EmotionSense: An Adaptive Emotion Recognition System Based on Wearable Smart Devices

ZHU WANG, ZHIWEN YU, BOBO ZHAO, and BIN GUO, Northwestern Polytechncial University
CHAO CHEN, Chongqing University
ZHIYONG YU, Fuzhou University

With the recent surge of smart wearable devices, it is possible to obtain the physiological and behavioral data of human beings in a more convenient and non-invasive manner. Based on such data, researchers have developed a variety of systems or applications to recognize and understand human behaviors, including both physical activities (e.g., gestures) and mental states (e.g., emotions). Specifically, it has been proved that different emotions can cause different changes in physiological parameters. However, other factors, such as activities, may also impact one's physiological parameters. To accurately recognize emotions, we need not only explore the physiological data but also the behavioral data. To this end, we propose an adaptive emotion recognition system by exploring a sensor-enriched wearable smart watch. First, an activity identification method is developed to distinguish different activity scenes (e.g., sitting, walking, and running) by using the accelerometer sensor. Based on the identified activity scenes, an adaptive emotion recognition method is proposed by leveraging multi-mode sensory data (including blood volume pulse, electrodermal activity, and skin temperature). Specifically, we extract fine-grained features to characterize different emotions. Finally, the adaptive user emotion recognition model is constructed and verified by experiments. An accuracy of 74.3% for 30 participants demonstrates that the proposed system can recognize human emotions effectively.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **Applied computing → Health care information systems**;

Additional Key Words and Phrases: Emotion recognition, activity identification, scene-adaptive, wearable devices, multi-mode signals

ACM Reference format:
Zhu Wang, Zhiwen Yu, Bobo Zhao, Bin Guo, Chao Chen, and Zhiyong Yu. 2020. EmotionSense: An Adaptive Emotion Recognition System Based on Wearable Smart Devices. *ACM Trans. Comput. Healthcare* 1, 4, Article 20 (September 2020), 17 pages.
https://doi.org/10.1145/3384394

This work was partially supported by the National Key R&D Program of China (no. 2016YFB1001401), the National Natural Science Foundation of China (nos. 61725205 and 617772428), the Innovative Talents Promotion Program of Shaanxi Province (no. 2018KJXX-011), and the Fundamental Research Funds for the Central Universities (no. 3102019AX10).
Authors' addresses: Z. Wang, Z. Yu, B. Zhao, and B. Guo, Northwestern Polytechnical University, West Youyi Road 127, Xi'an, China; emails: {wangzhu, zhiwenyu}@nwpu.edu.cn, zhaobobo@mail.nwpu.edu.cn, guob@nwpu.edu.cn; C. Chen, Chongqing University, Shazheng Road 174, Chongqing, China; email: cschaochen@cqu.edu.cn; Z. Yu, Fuzhou University, Wulongjiang North Road 2, Fuzhou, China; email: yuzhiyong@fzu.edu.cn.

2637-8051/2020/09-ART20 $15.00
https://doi.org/10.1145/3384394

1 INTRODUCTION

Emotions are psychophysiological experiences that affect every aspect of our daily life, which are triggered by unconscious or conscious perceptions to something, and often associated with mood, personality, temperament, disposition, and motivation [7]. Additionally, they are a series of processes directed toward specific internal or external objects or events, which result in changes of both behavioral and physiological states [6]. Particularly, emotional health is closely related to the quality of personal life, as well as the security and stability of public communities [3]. A bad emotional state is not only harmful to personal health but also likely to be an early sign or cause of some serious mental illness [4, 14]. Especially for some special groups (e.g., drivers, pilots, and enginemen), emotional health even have serious influence with regard to public security [23, 24]. Therefore, developing an effective emotion recognition system is very important.

Recently, numerous studies have focused on designing effective human emotion recognition systems, which can identify implicit features contained in human communications such as facial expression, gestures, or speech in different experimental setups [10, 17, 18]. However, the features of audio/visual emotion channels usually are not adequate to obtain emotion classification results, as humans can disguise their emotions with artifacts of social masking [5]. For example, people may have a "poker face" and may not express emotion changes via intuitive body languages when they are in the mood [18]. Similarly, using traditional physiological measurements, including electroencephalography (EEG), electromyography (EMG), respiration (RSP), and blood oxygen saturation for emotion classifications has some limitations [1, 8, 15, 20, 22]. First, the data acquisition equipment is medical level, which is very expensive and not suitable for daily use. Second, physiological patterns cannot be mapped into specific emotional states uniquely because emotions could be influenced by many other factors, such as time, context, space, and culture [7]. To this end, this work aims to deal with the preceding issues by developing an adaptive emotion sensing system based on a sensor-enriched wearable smart watch. On the one hand, the smart watch is able to capture physiological signals in a non-invasive manner and thus is suitable for real-life use. On the other hand, fine-grained features can be extracted from multi-mode physiological signals, which take full advantage of the information contained in emotion changes, and further improve the efficiency of emotion recognition.

However, effective recognition of human emotions remains a challenging problem for the following reasons:

- In this article, we attempt to recognize user emotions by exploring the physiological signals. However, the user's physiological parameters not only depend on emotions but also other factors such as activities. In other words, other factors may interfere with physiological signals, making it difficult to recognize emotions adaptively by using only physiological signals.
- As emotions affect several aspects of the user's physiological signs, such as heartbeat, blood pressure, and skin temperature, it is hard to identify emotions by exploring one single attribute or attributes from one single physiological sign. Thereby, if we only focus on features from one single aspect, it is hard to achieve high recognition accuracy.

To tackle the preceding challenges, we propose an adaptive emotion recognition method based on a wearable smart watch, which takes the user's real-time activity scene into consideration and leverages more fine-grained features to identify emotion patterns by systematically exploring the multi-mode physiological signals. In particular, the proposed system is suitable for real-life environments and does not need to limit the subject in certain specific places. The main contributions of this work are threefold:

- To address the issue that physiological parameters not only depend on emotions but also other factors such as activities, we put forward an adaptive emotion recognition framework, which first identifies the activity scene by using the accelerometer sensory data and then facilitates context-aware and real-time emotion analysis by exploring multi-mode physiological signals.

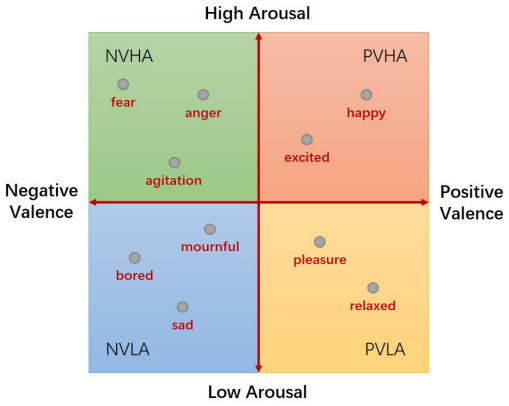


Fig. 1. The 2-D emotion model.

