Cluster-Based Analysis for Personalized Stress Evaluation Using Physiological Signals

Qianli Xu, *Member, IEEE*, Tin Lay Nwe, *Member, IEEE*, and Cuntai Guan, *Senior Member, IEEE*

***Abstract*—Technology development in wearable sensors and biosignal processing has made it possible to detect human stress from the physiological features. However, the intersubject difference in stress responses presents a major challenge for reliable and accurate stress estimation. This research proposes a novel cluster-based analysis method to measure perceived stress using physiological signals, which accounts for the intersubject differences. The physiological data are collected when human subjects undergo a series of task-rest cycles, incurring varying levels of stress that is indicated by an index of the State Trait Anxiety Inventory. Next, a quantitative measurement of stress is developed by analyzing the physiological features in two steps: 1) a k-means clustering process to divide subjects into different categories (clusters), and 2) cluster-wise stress evaluation using the general regression neural network. Experimental results show a significant improvement in evaluation accuracy as compared to traditional methods without clustering. The proposed method is useful in developing intelligent, personalized products for human stress management.**

***Index Terms*—Clustering, physiological signal processing, stress evaluation.**

I. Introduction

ESTIMATION of event-based stress is a fundamental issue of stress management. As known in psycho-physiological studies, a person's physiological features are correlated with his/her stress levels [1]–[3]. Therefore, they can be used to construct an objective measurement of the stress. Prior work in stress measurement has been focusing on the collection and analysis of physiological data and the identification of the correlation between perceived stress and multiple physiological features [4], [5]. However, reliable and accurate estimation of stress levels using physiological data is challenging due to the complex, multivariate relationship between the stress and human physiology [6], [7]. This necessitates the identification of appropriate physiological features, and the development of reliable methods for predicting stress levels based on these features.

Among the many barriers to reliable stress evaluation, the individual subjects' differences in stress response (i.e., how a person's physiology changes in response to stressful events) is one that has not been effectively tackled [8], [9]. A common method to deal with this issue is to calibrate the biosignals against a set of baseline data, where a certain stress level (typically low or absence of stress) is induced in a stress-free environment [4], [9]. However, the usage of a baseline can be tricky depending on if the baseline is built for individual users or for a group of users. Collecting and utilizing the baseline data for an individual user may improve the accuracy of stress evaluation for that particular user, but not for the other users, whereas a baseline for the entire population may not be adequate to deal with intersubject differences. Some research introduces person-specific parameters to deal with variations in individual stress response [9]. However, such a strategy relies on the training of models with respect to individual users, which hampers its generalizability. Alternatively, to develop generic methods for pattern recognition, some studies adopt a feature selection process to enhance the predictive accuracy [10]–[12]. Nonetheless, the feature selection is usually carried out for the whole population. Therefore, they had limited power in personalized stress evaluation.

The objective of this research is to accurately evaluate human stress while accounting for the intersubject differences. Data are collected in an experimental environment where subjects underwent a series of task and rest cycles, incurring varying levels of physical and emotional stress. This study adopts the State-Trait Anxiety Inventory (STAI)-Form Y1 for annotating the event-based temporary stress [13]. Next, a two-stage procedure for stress measurement is designed for data analysis. 1) A cluster analysis is carried out using a subset of data samples, so that individual subjects are assigned to one of a few clusters. 2) Cluster-wise stress evaluation is performed using a general regression neural network (GRNN). The novelty of this research lies in the clustering process that assigns subjects into subgroups, so as to exploit the inherent homogeneity of subjects' stress response within the clusters (i.e., subjects within the same cluster are expected to share similar patterns of stress response). Thus, the intersubject differences are automatically accommodated, and the overall accuracy of the stress evaluation is improved. The performance of cluster-based analysis is evaluated using cross-validation methods. It is shown that the method is effective in dealing with the intersubject differences. To the best knowledge of the authors, this is the first attempt to design and evaluate a cluster-based analysis method for dealing with the intersubject difference in stress evaluation.

II. Related Work

Under stressful conditions, measureable changes can be found in human physiology, such as increased heart rate and blood pressure, reduced skin resistance, enlarged pupil size, etc. This

Manuscript received October 3, 2013; revised January 8, 2014; accepted March 1, 2014. Date of publication March 11, 2014; date of current version December 30, 2014. This work was supported in part by the Singapore A*STAR SERC TSRP grant on Human Factors Engineering (#0921530095) and Neural and Biomedical Technology Department, Institute for Infocomm Research.

