Cross-task emotion recognition using EEG measures: first step towards practical application

LIU Shuang[1], MENG Jiayuan[1], ZHAO Xin[1], YANG Jiajia[1], HE Feng[1],
QI Hongzhi[1], ZHOU Peng[1], HU Yong [2], MING Dong [1,*]
(1. *College of Precision Instruments & Optoelectronics Engineering*, *Tianjin University*, *Tianjin* 300072, *China*;
2. *Li KaShing Faculty of Medicine*, *The University of Hong Kong*)

Abstract: Electroencephalographic (EEG)-based emotion recognition has received increasing attention in the field of human-computer interaction (HCI) recently, there however remains a number of challenges in building a generalized emotion recognition model, one of which includes the difficulty of an EEG-based emotion classifier trained on a specific task to handle other tasks. Little attention has been paid to this issue. The current study is to determine the feasibility of coping with this challenge using feature selection. 12 healthy volunteers were emotionally elicited when conducting picture induced and videoinduced tasks. Firstly, support vector machine (SVM) classifier was examined under within-task conditions (trained and tested on the same task) and cross-task conditions (trained on one task and tested on another task) for pictureinduced and videoinduced tasks. The within-task classification performed fairly well (classification accuracy: 51.6% for picture task and 94.4% for video task). Cross-task classification, however, deteriorated to low levels (around 44%). Trained and tested with the most robust feature subset selected by SVM-recursive feature elimination (RFE), the performance of cross-task classifier was significantly improved to above 68%. These results suggest that cross-task emotion recognition is feasible with proper methods and bring EEG-based emotion recognition models closer to being able to discriminate emotion states for any tasks.

Key words: Emotion recognition; Electroencephalographic (EEG), cross-task recognition, support vector machine-recursive feature elimination (SVM-RFE)

1 Introduction

Emotion is a psycho-physiological process triggered by conscious and/or unconscious perception of an object or a situation, which is often associated with mood, temperament, personality disposition and motivation[1]. Recently emotion recognition has received increasing attention in the field of human-computer interaction (HCI), since it would be valuable to develop an instrument capable of recognizing a person´s emotional states. If machines could perceive a person's emotional state, HCI may become more intuitive, smoother, and more efficient. Additionally, it is extremely necessary to recognize negative emotions for some special workers, such as astronauts and drivers. The accumulation of negative emotions not only affects their work quality and efficiency severely, but also impedes the work of immune system, making people more vulnerable to viral infections and increasing recovery times from surgery or disease[2].

Currently increasing efforts have been deployed to recognize human emotions using physiological responses, among which electroencephalogram (EEG), electrocardiogram (ECG), respiration and skin conductance are the four most prevailing physiological signals used in emotion studies. Many scientific reports used EEGs as an independent measure to recognize emotion states but others always perform together, two or more is common. EEG signal also provides considerable advantages, being largely involuntary and generally less mediated by cognitive and social influences, opposedto audiovisual recordings such as facial expressions, voice and gesture, thus EEG has been a primary option for the development of online emotion recognition systems. In these existing studies, moreover, many different materials have been used to elicit emotions in the laboratory such as facial expressions[3], pictures[4], texts[5], music [6], and movies[7]. Specially, affective pictures and videos are the two most popular evoked stimuli. Yohanes et al. proposed discrete wavelet transform coefficients for emotion recognition from EEG signals in response to emotional pictures. Their study achieved a maximum recognition rate of above 84.6% for two emotional states[8]. Koelstra et al. [1] presenteda multimodal data set of electroencephalogram (EEG) and peripheral physiological signals for the analysis of human affective states, and the average accuracy of the valence and arousal estimation is about 60% for two-level class case. Meanwhile the database is made publicly available, then Yoo et al. [9] used the open database and performed EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm, for the two-level class case, the average accuracy of the valence and arousal estimation is 70.9% and 70.1%, respectively. For the three-level class case, the average accuracy is 55. 4% and 55.2%, respectively.

Nearly all existing studies tend to provide evidence for the feasibility of using EEG measures in the monitoring of emotional states. However, there has been few study reports on real-world applications to date, and to be used in practical settings, EEG-based emotion recognition systemwould require a certain flexibility in recognizing emotions so that it could generalize to other time, other subjects and even a new situation to induce the same emotions[10]. If a model trained by the data of a specific individual performing a specific task at the specific time can recognize emotion states of the same individual performing the same task at the same time. It can be thought that the classification was indeed based on the individual, task type and the time[11]. Of these three challenges, matters of subject-dependent and time-dependenthave been addressed in only a few papers[2, 12]. However, the task-dependentissueappears to be a more intractable challenge and there has been no effort to investigate this issue. Thus this paper was to cope with the problematic effects of task-to-task variations, and then tried to find out whether the performance of cross-task emotion recognition can be improved with proper feature selection.