- We extract a set of fine-grained features from multi-mode physiological signals, which are closely related with emotion changes. Moreover, to validate the effectiveness of the extracted features, different emotion classification models are tested.
- To evaluate the proposed system, we construct an emotion stimuli corpus based on which different emotions can be evoked. We recruited 30 volunteers to collect a real physiological dataset together with the corresponding ground-truth emotion labels. Experimental results show that the system achieves an emotion recognition accuracy of 74.3%, which demonstrates the efficiency of the proposed system.

The rest of the article is organized as follows. We first review the related work in Section 2. Then, Section 3 presents an overview of the proposed system, followed by activity scene identification in Section 4. The adaptive emotion recognition method is presented and evaluated in Sections 5 and 6. Finally, we conclude in Section 7.

2 RELATED WORK

Lots of studies have been conducted in the field of effective computing, and great achievements have been made. In this section, we review these related works from three aspects.

The first line of research is emotion classification. Generally, it is difficult to judge or model human emotions because people usually express their emotions in different ways. Currently, there are two widely used emotion models. The first approach is discrete emotion models, in which we must choose a specific list of word labels to describe emotional states, such as joy, fear, sadness, anger, tension, and surprise. Examples of this kind of model include the six basic emotion model [19] and the tree structure emotion model [26]. However, one shortcoming of this approach is that the stimuli may elicit blended emotions that cannot be adequately expressed in words since the meanings of the chosen words are too restrictive or culturally dependent [7]. The other approach is dimensional models, which categorize emotions with multiple dimensions or scales instead of discreet words or labels. Arousal and valence are two commonly used scales for emotion classification [11], mapping all of the emotions onto a 2-D plane, as shown in Figure 1. The valence dimension represents the pleasantness of emotions (positive and negative). For example, joy and happiness are positive valence, whereas fear and sadness are negative valence [16]. The arousal dimension represents the level of emotions (i.e., low and high) [12]. For instance, sad and relaxed are low arousal, whereas fear and excited are high arousal. In this article, we use the 2-D model to describe human emotions with the labels LANV, LAPV, HANV, and HAPV. Today, psychophysiologists are still studying how to categorize human emotions accurately and specifically.

The second line of research focuses on emotion recognition based on physiological signals. Several studies have been conducted to recognize human emotion by utilizing physiological data, such as EEG, ECG, muscle

activity, skin conductivity, and RSP velocity [1, 5, 9, 18, 28]. For example, Koelstra et al. [9] investigated emotion recognition in terms of the levels of arousal, valence, like/dislike, dominance, and familiarity by exploring EEG data collected from 32 participants who had watched a 1-minute excerpt of a music video to evoke emotion. Although the work achieved an accuracy of 75%, EEG devices are neither too expensive nor inconvenient for daily use. Hsu et al. [5] adopted a musical induction method to evoke emotions of the participants and collected their ECG signals for emotion recognition. By extracting a set of ECG features, four types of emotions (i.e., joy, tension, sadness, and peacefulness) were classified with the LS-SVM algorithm. Setz et al. [21] explored galvanic skin response to distinguish stress and cognitive load. An arithmetic task was solved by 32 participants to elicit emotions, and the captured data was normalized based on the baseline period to address the issue of individual differences. Although an accuracy of 82% was achieved, it is difficult to ensure that the emotional states of different participants are comparable, as different participants may react differently to the designed tasks and the boundary between stress and workload is fuzzy. Even though all of the preceding studies achieved high accuracies, the performance of such systems heavily depends on the used emotion evoke method and the collected dataset. Our previous work [27] developed an emotion recognition system by exploring multi-mode physiological signals obtained with a smart watch. However, it had not considered the fact that other factors (e.g., activities) may also impact one's physiological parameters, which could severely influence the emotion recognition performance. Therefore, to address the issue that physiological parameters not only depend on emotions but also other factors such as activities, we aim to propose an activity scene-adaptive emotion recognition framework in this study.

The third line of research includes studies that try to recognize human emotions with audio/visual signals, such as facial expression, postures, and speech. Sarode and Bhatia [19] collected and explored facial expressions to recognize human emotions and proposed a 2-D appearance-based approach to extract intrinsic facial features. Radial symmetry transform and edge projection analysis were further used for feature extraction, and an accuracy of 81% was achieved for facial expression recognition from grayscale images. Wu et al. [26] adopted modulation spectral features (MSFs) to automatically recognize human affective information from speech. The authors used a modulation filter bank and an auditory filter bank for speech analysis and captured both temporal modulation frequency and acoustic frequency components. An overall recognition accuracy of 91% was achieved for classifying seven different emotions. However, as mentioned in the previous section, people may have a "poker face" or may not express their emotions via intrusive body languages. Therefore, compared with audio/visual signals, physiological signals are more reliable for the recognition of emotions.

3 SYSTEM OVERVIEW

As illustrated in Figure 2, the proposed system consists of three components: data collection, activity identification, and adaptive emotion recognition. We first collect physiological data from participants with emotion changes triggered with the emotion stimuli corpus, using the sensor-enriched Empatica E4 smart watch. After data preprocessing, human activities are identified by exploring the accelerator data, based on which we further build an adaptive emotion recognition model by leveraging three different physiological signals: the blood volume pulse (BVP) signal, the electrodermal activity (EDA) signal, and the skin temperature (SKT) signal.

Data capture. Human emotion is a complex progress, which consists of a series of reactivities. To facilitate the collection of emotion-related physiological data, we construct a stimuli corpus to evoke the participant's emotion changes [2]. At the same time, we ask all participants to report their emotional states by filling out a questionnaire, which can be used as the ground truth for emotion recognition. The detailed description of the experimental setups will be described in Section 6.

Activity identification. By adopting a sliding window–based segmentation method, we separate the accelerometer signal into a set of segments, and extract features for each of them from both time and frequency domains accordingly. Principal component analysis is used to optimize the extracted features, based on which an activity identification model is developed to identify different activities.

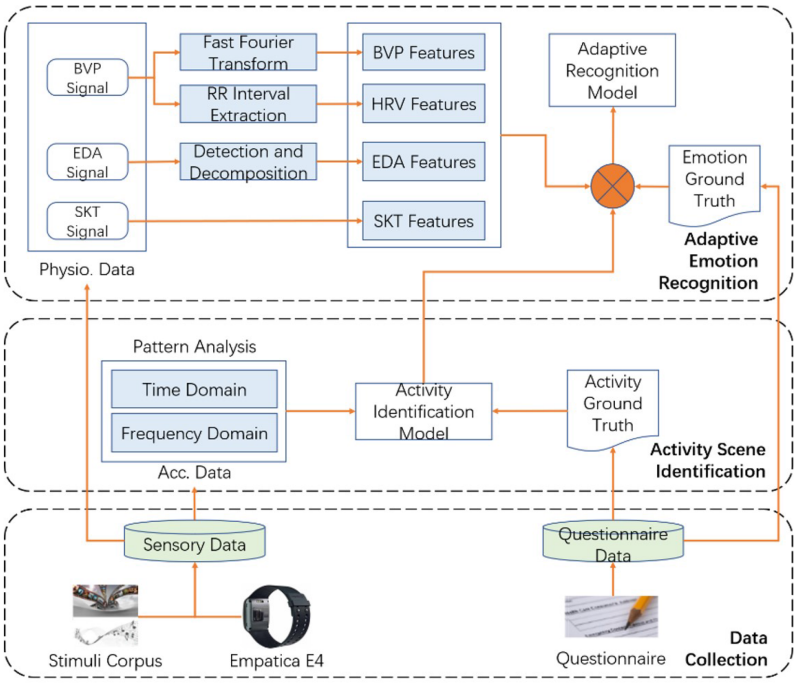


Fig. 2. Overview of the proposed adaptive emotion recognition system.

Adaptive emotion recognition. First, to eliminate the power frequency interference, we use the band-pass filter to remove noise from the BVP and SKT signals. Afterward, by exploring the change patterns of physiological signals, we extract emotion features from time, frequency, and non-linear domains. Finally, an adaptive emotion recognition model is constructed based on support vector machines (SVM).

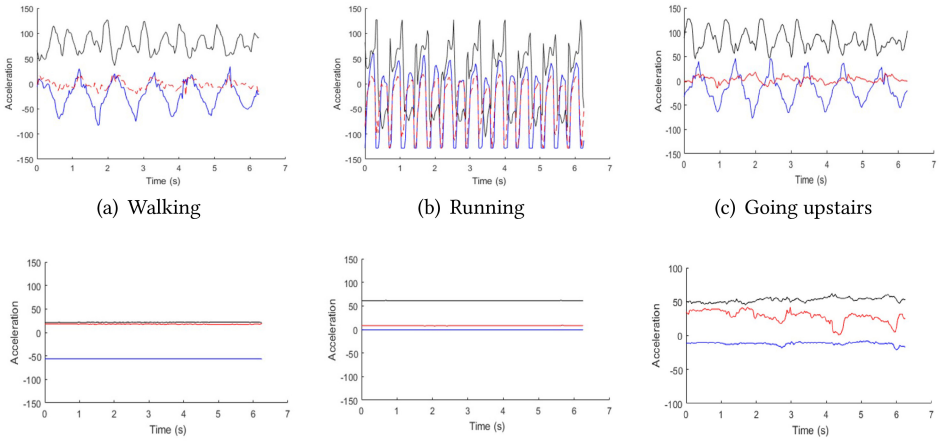
4 ACTIVITY IDENTIFICATION

In this work, we aim to detect user emotions by exploring sensor-enriched smart watches, which can collect the BVP signal, the EDA signal, and the SKT signal. However, as mentioned previously, other factors, such as activities, may also impact one's physiological parameters. Specifically, medical studies have found that changes in human physiological signals are normally affected by two factors: one is the individual's in vitro activities (e.g., sitting, walking, running, and riding a bicycle), whereas the other is the change of the human body itself caused by external stimuli (e.g., physical lesions or external emotional stimuli). For instance, if one is running or cycling, her heart rate will increase; if one is sitting quietly, her heart rate will slow down.

To address this issue, we first need to identify the user's current activity and then recognize her emotions accordingly. Specifically, our study is based on the Empatica E4 smart watch, as shown in Figure 2, which has a built-in three-axis accelerometer sensor (32 Hz) to collect the accelerometer signal, a photoplethysmography (PPG) sensor (64 Hz) to collect the BVP signal, a skin conductance sensor (4 Hz) to collect the EDA signal, and a skin temperature sensor (1 Hz) to collect the SKT signal. To identify user activities, we mainly use the accelerometer sensor, whereas the other sensors will be used for emotion recognition.

4.1 Accelerometer Data Analysis

Before feature extraction, we first analyze the collected three-axis accelerometer recordings. As the user moves, the x-axis of the accelerometer represents her horizontal acceleration, the y-axis represents her up-down acceleration, and the z-axis represents her forward acceleration, as shown in Figure 3.



(a) Walking (b) Running (c) Going upstairs

(d) Standing (e) Sitting (f) Riding bicycle

Fig. 3. Accelerometer recordings during different activities.

We can find that for most of the activities, the accelerometer recordings show periodic fluctuations. To leverage this characteristic for further analysis, a peak detection algorithm is applied to obtain the periodicity of the recordings. In the case of walking, as shown in Figure 3(a), all three-axis acceleration time series show periodic fluctuations, where the time interval of two adjacent peaks of the x-axis (red line) and the z-axis (blue line) is about 1 second, and that of the y-axis (black line) is about 0.5 second—for instance, the fluctuation frequency of up-down acceleration (y-axis) is twice as high that of forward acceleration (z-axis). Meanwhile, peaks of the x-axis and the z-axis are almost overlapping (the magnitude of the x-axis is much lower), and the time interval of such two adjacent peaks represents one step.

In the case of running, as shown in Figure 3(b), we can see that the y-axis and z-axis have similar fluctuation trends, and the time interval between two adjacent peaks is much shorter than that of walking (about 0.25 second)—for instance, the motion frequency of running is much higher than that of walking. Meanwhile, compared with walking, the y-axis acceleration in the negative direction during running is more obvious.

In the case of going upstairs, as shown in Figure 3(c), the magnitude of the y-axis is small and the time interval between adjacent peaks is larger (about 0.75 second), indicating that the motion frequency of going upstairs is lower than that of walking. Meanwhile, the negative direction of the z-axis indicates the user's tendency to go upstairs. The x-axis is a semi-regular periodic fluctuation, and the acceleration values swing between positive and negative.

In the case of sitting and standing, as shown in Figure 3(d) and (e), the recordings of the three-axis accelerometer do not show regular periodic fluctuations, and all acceleration values are relatively stable. Nevertheless, we can still observe obvious differences between the patterns of these two kinds of activities.

Finally, in the case of riding a bicycle, as shown in Figure 3(f), we can see that its fluctuation patterns are different from all of the preceding five activities. In particular, as arms usually keep relatively stable during riding, none of the three-axis acceleration time series show evident periodic fluctuations.

Based on the preceding analysis, we can find that the accelerometer recordings during different types of activities have distinct fluctuation patterns, and it is possible to distinguish different activities accordingly. For example, walking, running, and going upstairs can be distinguished based on the time interval between two adjacent peaks and the relative magnitude of acceleration. In particular, accelerometer recordings during most daily activities are quasi-periodic, fluctuating with the time. Thereby, the time domain characteristics of the

Table 1. Activity Identification Performance

	Walking	Running	Riding a Bicycle	Going Upstairs	Sitting	Standing	Accuracy
Walk	984	8	3	5	0	0	98.4%
Running	4	989	5	2	0	0	98.9%
Riding a Bicycle	1	4	994	1	0	0	99.4%
Going Upstairs	8	4	2	986	0	0	98.6%
Sitting	0	0	0	0	967	83	96.7%
Standing	0	0	0	0	54	976	97.6%
Overall Accuracy							98.27%

accelerometer data are useful for activity identification. Meanwhile, when performing different activities, a user's arms usually swing in different ways, resulting in distinct frequency domain characteristics.

In the next section, we will try to characterize such patterns by extracting both time domain and frequency domain features.

4.2 Accelerometer Feature Extraction

The accelerometer recordings during different activities have distinct statistical characteristics, and thus it is possible to identify them by exploring such statistical characteristics. In this article, *mean*, *variance*, *minimum*, *maximum*, and *standard deviation* of the acceleration time series are extracted as time domain features.

However, based on experimental results, in the frequency domain, we mainly use the power spectral density (PSD) to distinguish different activities, which indicates the energy distribution of the signal at different frequencies.

4.3 Activity Identification

To identify different activities, we adopt the AdaBoost algorithm with the BP network as the basic classifier. After parameter tuning, the results are summarized in Table 1, where the recognition accuracies for six different activities (i.e., walking, running, riding a bicycle, going upstairs, sitting, and standing) are 98.4%, 98.9%, 99.4%, 98.6%, 96.7%, and 97.6%, respectively, with an overall recognition accuracy of 98.27%.

Specifically, in the case of sitting, we can find that 967 of 1,000 sitting samples were classified correctly, whereas the other 93 samples were misclassified as standing. The classification result of standing was similar (i.e., a small number of standing samples were classified as sitting). It is worth mentioning that these two types of activities had not been misclassified as any other activities. In the case of walking, we found that it was most likely to be misjudged as running, which may be due to the commonalities of these two activities.

In general, satisfactory activity identification performance was obtained, which laid the foundation of adaptive emotion recognition.

5 ADAPTIVE EMOTION RECOGNITION

In general, people with the same emotions often show certain individual differences in terms of physiological signals, and the boundary between different emotions could be vague; therefore, it is necessary to design effective data processing and mining approaches to weaken the impact of individual differences on emotion recognition.

5.1 Physiological Signals

5.1.1 Photoplethysmography (PPG). There are photosensitive diode LEDs in the used Empatica E4 smart watch, which is able to capture the light reflected by the skin tissue and convert it into electrical signals. The electrical signals can be further converted into digital signals using an AD converter, resulting in the BVP signal

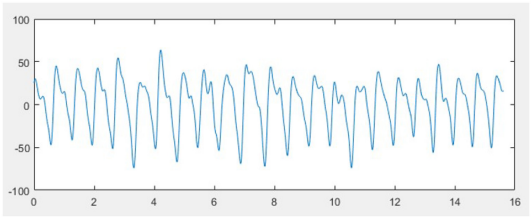
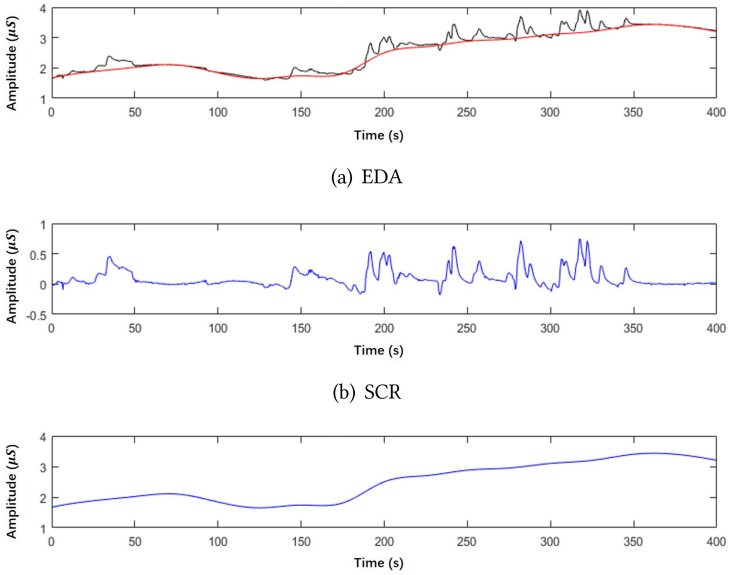


Fig. 4. The BVP signal.



(a) EDA

(b) SCR

(c) SCL

Fig. 5. The EDA signal.

used in this work. The BVP signal is a kind of oscillating movement of the blood and blood vessel wall caused by heart contraction. It forms at the root of the aorta and spreads rapidly to the peripheral blood vessel along the artery, which is the manifest fluctuation of various body parts. A slice of the BVP signal (16 seconds) is shown in Figure 4, where two peaks can be observed in each cycle, corresponding to the volume pulse wave and the pressure pulse wave, respectively.

5.1.2 Electrodermal Activity (EDA). A pair of electrodes are embedded in the Empatica E4 smart watch, which is able to obtain the skin conductivity signal by measuring the weak current in human skin. The skin electrical signal changes along with the sweat gland function, which is called the *skin electric response*. The skin electrical signal consists of two parts: skin conductance level (SCL) and skin conductance response (SCR). The SCL reflects the basic value of physiological activity in a quiet state, whereas the SCR is a series of transient and fast signal fluctuations due to physiological activation caused by stimulation. We illustrate the skin electrical signal in Figure 5.

Figure 5(a) is a typical skin electrical signal (the black line), from which we can extract the SCL signal (the red line in Figure 5(a), as well past the line in Figure 5(c)) and the SCR signal (as shown in Figure 5(b)).

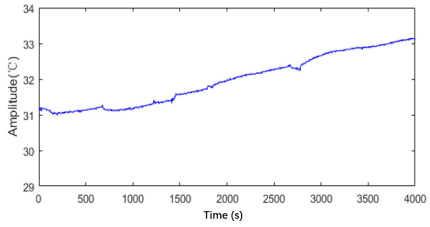


Fig. 6. The SKT signal.

Studies have shown that the skin electrical response is closely related to emotional arousals. When the human sympathetic nervous system is excited, sweat gland activity is enhanced, resulting in increased skin conductivity. Sweat gland secretion also increases while one is experiencing emotions of stress, fear, or anxiety, leading to elevated skin electricity. Therefore, the skin electrical signal is considered as one of the objective indicators to measure human emotions.

5.1.3 Skin Temperature (SKT). The E4 smart watch measures human skin temperature based on a temperature-sensitive thermistor. A typical slice of the SKT signal is shown in Figure 6, where changes of temperature can be observed.

5.2 Signal Preprocessing

5.2.1 RR Interval Extraction and Correction. Heart rate variability describes the slight changes in the human heart rate and the continuous heart rate cycle, which has been proved to be very sensitive to the sympathetic nervous system and can effectively convey the state of the sympathetic nervous stimulation. Emotion changes will affect the user's heart rate variability through the sympathetic nervous system, which can be used to characterize the emotion-changing process.

To calculate heart rate variability features, we need to extract the time series of RR intervals from the raw BVP signal, which measure the duration of heartbeat cycles. A typical BVP signal usually has two peaks (i.e., a main peak and a secondary peak) in each circle, as shown in Figure 4. The traditional peak detection method may confuse these two peaks and thus leads to a large number of fault-checked RR intervals. To address this issue, we adopt the RR interval extraction and correction algorithm proposed in our previous work [13, 25]. In particular, we first apply an overlapped sliding window method to obtain the preliminary time series of RR intervals, and then refine possible leak checks and fault checks of RR intervals based on the fact that a normal RR interval should neither be too short nor too long.

5.2.2 Detection of Skin Electrical Responses. When we experience emotion changes, there will be a skin electrical response with the following characteristics: the response signal lasts for 3∼10 seconds and usually has a slow rising edge and a fast falling edge. In particular, such responses can be divided into stimulus response and spontaneous response according to the difference in peak values, as shown in Figure 7, where the red parts of the signal are electrical responses caused by emotion changes, waveforms with higher peak values correspond to stimulus responses, and waveforms with lower peak values correspond to spontaneous responses.

The accurate detection of skin electrical responses is very important for the recognition of emotions. Therefore, we propose the following detection method based on the short-term energy and zero-crossing rate algorithm:

- *Step 1*: Remove the baseline level of the EDA signal using the third-order Butterworth filter.
- *Step 2*: Divide the signal into a set of overlapping segments; the length of each segment is 1 second (i.e., four sample points) with an overlapping length of 0.5 second (i.e., two sample points).
- *Step 3*: Calculate the signal energy of each segment as $E = \sum_{i=1}^{n} {x_i}^2$.

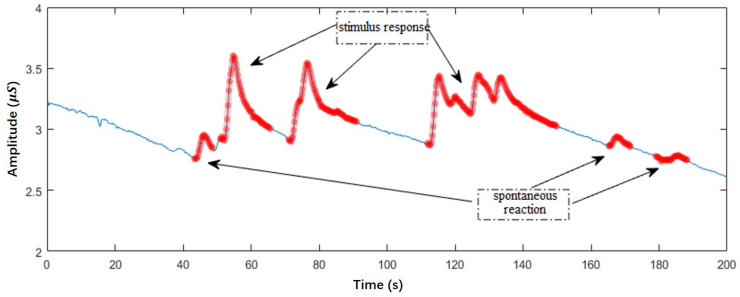


Fig. 7. Stimulus and spontaneous skin electrical responses in the EDA signal.

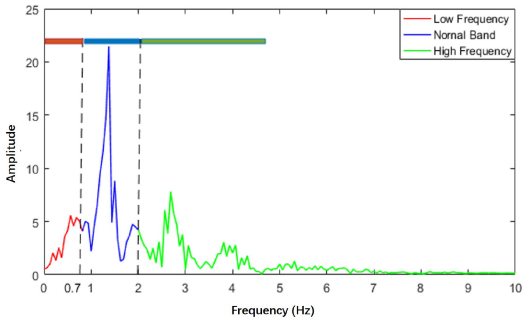


Fig. 8. FFT of the BVP signal.

- *Step 4*: Define the upper and lower energy threshold as *ETH* and *ETL*, and the upper and lower zero-crossing rate threshold as *ZCRH* and *ZCRL*. Set the maximum calm time of a skin electrical response as 3 seconds, and the minimum and maximum length of a skin electrical response as 4 seconds and 10 seconds.
- *Step 5*: Detect the starting point and end point of skin electrical responses accordingly.

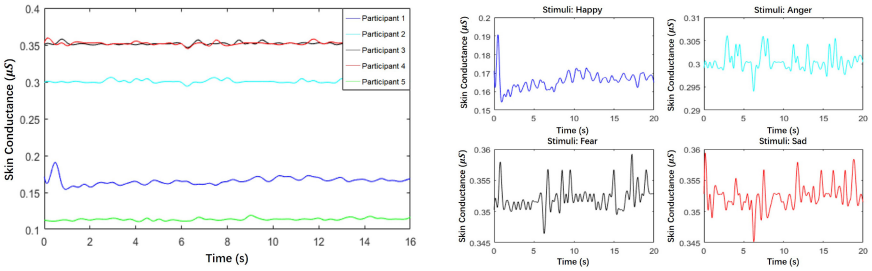
5.3 Feature Extraction

To extract emotion-related features from the obtained physiological signals, we choose to adopt a sliding window–based method. In particular, the size of the window is set as 16 seconds (i.e., we divide a time series of signal into a set of segments of 16 seconds). As different sensors have different sampling frequency, the length of each segment of the BVP signal (64 Hz), the EDA signal (4 Hz), and the SKT signal (1 Hz) are 1024, 64, and 16, respectively.

5.3.1 Blood Volume Pulse Features. A typical BVP signal is shown in Figure 4, which is quasi-periodic and thus we mainly extract frequency domain features. In particular, we utilize the 1,024-point fast Fourier transform (FFT) and partition the coefficients with the frequency range of 0∼10 Hz into three non-overlapping subbands. As the normal heart rate range is 40∼120 beats/minute (i.e., 0.7∼2 Hz), we set the subband 0∼0.7 Hz as the low-frequency (LF) band, 0.7∼2 Hz as the middle-frequency (MF) band, and 2∼10 Hz as the high-frequency (HF) band, as shown in Figure 8. We can find that the subband that corresponds to the normal heart rate range (i.e., the MF band) has much higher amplitude (i.e., energy) than the other two subbands.

First, the mean energy and the maximum energy of each subband are calculated as features, which are denoted as *P_power_LF*, *P_power_MF*, *P_power_HF*, *P_peak_LF*, *P_peak_MF*, and *P_peak_HF*.

Furthermore, to measure the uncertainty of each subband, we calculate the subband spectral entropy (SSE). In particular, to compute SSE, it is necessary to convert each spectrum into a probability mass function (PMF).



(a) EDA signals of different subjects under the same stimuli

(b) EDA signals of the same subject under different stimulus

Fig. 9. EDA signals under different conditions.

To this end, we need to normalize the spectrum energy as follows:

$$x_i = \frac{X_i}{\sum_{i=1}^{N} X_i}, i = 1, \ldots, N, \tag{1}$$

where X_i represents the energy of the i^{th} frequency component and $\tilde{x} = \{x_1, x_2, \ldots, x_n\}$ is computed as the PMF of the spectrum. Finally, the SSE is defined as

$$H_{sub} = -\sum_{i=1}^{N} x_i \cdot \log x_i. \tag{2}$$

5.3.2 Heart Rate Variability Features. Based on the time series of RR intervals, we extract a set of heart rate variability features in both the time domain and the frequency domain. Time domain features include the mean value ($H\_meanValue$), the maximum value ($H\_maxValue$), the minimum value ($H\_minValue$), the standard deviation (H_SDNN), the proportion of pairs of successive RR intervals that differ by more than 50 ms (H_PNN50), and the standard deviation of the first derivative of heart rate variability (H_STDD).

In the frequency domain, we first calculate the PSD of the RR interval time series and then extract features at three different frequency bands: the very low frequency (VLF) band (0.003~0.04 Hz), the LF band (0.04~0.15 Hz), and the HF band (0.15~0.4 Hz). Specifically, we extract the following frequency domain features: the mean power of the VLF, LF, and HF band; the ratio of power within the LF band to that within the HF band (LF/HF); the frequency of the highest peak in the VLF band (H_peak_VLF); the frequency of the highest peak in the LF band (H_peak_LF); and the frequency of the highest peak in the HF band (H_peak_HF).

5.3.3 Electrodermal Activity Features. As mentioned previously, there are two types of EDA signals: one is the SCL, which indicates the basic physiological level of the subject, and the other is the SCR, which is considered to be useful as it signifies a response to external stimuli. In this work, we use a three-order Butterworth filter to decompose the two types of signals with a cut-off frequency of 0.5 Hz. In Figure 9(a), we present the EDA signals of five different subjects under the same stimulus condition, and we can see that the five waveforms locate in different scales, which should be due to the difference of basic physiological statuses. In addition, the physiological signals of one subject under different stimuli conditions are given in Figure 9(b).

Features we extracted from the EDA signal are as follows: the mean value of SCL (E_mean_SCL), the standard deviation of SCL (E_std_SCL), the mean value of SCR (E_mean_SCR), the standard deviation of SCR (E_std_SCR), and the occurrence of SCR detected by finding two consecutive zero-crossings (E_num_SCR) (i.e., from negative to positive and positive to negative).

5.3.4 Skin Temperature Features. Variations in the skin temperature mainly come from localized changes in the blood flow, caused by the vascular resistance or the arterial blood pressure. The local vascular resistance is

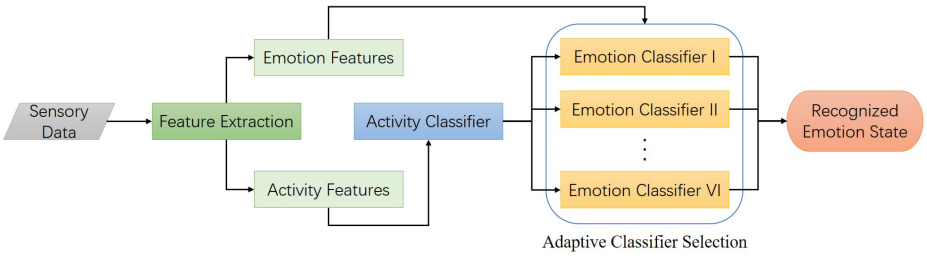


Fig. 10. Adaptive emotion recognition model.

modulated by smooth muscle tone, which is mediated by the sympathetic nervous system. The mechanism of the arterial blood pressure variation can be described by a complicated model of cardiovascular regulation by the autonomic nervous system. Thus, the skin temperature variation reflects the activity of the autonomic nervous system and is another effective measure of the emotional status. In this work, the maximum, the minimum, the mean, and the standard deviation values within a time interval of 16 seconds are extracted as the features of skin temperature (i.e., $T\_maxValue$, $T\_minValue$, $T\_meanValue$, and T_STD).

5.4 Adaptive Emotion Recognition

To recognize human emotions based on wearable devices, an adaptive emotion recognition model is proposed, which assembles multiple emotion classifiers to achieve adaptive recognition, as shown in Figure 10.

First, based on the sensory data collected with the smart watch, both activity features and emotion features are extracted. Afterward, with the activity features, the user's activity scene can be identified with the AdaBoost-based activity classifier. Then, according to the identified activity scene, we select the corresponding emotion classifier from Emotion Classifiers I through VI, which are trained based on sensory data collected under six different scenes (i.e., walking, running, riding a bicycle, going upstairs, sitting, and standing). In particular, for the recognition of emotions, we adopt four different classification algorithms: Random Forest, Neural Network, SVM, and Naive Bayes. Finally, one's real-time emotional state is recognized using the selected emotion classifier.

6 PERFORMANCE EVALUATION

In this section, we first report the experimental setups including the stimuli corpus we developed, the subjects we recruited, and the experimental design. Then, we present the evaluation results of the proposed system.

6.1 Experimental Setup

To capture the relationship between physiological signals and different human emotions, we developed an emotion stimuli corpus to evoke emotions, which consists of several different types of film clips, including comedy film, documentary film, horror film, and war film.

When constructing the stimuli corpus, we mainly selected Chinese film clips, as native culture factors may facilitate the elicitation of human emotions. The used film clips were selected by the participants themselves, which they believe can truly evoke the four types of emotions corresponding to the four quadrants in the 2-D emotion model (as shown in Figure 1). In addition, the criteria we used for selecting film clips are as follows: (1) the time duration of the whole experiment should not be too long so as to avoid visual fatigue, (2) the film clips should be easy to understand, and (3) each film clip should evoke a single target emotion. In total, we selected 20 film clips with a duration of about 2.5 hours, consisting of four happiness clips, four sadness clips, two fear clips, two anger clips, four sensation clips, and four neutrality clips, each of which lasts for about 5 minutes.

We recruited 30 volunteers (18 males and 12 females, 20–28 years old) for the experiment; most of them are undergraduate and postgraduate students, and none of them has reported cardiovascular, neurological, epileptic,

Fig. 11. The experimental setting.

or hypertensive disease. The volunteers were requested not to use caffeine, salty or fatty food 1 hour before the experiment.

We designed two different experiments for data collection. In the first one, 15 participants were asked to sit in a chair (i.e., the sitting activity scene, which is the most convenient one for the stimuli of emotions) and watch film clips one by one to complete the experiment, as illustrated in Figure 11. At the end of each film clip, the participant would have a 1-minute break, during which one could fill in the questionnaire to describe her real-time emotional status. We repeated the preceding process six times for all 20 film clips, resulting in a dataset with 1,800 samples (Data Set I), each of which consists of the sensory recording and the ground truth provided by the questionnaire.

In the second experiment, all 30 participants were divided into six equal groups, and each group would complete the experiment in one of the six activity scenes (i.e., walking, running, riding a bicycle, going upstairs, sitting, and standing). Specifically, the participant was first asked to perform a certain activity for about 1 minute so as to enter the target activity state. Afterward, a film clip would be played to evoke the participant's emotion. The preceding process was repeated six times for all 30 participants and 20 film clips, resulting in another dataset with 3,600 samples (Data Set II).

6.2 Emotion Recognition Performance

6.2.1 Overall Recognition Performance. The emotion recognition performance is evaluated by means of the correct classification ratio (CCR), which is defined as

$$CCR = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%. \tag{3}$$

To validate the performance of individual emotion recognition, we used the SVM method to train an individual classifier for each participant based on Data Set I, and the results are shown in Table 2. In particular, we first divided the samples into four sub-groups according to the 2-D emotion model (i.e., LANV, LAPV, HANV, and HAPV). Based on such a classification strategy, we can analyze the arousal and the valence dimension, respectively. For the classification of the arousal emotion, the average CCR is 87.00%, where participant ♯9 has the highest CCR of 93.00%, and participant ♯13 has the lowest CCR of 78.00%. For the classification of the valence emotion, the average CCR is 84.56%, where participant ♯9 has the best CCR of 94.67%, and participant ♯1 has the lowest CCR of 74.67%. We can find that the recognition performance of the arousal emotion is higher than that of the valence emotion, and the reason might be that human emotion changes are more sensitive to the arousal dimension.

Table 2. Individual Recognition Performance of Four Types of Emotions, Arousal, and Valence

Participant ID	Four Types of Emotions (%)	Arousal (%)	Valence (%)
♯1	73.00	79.67	74.67
♯2	81.33	81.33	83.00
♯3	83.00	88.00	84.67
♯4	74.67	79.67	78.00
♯5	86.33	93.00	89.67
♯6	84.67	89.67	86.33
♯7	81.33	86.33	83.00
♯8	86.33	84.67	88.00
♯9	91.33	93.00	94.67
♯10	89.67	91.33	86.33
♯11	84.67	91.33	79.67
♯12	86.33	89.67	91.33
♯13	83.00	78.00	83.00
♯14	84.67	86.33	79.67
♯15	86.33	93.00	86.33
Average	83.78	87.00	84.56

Table 3. Overall Recognition Performance of Four Types of Emotions

	LANV	LAPV	HANV	HAPV
LANV	0.853	0.084	0.009	0.049
LAPV	0.102	0.804	0.067	0.093
HANV	0.009	0.058	0.907	0.080
HAPV	0.036	0.053	0.018	0.778
Overall (%)				83.56

To further validate the overall performance of emotion recognition, we used all samples in Data Set I to build a general model. The overall recognition performance of four types of emotions is shown in Table 3, where the best recognition accuracy is 90.7% for HANV and the lowest is 77.8% for HAPV. In particular, we can find that most of the misclassified HAPV are recognized as LAPV, and there is an extreme uncertainty between LANV and LAPV.

According to the results of individual emotion recognition model (Table 2) and the general emotion recognition model (Table 3), we find that the performance of the general model is similar to the average performance of each individual model, which demonstrates that the proposed model is able to address the issue of individual differences. In other words, it is possible to construct a general model that can be used to recognize emotions of untrained subjects.

6.2.2 Comparison of Different Classification Methods. Based on Data Set I, the classification performance of different classifiers (i.e., Random Forest, Neural Network, SVM, and Naive Bayes) are compared in this section. Considering individual differences, all of the classifiers are evaluated using leave-one-out cross validation. Figure 12 shows the classification results of arousal, valence, and four types of emotions using different classifiers. We can see that SVM achieves the best performance for the classification of arousal, valence, and four types of emotions, indicating that SVM is more suitable for our dataset. Other classifiers have quite similar

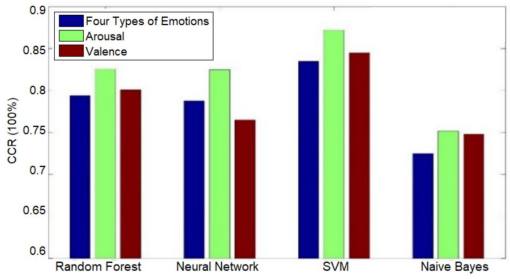


Fig. 12. Comparison of different classification methods.

Table 4. Adaptive Emotion Recognition vs. Non-Adaptive Emotion Recognition

	Precision	Recall	AUC
Adaptive method	0.743	0.771	0.769
Non-adaptive method[a]	0.461	0.483	0.502

[a]The non-adaptive approach constructs a single recognition model without considering the activity scene.

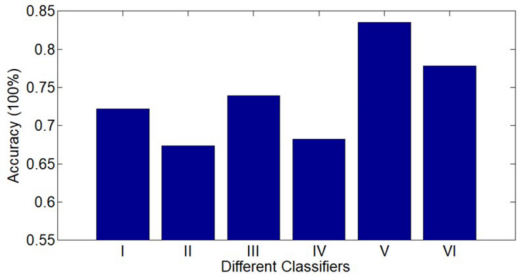


Fig. 13. Overall recognition performance of different scene-adaptive emotion classifiers.

performance especially for Random Forest and Neural Network, whereas Naive Bayes has the lowest CCR among all of the classifiers.

6.2.3 Performance of Adaptive Emotion Recognition. Based on the framework shown in Figure 12, we evaluate the adaptive emotion recognition approach using Data Set II as follows. First, the user's activity scene is identified, and then one's emotion is recognized using the emotion classifier corresponding the identified activity scene, achieving the purpose of scene-adaptive emotion recognition. According to Table 4, we can see that the accuracy, recall, and AUC of the adaptive approach are 28.2%, 28.8%, and 26.7% higher than those of the non-adaptive approach. This is because the non-adaptive approach does not consider the influence of different activities scenes, which may completely submerge the emotion-related information contained in the sensory data, making it difficult to distinguish different emotions.

In particular, we give the overall recognition performance of the six scene-adaptive emotion classifiers in Figure 13 (Classifier-I: walking, Classifier-II: running, Classifier-III: riding a bicycle, Classifier-IV: going upstairs, Classifier-V: sitting, and Classifier-VI: standing), which are 0.722, 0.674, 0.739, 0.682, 0.835, and 0.778, with an average accuracy of 0.743. We can find that the recognition accuracy of these six classifiers is different, where the highest and lowest accuracy correspond to the scene of sitting and running, respectively. The reason might

be due to the fact that data captured in the sitting scene is of high quality, whereas data captured in the running scene may contain much more noise.

6.3 Discussion

In this work, we only considered several activity scenes, including walking, running, riding a bicycle, going upstairs, sitting, and standing, and developed corresponding emotion recognition models for each of them. Possible activity scenes are more complicated, which means that new models should be trained and added to the adaptive emotion recognition framework.

Moreover, during the data collection process, it is difficult to evoke emotions in some of the activity scenes (e.g., running), which leads to the problem of data sparsity. As a result, the emotion recognition accuracy in such scenes is lower than those in other scenes.

Meanwhile, when annotating the sensory data, we used the Arousal-Valence 2-D emotion model, in which each dimension is divided into two levels, and the participant's rating is used as the ground truth. However, the boundary between some of the emotions could be vague, which makes it difficult to distinguish such emotions subjectively, and brings some errors to the annotations. Therefore, we plan to divide each dimension into more levels to reduce the impact of ambiguity in emotional boundaries on recognition performance.

7 CONCLUSION

In this work, we used a sensor-enriched smart watch to collect multi-mode physiological signals, based on which we extracted a set of fine-grained features to characterize and recognize different emotions. Specifically, to deal with the problem that other factors (e.g., activities) may also impact one's physiological signals, we proposed an adaptive emotion recognition framework, which is able to recognize human emotions according to the user's activity scenes. We first developed an activity scene identification method by exploring the accelerometer sensory data to distinguish six different activities. Afterward, based on the result of activity scene identification, we further built a set of emotion classifiers to recognize different emotions accordingly, and mainly explored the BVP, EDA, and SKT signals to characterize different emotions by extracting a set of emotion-related features. An overall accuracy of 74.3% for 30 participants demonstrates that the proposed system can recognize human emotions effectively.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to the volunteers for participating in the experiment.

REFERENCES

[1] R. Bailon, L. Sornmo, and P. Laguna. 2006. A robust method for ECG-based estimation of the respiratory frequency during stress testing. *IEEE Transactions on Biomedical Engineering* 53, 7 (2006), 1273–1285.

[2] J. Cabibihan and S. S. Chauhan. 2017. Physiological responses to affective tele-touch during induced emotional stimuli. *IEEE Transactions on Affective Computing* 8, 1 (2017), 108–118.

[3] Yixiang Dai, Xue Wang, Pengbo Zhang, and Weihang Zhang. 2017. Wearable biosensor network enabled multimodal daily-life emotion recognition employing reputation-driven imbalanced fuzzy classification. *Measurement* 109 (2017), 408–424.

[4] Guido H. E. Gendolla. 2000. On the impact of mood on behavior: An integrative theory and a review. *Review of General Psychology* 4, 4 (2000), 378–408.

[5] Y. Hsu, J. Wang, W. Chiang, and C. Hung. 2020. Automatic ECG-based emotion recognition in music listening. *IEEE Transactions on Affective Computing* 11, 1 (2020), 85–89.

[6] Eun-Hye Jang, Byoung-Jun Park, Mi-Sook Park, Sang-Hyeob Kim, and Jin-Hun Sohn. 2015. Analysis of physiological signals for recognition of boredom, pain, and surprise emotions. *Journal of Physiological Anthropology* 34, 1 (2015), Article 25, 12 pages.

[7] J. Kim and E. Andre. 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 12 (Dec. 2008), 2067–2083.

[8] A. Kleinsmith and N. Bianchi-Berthouze. 2013. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing* 4, 1 (Jan. 2013), 15–33.

[9] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. 2012. DEAP: A database for emotion analysis ;using physiological signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31.
[10] D. Kulic and E. A. Croft. 2007. Affective state estimation for human-robot interaction. *IEEE Transactions on Robotics* 23, 5 (2007), 991–1000.
[11] P. Kuppens, F. Tuerlinckx, J. Russell, and L. Barrett. 2013. The relation between valence and arousal in subjective experience. *Psychological Bulletin* 139, 4 (2013), 917–940.
[12] M. Kusserow, O. Amft, and G. Troster. 2013. Modeling arousal phases in daily living using wearable sensors. *IEEE Transactions on Affective Computing* 4, 1 (2013), 93–105.
[13] Fan Liu, Xingshe Zhou, Zhu Wang, Jinli Cao, Hua Wang, and Yanchun Zhang. 2019. Unobtrusive mattress-based identification of hypertension by integrating classification and association rule mining. *Sensors* 19, 7 (2019), Article 1489, 25 pages.
[14] I. Mauss, R. Levenson, L. McCater, F. Wilhelm, and Gross J. 2005. The tie that binds—Coherence among emotion experience, behavior, and physiology. *Emotion* 5, 2 (2005), 175–190.
[15] Michele Orini, Raquel Bailón, Ronny Enk, Stefan Koelsch, Luca T. Mainardi, and Pablo Laguna. 2010. A method for continuously assessing the autonomic response to music-induced emotions through HRV analysis. *Medical & Biological Engineering & Computing* 48, 5 (2010), 423–433.
[16] Markus Quirin, Miguel Kazén, and Julius Kuhl. 2009. When nonsense sounds happy or helpless: The implicit positive and negative affect test (IPANAT).*Journal of Personality and Social Psychology* 97, 3 (2009), 500–516.
[17] Pierre Rainville, Antoine Bechara, Nasir Naqvi, and Antonio R. Damasio. 2006. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology* 61, 1 (2006), 5–18.
[18] Georgios Rigas, Christos D. Katsis, George Ganiatsas, and Dimitrios I. Fotiadis. 2007. A user independent, biosignal based, emotion recognition method. In *Proceedings of the International Conference on User Modeling*. 314–318.
[19] N. Sarode and S. Bhatia. 2010. Facial expression recognition. *International Journal on Computer Science and Engineering* 2, 5 (2010), 1552–1557.
[20] K. Schaaff and T. Schultz. 2009. Towards emotion recognition from electroencephalographic signals. In *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. 1–6.
[21] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert. 2010. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010), 410–417.
[22] K. Wac and C. Tsiourti. 2014. Ambulatory assessment of affect: Survey of sensor systems for monitoring of autonomic nervous systems activation in emotion. *IEEE Transactions on Affective Computing* 5, 3 (2014), 251–272.
[23] J. Wang, Y. Wang, D. Zhang, and S. Helal. 2018. Energy saving techniques in mobile crowd sensing: Current state and future opportunities. *IEEE Communications Magazine* 56, 5 (2018), 164–169.
[24] J. Wang, Y. Wang, D. Zhang, Q. Lv, and C. Chen. 2019. Crowd-powered sensing and actuation in smart cities: Current issues and future directions. *IEEE Wireless Communications* 26, 2 (2019), 86–92.
[25] Zhu Wang, Xingshe Zhou, Weichao Zhao, Fan Liu, Hongbo Ni, and Zhiwen Yu. 2017. Assessing the severity of sleep apnea syndrome based on ballistocardiogram. *PLoS ONE* 12, 4 (2017), 1–24. https://doi.org/10.1371/journal.pone.0175351
[26] Siqing Wu, Tiago H. Falk, and Wai-Yip Chan. 2011. Automatic speech emotion recognition using modulation spectral features. *Speech Communication* 53, 5 (2011), 768–785.
[27] B. Zhao, Z. Wang, Z. Yu, and B. Guo. 2018. EmotionSense: Emotion recognition based on wearable wristband. In *Proceedings of the 15th IEEE International Conference on Ubiquitous Intelligence and Computing*. 346–355.
[28] M. D. Zwaag, J. H. Janssen, and J. M. Westerink. 2013. Directing physiology and mood through music: Validation of an affective music player. *IEEE Transactions on Affective Computing* 4, 1 (2013), 57–68.

Received July 2019; revised December 2019; accepted February 2020