The authors are with the Institute for Infocomm Research (A*STAR), Singapore (e-mail: qxu@i2r.a-star.edu.sg; tlnma@i2r.a-star.edu.sg; ctguan@i2r.a-star.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2311044

TABLE I
RELATED WORK

Reference	*Type of Stress*	*Stress Generation*	*Input Signal*	*Annotation*	*Data Processing Technique*	*Consideration of individual difference*
Shi et al. (2010) [9]	- Physical - Mental - Social	- Cold water - Arithmetic - Public speech	26 features in 6 categories: Heart rate; ECG; Respiration; Galvanic skin response (GSR); Body temperature	Questionnaire using ecological momentary assessment	Support vector machine (SVM)	A single parameter in SVM to account for person-specific information.
Jovanov et al. (2003) [16]	Physical	9D5 multi-place underwater egress trainer	Heart rate variability (HRV)	Salivary hormone response, cognitive and psychological functioning before, during, and after stressors.	[No information]	Not considered.
Healey & Picard (2005) [4]	Driving	On road driving with different road conditions.	22 features in 4 categories: EMG (9); Skin conductance (hand & foot) (4); Respiration (8); HRV (1): LF/HF or (LF+MF)/HF	Questionnaire and video coding	Fisher projection matrix and linear discriminant for classification.	Normalization against baseline measures.
Rani et al. (2002) [18]	Mental stress	Video game with varying difficulty	EKG (IBI), comparative magnitude of sympathetic vs. parasympathetic power spectrum	Probably self-assessment.	Wavelet packet decomposition; Fuzzy inferences	Subject variability observed for two subjects.
Rani et al. (2007) [2]	Anxiety	Anagram, math, sound discrimination.	18 features in 4 categories: ECG (7); Electrodermal activity (EDA) (5); EMG (4); Temperature (2)	Anxiety index (self-reported stress level in terms of anxiety, overload, calm)	Regression tree, fuzzy logic	Different features selected as predictors for individual participants. Personalized rules.
Plarre et al. (2011) [5]	- Physical - Mental - Social	- Cold water - Arithmetic - Pubic speech	4 features from ECG 7 features from respiration.	Self-reported stress rating.	Hidden Markov model.	Perceived stress (subjective) normalized considering accumulation and decay factors.
Faireclough & Venables (2006) [1]	Mental stress	Multi-attribute task battery (MATB) 4x20min	7 features in 4 categories. EEG (3); ECG (2); Skin conductance (foot) (1) Eye blink interval (EOG) (2) Respiration rate (1)	Dundee stress state questionnaire (DSSQ)	Multiple regression analysis (MANOVA)	Not considered.
Salahuddin and Kim (2006) [19]	Mental stress	Color stroop	6 features from HRV	Skin conductance and fingertip temperature used for validation	Non-parametric testing	Not considered.
Hasegawa, et al. (2001) [7]	Relaxation	Massage chair	HRV	Contents of glucose in blood	Maximum Lyapunov Exponent	Not considered.
Hirshfield et al. (2009) [20]	Working memory load	Working memory cognitive tasks	Functional near infrared spectroscopy (fNIR)	Oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (Hb).	ANOVA	Analysis for individual subjects. Normalized against each subject's baseline.
Pärkkä et al. (2009) [8]	Real life environment.	Real life environment.	High-level features extracted from wearable sensors and environment sensors	Bergen burnout indicator Derogates stress profile	Correlation analysis	Not considered.
Choi et al. (2012) [14]	Mental	Cognitive tasks and relaxed breathing	5 features EDA (3); HRV (2)	Self-reported stress level.	Generalized linear model	Not considered
Setz et al. (2009) [15]	Work related stress vs. cognitive load	Montreal imaging stress task for stress and cognitive load.	11 features from electrodermal activity (EDA)	Task-related.	Classification using: Linear discriminant analysis; SVM; - Nearest class center	Normalization against baseline feature.
Cinaz et al. (2013) [12]	Mental workload	Variants of Dual N-Back Task.	ECG: Time domain features (8); frequency domain features (3): LF, HF, LF/HF	NASA Task Load Index (TLX)	ANOVA, linear discriminant analysis, k-nearest neighbor, SVM	Calibration.

is the biological basis of stress measurement using physiological signals, which is an alternative way to stress evaluation based on subjective self-report. Along this line, many researchers develop the hardware and software solutions to stress measurement [1], [2], [4], [5], [14]–[16]. Hardware solutions are mainly concerned with the development of wearable and robust sensors for biodata collection. Software development focuses on the biosignal processing, i.e., to effectively extract meaningful patterns from the collected data. Along this line, the evaluation of the stress from the collected signal faces many challenges, such as the establishment of reliable "ground truth" of stress levels [5], [6], and the individual differences in stress responses [5]. To examine the prevalent solutions to these challenges, the study summarized related work in this field focusing on answers to the following questions (see Table I).

1) What type of stress does the study address?
2) How is the stress incurred?
3) How is the stress annotated?
4) What type of physiological signals is used as the predictors (i.e., the input feature)?
5) What is the data processing technique?
6) Does it consider the individual differences? If so, how?

As seen from the literature, while it has been widely recognized that individual difference is present in human's stress response, little has been done to mitigate such differences. A few studies resort to a calibration process against the baseline period [4], [5], [12], [15]. However, it makes the process complex and unpredictable especially outside a laboratory environment. A personalization parameter is used in the classification algorithm based on a support-vector machine [9]. However, the effectiveness of the method is not validated by performance evaluation. A self-organizing map is used in a stress study to form local clusters by preserving the topological relationships [17]. It is suggested that larger grids lead to better performance since they account for more patterns of stress responses. However, the validity and generalizability of the method are not examined because only three grid sizes are tested. In summary, existing work lacks effective mechanisms to tackle intersubject differences using a generic model for stress evaluation. Moreover, it usually involves a complicated calibration process that hampers its applications.

III. EXPERIMENTAL DESIGN

A. Experimental Procedure

A general experimental procedure includes three stages in which a subject carries out the following main activities: 1) initial rest (5 min), 2) task load (5 or 20 min), and 3) recovery

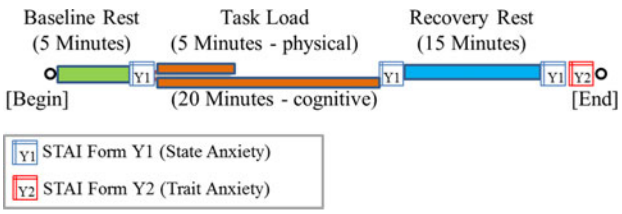

Fig. 1. Experimental procedure.

TABLE II
EQUIPMENT FOR THE EXPERIMENT

Signal	*Sensor*	*Sampling Rate*	*Placement*
EEG	MindBand™ - dual sensor headband	256Hz	2 electrodes contacting Fp1, Fp2.
ECG	Alive Heart & Activity Monitor	300Hz	2 electrodes attached to the chest.
EMG	Shimmer EMG sensor	512Hz	3 electrodes placed on the trapezius.
GSR	Shimmer GSR sensor	32Hz	2 electrodes on distal phalanges of index & ring fingers (non-dominant hand).

rest (15 min) (see Fig. 1). In the experiment, a participant went through one of two types of tasks, namely, physical or cognitive task. In the physical task, a subject did a squat-stand exercise continuously for 5 min at a pace that was considered challenging to him/her. In the cognitive task, a subject completed six computer-mediated tasks that involved different types of cognitive workload, such as attention, calculation, and memory. These tasks included go/no-go visual reaction time, color reading interference (stroop), fast counting, PASAT speed run, visual forward digit span, working memory test (n-back) (http://cognitivefun.net/). Each task lasted for 3–4 min, such that the total time of cognitive task was 20 min. During the task load period, the experimenter constantly challenged the subjects to achieve higher performance, so as to incur adequate stress.

The STAI form Y1 (state anxiety) was deployed immediately after the rest or task load period, where the order of the questionnaire items were randomized for each stage. Finally, the STAI form Y2 (trait anxiety) was deployed after the last Y1 form. A total of 44 male subjects were recruited, who were healthy adults without mental/emotional disorders, with no symptoms or history of cardiovascular or respiratory disorders. Subjects were randomly assigned to the physical- and cognitive-task groups, with equal number (22) in the each group. The mean age of the subjects was 28.6 years, with a standard deviation of 7.2 years.

B. Equipment Configuration

A set of noninvasive, wireless sensors was used to collect a subject's physiological signals including, electrocardiography (ECG), galvanic skin response (GSR), electroencephalography (EEG), electromyography (EMG), and saturation of peripheral oxygen (SpO2). A summary of the equipment configuration is shown in Table II. The sensors were connected to a PC via Bluetooth and data from different channels was synchronized. All events related to different stages of the experimental procedure were labeled using time stamps. Physiological features that potentially correlate with physical/mental stress were extracted from the raw data collected by the sensors.

IV. Data Analysis

A. Stress Indices

The subjects' stress levels were derived from the STAI-Y1 scores. Cronbach's alpha for the three experiment stages are 0.923 (baseline), 0.899 (task load), and 0.904 (treatment), showing high internal consistency of the questionnaire items. Next, the scores are checked for potential outliers by applying the following exclusion criterion. Let S_{ij} $(i = 1, 2, \quad , N;\ j =$ 1,2,3$)$ denote the STAI-Y1 scores of subject i in three stages, namely, baseline $(j = 1)$, task load $(j = 2)$, and treatment $(j = 3)$; N is the number of subjects. A record i is excluded if $S_{i2} < S_{i1}$ and $S_{i2} < S_{i3}$. This is because a subject is expected to feel more stressful after the task load (S_{i2}) than after the initial rest (S_{i1}) and recovery (S_{i3}) periods. Three (3) cases were excluded by applying this criterion. Next, the normalized stress indices (s_{ij}) for subject i is computed as

$$s_{ij} = \frac{S_{ij} - \min(S)}{\max(S) - \min(S)} \tag{1}$$

where $\max(S)$ and $\min(S)$ refer to the largest and smallest STAI scores of all subjects in three stages. Thus, the stress indices are within the range of [0, 1].

B. Feature Extraction

Prior work has shown that a person's perceived stress is correlated with various physiological features. In this research, 15 features were extracted for analysis.

1) Mean and standard deviation of GSR [4], [15], [21].
2) Mean SpO2.
3) Mean and standard deviation of EMG [4], [22], where the raw EMG data is full-wave rectified, averaged with a time interval of 0.1 s, and filtered using a band-pass of 5–300 Hz.
4) Mean heart rate (HR) and heart rate variability (HRV) [2], [5], [23]. First, the interbeat interval was computed from the raw ECG data. Next, five features were extracted using the Kubios HRV (http://kubios.uku.fi/): 1) mean HR, 2) pNN50–the portion of normal sinus intervals exceeding 50 ms, 3) LF power (the power of the low frequency band 0.04–0.15 Hz), 4) HF power (the power of the high frequency band 0.15–0.4 Hz), 5) LF/HF, which is the ratio of power in the low- and high-frequency bands.
5) Mean spectral power of EEG in four bands. These are 1) δ power (1–4 Hz), 2) Θ power (4–8 Hz), 3) α power (8–13 Hz), and 4) β power (13–20 Hz). In addition, the ratio of Θ and α is computed (Θ /α) [1]. To cancel out the effect of ocular artifacts in the raw EEG signal, this research adopted the wavelet transformation method [24].

Before computing these features, the raw signals were normalized against the baseline mean values of the respective features. Next, these 15 features were computed with respect to

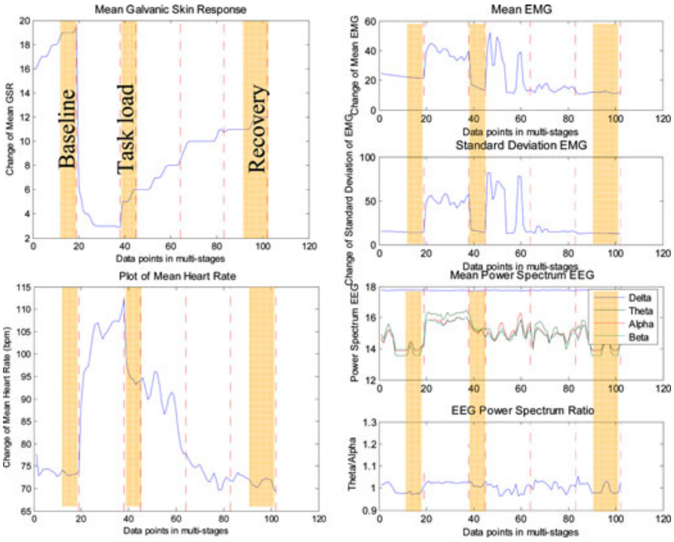

Fig. 2. Sample features (GSR, EMG, HR, EEG) for a subject. Shaded areas are time period from which the features are used for data analysis.

three experiment stages, namely: 1) the last 3 min of baseline rest, 2) the last 3 min of the task load period, and 3) the last 5 min of recovery rest. Presumably, the first data segment characterized a baseline low stress state. The second segment was related to a high stress state, and the third segment corresponded to a low stress state, i.e., recovery from a high stress level. For each data segment, the signal was divided into frames of 30 s with 50% overlap. Each frame was multiplied by a Hamming window to minimize signal discontinuities at the end of each frame [21].

For each subject, there were 41 frames (baseline: 11, task load: 11, and recovery: 19). Finally, for each frame of the feature, it was labeled with the STAI stress index based on the activity stage it belonged to (i.e., baseline, task load, and recovery rest). Fig. 2 illustrates some of the features used for stress estimation. Due to broken data during recording, two subjects were excluded for further analysis. Therefore, 39 subjects were included in subsequent analysis —21 belonging to the physical task group, and 18 the cognitive task group.

C. Cluster-based Analysis

The idea of cluster-based analysis originates from the observation that different physiological features have varying predictive power in stress evaluation for individual subjects, and that a few subjects may share certain similarity in their stress response. For example, some subjects may have identical HRV change patterns, while others may exhibit similar GSR change patterns. It follows that HRV might be a good predictor for the former, and GSR for the latter. If a single model is built using the cohort data of all subjects, which was the strategy adopted by most legacy methods, an implicit assumption is that all the subjects have homogeneous stress responses. Alternatively, if a model is built for an individual subject, the stress prediction might be superior for that particular subject, but poor for the other subjects. In addition, when individual models are used, one has to build many models for each and every subject. Accordingly, it would be difficult if not impossible to select an appropriate model for unseen datasets, thus jeopardizing its predictive power.

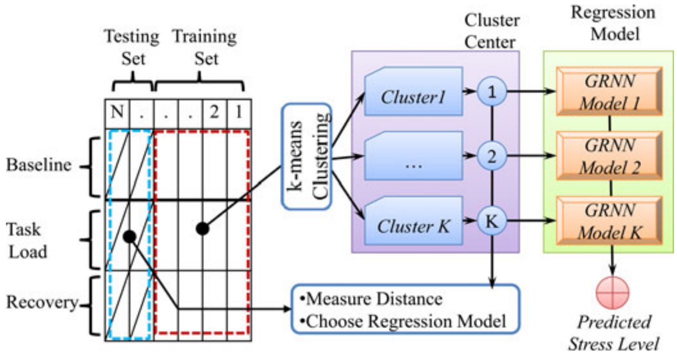

Fig. 3. Principle of cluster-based analysis.

To alleviate such a problem, a two-stage procedure is proposed in this research, where a clustering process is used to divide subjects into a set of subgroups. The principle of the cluster-based method is shown in Fig. 3. First, the collected feature data are divided into the training set and the testing set. The former is used to build the model and the latter is used to test the performance. For a training dataset, the feature data of the baseline stage is used to do the clustering. Let the baseline feature vector be denoted as $\mathbf{f}_{(i\ j)} = [f^1_{(i\ j)}\ f^2_{(i\ j)}\ \ \ f^n_{(i\ j)}]$, where $i(\in 1\ 2\ \ \ N))$ is the index of a subject, $j(\in 1\ 2\ \ \ m)$ is the frame index ($m = 11$ in this case), and n is the number of features ($n = 15$). The feature center of subject i is computed as

$$\vec{\mathbf{f}}^{\,C}_i = \left[\mathbf{f}^1_i\ \mathbf{f}^2_i\ \ \ \mathbf{f}^n_i\right] = \left[\frac{\sum_{j=1}^m f^1_{(i\ j)}}{m}\ \frac{\sum_{j=1}^m f^2_{(i\ j)}}{m}\ \ \ \frac{\sum_{j=1}^m f^n_{(i\ j)}}{m}\right]^C_i \quad (2)$$

Next, a k-means clustering is carried out so that the subjects are assigned to a subgroup based on the feature centers of the respective subjects ($\vec{\mathbf{f}}^{\,C}_1\ \vec{\mathbf{f}}^{\,C}_2\ \ \ \vec{\mathbf{f}}^{\,C}_N$), i.e., the feature center is used as the coordinate of a subject. A subject i is assigned to one of K clusters. Assuming a set of subjects $i' \in (1\ 2\ \ \ N_k)$ belongs to the cluster k, the center of cluster k is computed as

$$C_k = \left[\frac{\sum_{i'=1}^{N_k} \mathbf{f}^1_{i'}}{N_k}\ \frac{\sum_{i'=1}^{N_k} \mathbf{f}^2_{i'}}{N_k}\ \ \ \frac{\sum_{i'=1}^{N_k} \mathbf{f}^n_{i'}}{N_k}\right] \quad (3)$$

In the second stage, a regression analysis is carried out using the GRNN with respect to individual clusters based on the training dataset belonging to the task load and recovery stages. This leads to a set of K GRNN models that minimize the cluster-wise error.

Once the GRNN models are built for all clusters, the stress evaluation for a new subject is carried out as follows. The testing dataset is considered as a dataset of new subjects. The input to the model is a dataset consisting of a set of baseline features, and nonbaseline features. The baseline feature center is computed using eq. (2). The Euclidian distances between the feature

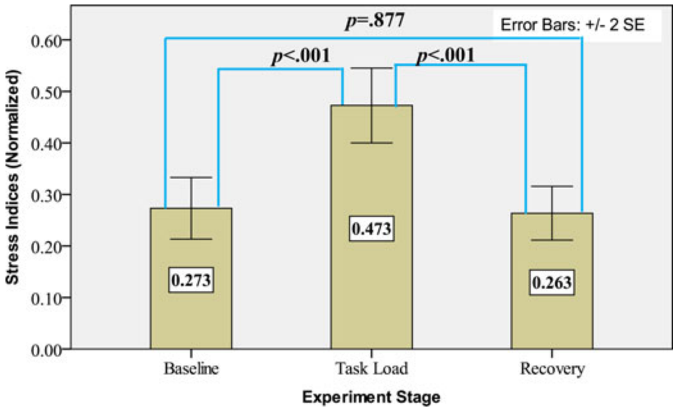

Fig. 4. Stress indices in three stages.

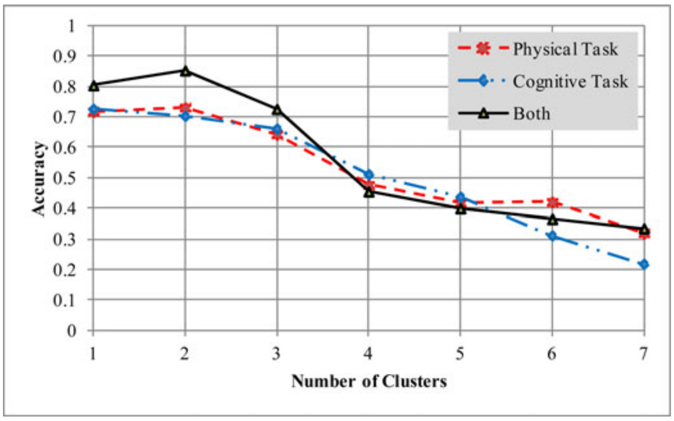

Fig. 5. Sensitivity analysis—task type and the number of clusters.

center and all K cluster centers are computed. The subject is assigned to the cluster of minimum distance. In so doing, the corresponding GRNN model is triggered to evaluate the stress level of the nonbaseline features. The subject's stress level can be predicted using the selected GRNN model. The system performance is computed as the difference between the predicated stress level and the stress indices captured by the normalized STAI score.

V. Results

A. Perceived Stress Level

The perceived stress effectively reflects the change of stress due to task load and recovery rest (see Fig. 4). The paired t-test showed a stress level after the task load stage was significantly higher than after the baseline stage ($t = 3.782$; $p < .001$) and recovery stage ($t = 3.989$; $p < .001$). The mean stress indices for the baseline and recovery stages were identical with $t = 0.155$ and $p = .877$. The mean trait anxiety was examined for the physical and cognitive task load groups. No significant difference was found in the trait anxiety ($t = .809$ and $p = .420$). This is desirable because it ruled out the influence of long-term stress on the subjects' responses.

B. Sensitivity Analysis

Since the number of clusters may influence the performance of the stress evaluation, it is necessary to compare the system performance with respect to the cluster numbers. The cluster number that achieved optimal overall performance is used in the final model. Sensitivity analysis is carried out by varying the number of clusters. Note that when the number of cluster is "1," it means that all training data are assigned to the same group, so that it is equivalent to doing a GRNN over all data frames without clustering. Moreover, the procedure is repeated with respect to three datasets differentiated by the task type: 1) the physical task (21 subjects), 2) the cognitive task (18), and 3) both tasks (39). The population-wise performance is tested using a leave-one-out cross validation. That is, for each training-testing round, feature data belonging to one subject was used for testing, and the rest was used for training.

Fig. 5 shows the accuracy of stress evaluation with respect to the three datasets and different number of clusters. Let R be the mean difference between the predicted stress level and the normalized stress indices associated with the feature vectors. The accuracy was computed as $1-R$. As seen from the result, the optimal performance was 0.852, which was achieved when all subjects (both task groups) were included for training, and when there were 2 clusters. The performance was reasonably good if only 1 cluster was used. This was especially true when data belonging to a single-task type was used, such that the performance of 1-cluster and 2-cluster was identical. This can be attributed to a more powerful GRNN model that enjoyed a larger dataset for training (Note that the GRNN models were built cluster-wise). When there were more than 3 clusters, the performance plummeted, irrespective of the task type.

C. Effect of Clustering

Although the overall performance was improved due to the clustering procedure, it is necessary to find out whether such an improvement was facilitated by the homogeneity of subjects within the clusters, which in turn contributes to better predictive power of the cluster-wise GRNN model. Note that in the testing stage, only one GRNN model was selected depending on which cluster center was closest to the subject's baseline feature center. It was expected that the performance of the selected GRNN model was better than the other GRNN models (referred to as alternative models). To verify this, the accuracy of the selected GRNN model was compared against the alternative models. In this study, only the 2-cluster case was investigated. The testing process was designed so that both GRNN models were used to predict the stress level of the testing dataset.

When all data were used (both physical and cognitive tasks), the mean accuracy of the selected model was 0.852 (std $=$ .08) and that of the alternative model was 0.732 (std $=$ .11). The paired t-test showed significant difference in the accuracy of the selected model and alternative model ($t = 29\ 4$; $p < .001$). Moreover, the test procedure was carried out for data belonging to individual task types (i.e., physical or cognitive task); it was observed that the selected model consistently outperformed the alternative model with $p < .001$. Therefore, the clustering

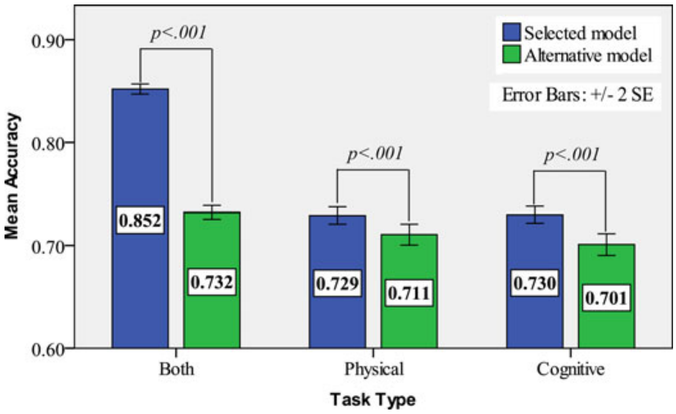

Fig. 6. Performance comparison of the selected model versus alternative model.

process indeed contributed to better system performance. The comparative result is shown in Fig. 6.

VI. Discussions

The proposed method was validated under conditions of both physical and cognitive stressors. To do so, the study included separated analysis for the physical and cognitive task conditions, as well as the combined conditions (see Fig. 5). Due to the small sample size, the significance of the method in the separated task conditions was not as obvious as in the combined condition. This is because cluster analysis improves system performance at the expense of reducing the predictive power of individual GRNN models, i.e., these models are built on a smaller dataset than if all samples are assigned into the same cluster. Therefore, one needs to ensure that sufficient samples are collected for each cluster. In the study, it is seen that despite the reduced sample sizes in the separated conditions, the clustering does not result in worse performance *per se*.

The performance of the model is dependent on the number of clusters. Thus, sensitivity analysis is needed to determine an appropriate number of clusters. Based on the analysis, it was shown that the number of clusters was typically small (in this case 2 clusters) in order to achieve optimal performance. So, it is advisable to test the system performance by restricting the number of clusters to a small value. However, the current result was based on a small dataset. For a large dataset with sufficient samples, one may test out more cluster numbers.

One limitation of the study is the naive strategy of clustering. The baseline features used for the clustering were collected when subjects were in a relaxed state. An alternative strategy is to trigger certain stress change in the baseline and use the change of physiological features to do the clustering. Nevertheless, this research proposes to use the relaxed state baseline data for clustering because it greatly simplifies the calibration process. In fact, it would be tedious and costly to include an additional stress-inducing period that is difficult to be replicated outside a laboratory. In this sense, the proposed method has greater application potentials.

Another limitation is related to the control of task difficulty for inducing stress. It is suspected that personal characteristics, such as emotional intelligence affect a person's endurance to stressors. In this study, this is partially alleviated by controlling the task difficulty during the experiment to suit the subjects' potentials.

Finally, for feature extraction, this study used 30-s window, which is a bit narrow especially for computing the low-frequency power of HRV. The choice of the time window has considered 1) the limited length of high-quality data for all task stages, 2) the desire to have more data points within the period of data collection, and 3) extraction of meaningful features for all signal channels. Despite the limitation, in case the low-frequency HRV power is unstable and unreliable, the GRNN model will reduce its weightage. Therefore, it will not have a significant effect on the final result.

VII. Conclusion

This paper presents a stress evaluation method that employs a novel clustering procedure to accommodate the individual differences. Effective evaluation of the stress level is a useful step toward computational intelligence where computers can "understand" users in an unobtrusive way. Unlike traditional methods that either built a single model using the cohort data from all subjects, or constructed specific models for individual subjects, the proposed method exploits the homogeneity of subjects in the change of their physiological features due to stress. The effectiveness of the system is examined through the sensitivity analysis. It is found that a small number of clusters represented a good balance between within-cluster homogeneity and between-cluster heterogeneity. The study conducted empirical validation on the effect of clustering—the model selection does improve the system performance. The method paves the way toward building customized models for sensor-based better human stress evaluation.

References

[1] S. H. Fairclough and L. Venables, "Prediction of subjective states from psychophysiology: A multivariate approach," *Biol. Psychol.*, vol. 71, no. 1, pp. 100–110, 2006.

[2] P. Rani, N. Sarkar, and J. Adams, "Anxiety-based affective communication for implicit human-machine interaction," *Adv. Eng. Inform.*, vol. 21, no. 3, pp. 323–334, 2007.

[3] S. D. Kreibig, "Automatic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.

[4] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[5] K. Plarre, A. Raij, S. M. Hossain, A. A. Ali, M. Nakajima, M. al' Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. E. Wittmers, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proc. Int. Conf. Information Process. Sens. Netw.*, Chicago, IL, USA, 2011, pp. 97–108.

[6] S. H. Fairclough, "Fundamentals of physiological computing," *Interact. Comput.*, vol. 21, no. 1–2, pp. 133–145, 2009.

[7] Y. Hasegawa, T. Ootsuka, T. Fukuda, F. Arai, and M. Kawaguchi, "A relaxation system adapting to user's condition-identification of relationship between massage intensity and heart rate variability," in *Proc. IEEE Int. Conf. Robot.*, Seoul, Korea, 2001, pp. 3195–3200.

[8] J. