vol. 1, Issue 1 (2015), pp. 12-21.
Künecke, J., Hildebrandt, A., Recio, G., Sommer, W., Wilhelm, O., Facial EMG responses to emotional expressions are related to emotion perception ability, PLOS ONE, Vol. 9, Issue 1(2014), e84053.
LeCun, Y, Bengio, Y., Hinton, G., Deep learning. Nature, Vol. 521 (2015), pp. 436-444.
Lin, Y., Wang, C., Jung, T., Wu, T., Jeng, S., Duann, J., Chen, J., EEG-Based emotion recognition in music listening.
IEEE Transactions on Biomedical Engineering, Vol. 57, Issue 7 (2010), pp. 1798-1806.

McCraty, R., Shaffer, F., Heart rate variability: new perspectives on physiological mechanisms, Assessment of Self-Regulatory Capacity, and Health risk.” Glob Adv Health Med, Vol. 4, Issue 1 (2015), pp.46-61.

Mehmood, M. R., Du, R., Lee, H. J., Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors. IEEE Access, Vol. 5 (2017), pp. 14797-14806.

Oisaka Electronic Device Ltd., Feature and function (online), available from <http://www.oisaka.co.jp/feature_function.html>, (accessed on 31 August 2018).

Posner, J., Russell, J. A., Peterson, B. S., The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development and psychopathology development and psychopathology, Vol. 17, Issue 3 (2005), pp. 715-734.

Samson, A. C., Kreibig, S. D., Soderstrom, B., Wade, A. A., Gross, J. J., Eliciting positive, negative and mixed emotional states: a film library for affective scientists, Cognition and Emotion, Vol. 30, No. 5 (2015), pp. 827-856.

Sutskever, L., Vinyals, O., Le, Q., Sequence to sequence learning with neural networks, Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol. 2 (2014), pp.3104-3112.

Taguchi, G., Objective function and generic function (6), Journal of Quality Engineering Society, Vol. 13, Issue 3 (2005), pp. 5-10 (in Japanese).

Tanaka, E., Muramatsu, K., Watanuki, K., Saegusa, S., Yuge, L., Development of a walking assistance apparatus for gait training and promotion of exercise, 2016 IEEE International Conference on Robotics and Automation (ICRA 2016), (2016a), pp. 3711-3716.

Tanaka, E., Muramatsu, K., Osawa, Y., Saegusa, S., Yuge, L., Watanuki, K., A walking promotion method using the tuning of a beat sound based on a two-dimensional emotion map, Proceedings of the AHFE 2016 International Conference on Affective and Pleasurable Design, (2016b), Springer, pp. 519-525.

Tanaka, E., Muramatsu, K., Osawa, Y., Watanuki, K., Saegusa, S., Yuge, L., Development of a walking assistance apparatus for promotion of exercise, The 4th International Conference on Design Engineering and Science (ICDES2017), (2017), Paper No. 223.

Tanaka, E., Yuge, L., Walking assistance according to the emotion, Emotional Engineering Vol. 6: Understanding Motivation (2018), Springer, pp. 140-152.

Union Tool Co., Wireless heartbeat sensor, Type WHS-1/ RRD-1 (online), available from <http://www.uniontool-mybeat.com/SHOP/8600010.html>, (accessed on 31 August 2018).

Zhang, Z., Tanaka, E., Affective computing using clustering method for mapping human's emotion, 2017 IEEE International Conference on Advanced Intelligent Mechatronics, (2017), Munich, Germany, pp. 235-240.

Zhuang, J. R., Guan, Y. J., Nagayoshi, H., Yuge, L., Lee, H., Tanaka, E., Two-dimensional emotion evaluation with multiple physiological signals. Proceedings of the AHFE 2018 International Conference on Affective and Pleasurable Design, Vol. 774 (2018), Springer-Verlag, pp. 158-168.