This paper is organized as follows. Section 2 addresses the methodology, including the experimental setup and data analysis. Section 3 details the performance with within-task and cross-task respectively, and further cross-task classification using Support vector machine-recursive feature elimination. The discussion and conclusion are stated in sections 4 and 5, respectively.

2 Materials and methods

2.1 Emotion model

As all people express their emotions differently, it is not an easy task to judge or to model human emotions. Researchers often use two different models to express emotions. One approach is to label emotions in discrete categories, like anger, happiness, love, etc[13]. Some psychologists, however, pointed out that emotions are not discrete but rather continuous phenomena. Another way is to have multiple dimensions or scales to categorize emotions. A common continuous model is the 2-D model which consists of the affective dimensions of valence and arousal[14]. Valence represents the degree of personal pleasure, from unpleasantness to pleasantness. For example, happiness has a positive valence, while disgust has a negative valence. Arousal expresses the degree of excitement felt by subjects, from calmness to excitement. For example, sadness has low arousal, whereas fear has a high arousal level. Different emotional labels could be plotted at various positions on this 2-D space.

2.2 Experimental setup

In the current study, a specifically designed protocol including two tasks, picture induced task and video induced task, was conducted through an emotion elicitation process. Two tasks were conducted on different days to avoid subjects being tired and influenced by the form of stimulation. Concretely, emotions were induced by the picture task onthefirst day, and video task on at roughly the same time of the other day. The interval between two tasks was about one week. Both utilized the stimuli containing the categories of neutral, negative and positive emotions. The experimental setup consists of the recruitment of eligible subjects, emotional elicitation and the acquisition of EEG data.

2.2.1 Subjects

A group of 12 healthy participants (6 female, 6 male) were enrolled in this study. All participants had normal or corrected-to-normal vision and normal hearing, and none of them had a history of severe medical treatment, psychological or neurological disorders. A signed consent was obtained from each subject before the experiment was carried out.

2.2.2 Emotional Elicitation

To evaluate whether the affective pictures and movie clips excite specified emotional states or not, 58 subjects who did not take part in the experiment participated in a questionnaire survey to verify the effectiveness of these elicitors before the experiment. Finally, 145 of 300 pictures and 15 of 72 movie clips were selected with 3 clips for each emotional state, 45 pictures for happy while 25 pictures for each emotional state of neutral, tense, sad and disgust states, respectively. These pictures and movie clips could elicit single desired emotion. The experimental protocol included the series of steps schematically depicted in Fig.2 and Fig.3 for the picture induced and video induced tasks, respectively.

In the picture induced task, affective elicitation was performed by projecting a set of pictures onto a PC monitor.As can be seenin Fig.2, the slide show was comprised of five sessions of pictures, each containing 29 trials. In a trial, a 1-second cue was displayed to imply the commencement of the trial and a 5-second picture display followed by self-assessment on a discrete 9-point scale for valence and arousal then took place.The subjects were required to rate valence, arousal and the specific emotion they had experience during picture viewing. It is worthy to note that, furthermore, of the 29 trials in each session, five neutral pictures were displayed first, followed by nine happy ones, five tense ones, five sad ones and, last, five disgust pictures. It sequences five emotions in a way that supposedly makes it easier for many people to transition from an emotion state to another.

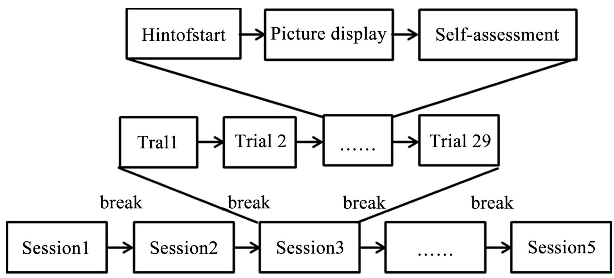

Fig. 1 Schematic representation of the picture induced task

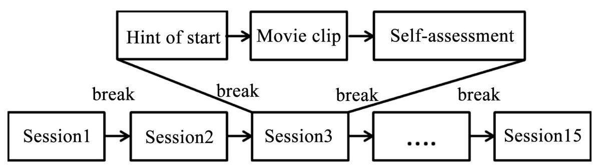

Fig. 2 Schematic representation of the video induced task

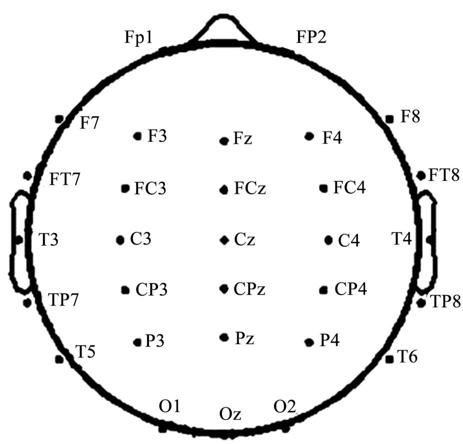

Fig. 3 EEG cap layout of 30 channels

In the video induced task, a group of movie clips were used as elicitors. The experiment consisted of 15 sessions, and in each session, a movie clip was displayed for about 5 minutes, preceded by a 5s red circle as the hint of start. At the end of each clip, the subjects were also required to rate valence, arousal and the specific emotion they had experience during movie viewing. Each session was followed by a short break, and the recordings took place whenever the subject was ready to watch the next video. During the video task, StereoPhilips speakers were used and the video volume was set at relatively loud level. However, each participant was asked before the task whether the volume was comfortable and it was adjusted when necessary.

It is should be noted that the subjects' ratings during the experiment are used as the real data label for the emotion classifier. Furthermore, it was then necessary to decide upon a procedure which enabled the elicitation of emotions in such a way that they were neither biased by the original mood of the participant nor by emotions elicited early on in the experiment. To achieve this, we adopted a procedure whereby participants would rest for 5 minutes before beginning the experiment. Each participant was then shown materials in the order of neural, happy, tense, sad, and disgust.

For presentation of the stimuli and the recording of the subjects' ratings, Eprime2.0 software was used. The pictures and videos were presented on a 21-inch screen and in order to minimize eye movements, all stimuli were displayed at the resolution of 1024×768.

2.2.3 Data collection

For the experiment, a quiet listening room was prepared in order to ensure that the subjects could experience the emotions evoked by the pictures or the videos undisturbed. 30-channels EEG signals were recorded continuously using a Neuroscan4.5 amplifier system. The electrodes were placed on the scalp according to the extension of the international 10-20 electrode positioning system[15]. EEG cap layout of 30 channels is shown in Figure 5. All channels were referenced to right mastoid and grounded central region. The signals were filtered at 1-45 Hz, digitized at 1,000 Hz and stored in a PC for offline analysis. Eye movements and blinks were monitored by recording the horizontal and vertical EOG.

2.3 Data Processing

2.3.1 Preprocessing

Prior to calculating features, a preprocessing stage of the EEG signals is required. All channels were re-referenced to bilateral mastoid, down-sampled to 128 Hz, and filtered at 1-45H. EOG artifacts were removed using independent component analysis (ICA). In the picture induced task, 1s-measurement from 1s to 2s after each picture onset was intercepted for further analysis. 1s-measurement was considered as a valid sample in the subsequent analysis. In the video induced task, however, data segments were picked out according to the subjects' self-report about which period of time they felt the emotion strongly. 1s-measurement was also regarded as a sample in subsequent analysis, thus dozens of samples can be drawn in each video clip.

2.3.2 Feature extraction

The spectral power of EEG signals in different bands was found to correlate with emotions[16, 17]. In this method, power spectral density (PSD) from different bands was computed using AR model with the Burg algorithm. In AR model, the series were estimated by a linear difference equation in time domain:

$$x(n) = -\sum_{i=0}^{p} \alpha_i x(n-i) + w(n)$$

Here a current sample x(n) is a linear function of pprevious samples plus an independent and identically distributed white noise input w(n). The Burg approach was employed to estimate AR parameters since AR model obtained with this approach was always stable. Additionally, it has a relatively better resolution compared with auto correlation method.

Average spectral power from six frequency bins were picked out as features in the following analysis, including θ(4~8 Hz), α(8~13 Hz), β1 (13~18 Hz), β2(18~30 Hz), γ1 (30~36Hz) and γ2 (36~44 Hz). Therefore, 180 (6 per channel × 30 channels) features were included in each feature set.

2.3.3 Support vector machine

SVM is a supervised learning algorithm which uses a discriminant hyperplane to identify classes. The goal of SVM is to find an optimal hyperplane with the maximal margin between two classes of data [25]. In the emotion recognition process, the features were mapped using Gaussian radial basis function into high-dimensional kernel space.

$$k(x,y) = exp\ (\|x-y\|^2/2\,\sigma^2)$$

The penalty parameter C was set to 1 as the default value in the LibSVM toolkit.

In this paper, the data from both tasks (picture induced task, P; video induced task, V) were used to train and test multiple SVMs for each participant. Each combination of training and testing sets was labeled accordingly (PP, VV, PV, VP). "PP" labelled the within-task condition, in which the training and testing sets came from the picture induced task. Similarly, "VV" labelled the condition under which both the training and testing sets came from the video induced task. The condition under which the SVM was trained on picture induced task (P) and tested on video induced task (V) was labeled with "PV", while the condition under which they were trained on video induced task (V) and tested on picture induced task (P) was labeled with "VP". 180 (6 per channel × 30 channels) features were included in each feature set that was mentioned above.

2.3.4 Recursive feature elimination (RFE)

Since not all the features carry significant information, features selection is necessary for decreasing and discarding redundant features that can potentially deteriorate classification performance. SVM-RFE (Lal et al., 2004) consists of a wrapped method of feature selection and is based on the powerful SVM classification, which has been successfully used for numerous applications. This feature selection method was proposed by Guyon et al. (Guyon and Elisseeff, 2003) and was based on the concept of margin maximization. The importance of a dimension is determined by the influence it has on the margin of a trained SVM. Let W be the inverse of the margin:

$$W(X,Y,C) = \frac{1}{\gamma(X,Y,C)} = \|w\|_2$$

At each iteration one SVM is trained and the features $\widehat{j}$ which minimize $|W(X,Y,C) - W(X^{\hat{j}},Y^{\hat{j}},C)|$ are removed (typically that is one feature only); this is equivalent to removing the dimensions $\widehat{j}$ that correspond to the smallest $|w_j|$.

In this case, however, the ranking criterion was modified as follows: In an N-dimension feature set, each feature was removed once and then got N performances with the remaining N-1 features. The feature, without which the feature set obtained the best accuracy, was considered as the one with the minimum contribution. At each iteration step we remove the feature with the minimum contribution from the feature set until only one feature remained. The features were removed one at a time, and there was a corresponding feature ranking. But it should be noted that the features that are top ranked are not necessarily the ones that are individually most relevant. Only taken together the features of a subset are optimal in some sense.

3 Results

3.1 Within- and cross-task classification using SVM

In valence-arousal model, the samples of five emotions were divided into three groups, negative valence (disgust, sad, and tense), positive valence (happy) and neutral valence. The 3-class classification was firstly performed using all the 180 features under both within-task (PP, VV. 5-fold cross-validation) and cross-task (VP, PV) conditions. As illustrated in Fig. 4, within-task performances were 51.58% and 94.37% for the picture induced and video induced tasks. But cross-task performances were just 43.34% and 44.55% for the VP and PV conditions. These results suggest that the degradation of classification performance for emotion recognition may occur when training sets and testing sets come from different tasks. This may lie in EEG features that have been deteriorated by emotion-irrelevant information, such as task type effects. Overcoming these facts maybe the most difficult but important step in generalizing anemotion recognition model to handle multiple tasks in the real world. Feature selection may be one of possibly feasible ways, which can pick out the most emotion-relevant features.

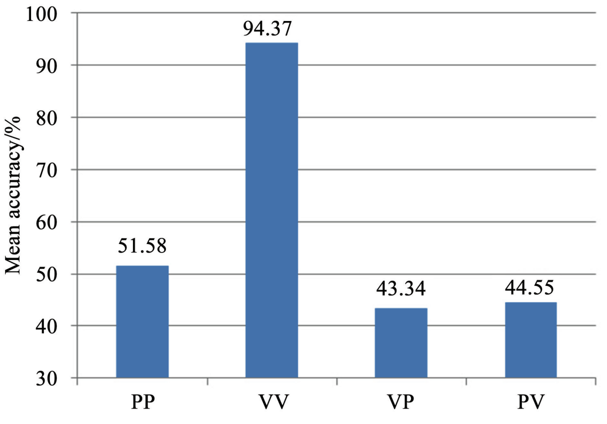

Fig. 4 Within- and cross-task classification rates using SVM

3.2 Cross-task classification using SVM-RFE

As mentioned above, cross-task classification performances may suffer from emotion-irrelevant effects, such as EEG patterns induced by specific tasks, mismatched emotional intensity between training and testing tasks and temporal effects. It was expected that SVM-RFE did at least partly avoid the emotion-irrelevant effects which causeddegradation of classification performance. Fig. 6 shows the mean cross-task classification rates using SVM-RFE. It can clearly be seen that SVM-RFE can significantly improve the mean accuracies and recognition rates jumped to average accuracy of 68.2% and 68.21% in the VP and PV conditions, respectively.

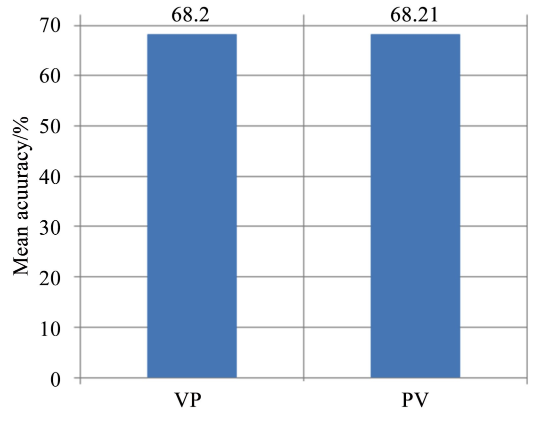

Fig. 5 Cross-task classification using SVM-RFE

Confusion matrices in Table 1 afforded a closer look at the sensitivity of three valence states (positive, negative and neutral valence). In these confusion matrices, the row represents the predicted label and column represents the true label. For example, the value of 0.071 in Table1 (a) represented there were 7.1% of neutral samples being classified as positive samples. It can be intuitively found that negative emotion was perfectly recognized, 90.6% and 91.4% for VP cross-task and PV cross-task classification, respectively. Furthermore, positive valence and neutral valence were falsely recognized as negative valence for a relative higher proportion.

Table 1 Confusion matrices achievedby SVM-RFE

a. VP cross-task classification

VP	Pos	Neu	Neg
Pos	0.361	0.082	0.557
Neu	0.071	0.423	0.506
Neg	0.030	0.064	0.906

b. PV cross-task classification

PV	Pos	Neu	Neg
Pos	0.558	0.040	0.402
Neu	0.077	0.223	0.700
Neg	0.059	0.027	0.914

4 Discussion

EEG-based emotion recognition is a promising new field of the HCI research. To be used in real-world settings, emotion recognition model would accurately recognizeemotion states no matter what kind of task that the operator would encounter. Regrettably there were few emotion studies available tocope with the issue. This work designed two tasks, picture induced and video induced tasks, to handle the cross-task challenge with the SVM-RFE that picked out the emotion-relevant and task-irrelevant features. The cross-task model may be still far away from ultimately practicalemotion classifiers, which can handle any task that the operator would encounter, but it is the first crucial step towards developing a generalized emotion recognition model. The degradation of cross-task classification performance may be at least due toseveral hypotheses, such as the mismatched emotional intensity, EEG patterns related to task type, and so on.

To date the labelling of emotions hasn't yet generated a consolidated standard, and researchers have used two major models to label the emotions: discrete emotions model and dimensional model. The former uses words and expressions to describe clearly separable states, such as anger, fear and joy, etc. However, there are issues with comparing results across different studies, as widely different sets of emotion labels are used. The second method, the dimensional approach, describes the emotion by their position in n-dimensional space and valence-arousal space is widely used in existing researches. However, the sizeable overlap between some emotions in two-dimensional space should be a shortcoming for this emotion model, such as anger and fearful[14]. Given the central role of emotion self-report in this research area, there should be efforts to develop new instruments that avoid some of the shortcomings of the existing approaches, like goingbeyond a simple valence-arousal space in order to be better able to differentiate qualitatively different states that share the same region in this space.

ACKNOWLEDGMENT

This paper was supported by National Natural Science Foundation of China (No. 81222021, 61172008, 81171423, 81127003,), National Key Technology R&D Program of the Ministry of Science and Technology of China (No. 2012BAI34B02) and Program for New Century Excellent Talents in University of the Ministry of Education of China (No. NCET-10-0618).

References

[1] KOELSTRA S, MUHL C, SOLEYMANI M, et al. DEAP: A Database for Emotion Analysis ; Using Physiological Signals[J]. Affective Computing, IEEE Transactions on, 2012, 3(1):18 - 31.

[2] PICARD R W, VYZAS E, HEALEY J. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(10): 1175 - 1191.

[3] J E, HJ J, PE. P. Emotion recognition via facial expression and affective prosody in schizophrenia - A methodological review[J]. Clinical Psychology Review, 2002, volume 22(6):789-832(44).

[4] H P, F L, C. D. Feature Selection Based On Mutual Information: Criteria Of Max-Dependency, Max-Relevance, And Min-Redundancy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8):1226 - 1238.

[5] ALM C O. Emotions from text: machine learning for text-based emotion prediction[J]. In Proceedings of HLT/EMNLP, 2005.

[6] KIM J, ANDRE E. Emotion Recognition Based on Physiological Changes in Music Listening[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008, 30(12):2067 - 2083.

[7] WANGX, NIE D, LU B. Emotional state classification from EEG data using machine learning approach[J]. NEUROCOMPUTING, 2014, 129(4):94-106.

[8] RE Y, W S, GB. H. Discrete Wavelet Transform. Coeffi-

cients For Emotion Recognition From Eeg Signals[J]. Conf Proc IEEE Eng Med Biol Soc, 2012, 49(4):2251 - 2254.

[9] HJ Y, SY. C. EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm.[J]. Computers in Biology and Medicine, 2013, 43(12):2230-2237..

[10] HAYNES, J.D. AND REES G.Decoding mental states from brain activity in humans[J]. Nature reviews. Neuroscience, 2006, 7(7):523-534.

[11] Z W, RM H, Z W, et al. Cross-subject workload classification with a hierarchical Bayes model.[J]. Neuroimage, 2012, 59(1):64-69.

[12] CHUEH T, CHEN T, LU H H, et al. STATISTICAL PREDICTION OF EMOTIONAL STATES BY PHYSIOLOGICAL SIGNALS WITH MANOVA AND MACHINE LEARNING[J]. INTERNATIONAL JOURNAL OF PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE, 2012, 26(4):1025-1035.

[13] HOSEINGHOLIZADE S, GOLPAYGANI M R H, MONFARED A S. Studying emotion through nonlinear processing of EEG[J]. Procedia - Social and Behavioral Sciences, 2012, 32(1):163-169.

[14] SCHERER K R. What are emotions? And how can they be measured? [J]. Social Science Information, 2005, 44(4):695-729.

[15] JASPER H H, The ten twenty electrode system of the international federation [J]. Electroencephalography and Clinical Neurophysiology, 1958..

[16] DAVIDSON R J. Affective neuroscience and psychophysiology: toward a synthesis.[J]. Psychophysiology, 2003, 40(5):655-665.

[17] CHANEL G, KIERKELS J J M, SOLEYMANI M, et al. Short-term emotion assessment in a recall paradigm [J]. International Journal of Human-Computer Studies, 2009, 67(8):607-627.

Authors' Biographies

LIU Shuang, born in 1989, received B.S. degree in biomedical engineeringat Tianjin Medical University in 2008. Now she is a PhD candidate inthe Neural Engineering and RehabilitationLaboratory of Tianjin University. Her current research interests include the physiological mechanism of emotion and emotion recognition.

MING Dong, born in 1976, received his B.S. andPh.D. degrees in biomedical engineeringat Tianjin University (TJU), China, in1999 and 2004, respectively. Heworked as a research associate in theDepartment of Orthopaedics and Traumatology, Li KaShing Faculty of Medicine, University of Hong Kong, from2002 to 2003 and was a visiting scholar in the Division ofMechanical Engineering and Mechatronics, University of Dundee,U.K., from 2005 to 2006. Now he is the Chair Professor of the Department of Biomedical Engineering of TJU and the head of the Neural Engineering and RehabilitationLaboratory of TJU. He is also the Chair of IEEE-EMBS Tianjin Chapter.His major research interests include neural engineering,rehabilitation engineering, sports science, biomedicalinstrumentation and signal/image processing, especially infunctional electrical stimulation, gait analysis, and brain-computer interface.He has also managed over 10 national and international research projects, organized and hosted over 10 international conferences as the Session Chairor Track Chair over the last 10 years and was the General Chair of the 2012 IEEE International Conference on Virtual Environments, Human -Computer Interfaces and Measurement Systems (VECIMS 12).Furthermore, he has been an International Advisory Board member of the The Foot, and theEditorial Committeemember of Acta Laser Biology Sinica, and International Journal of Biomedical Engineering in China.

E-mail: richardming@tju.edu.cn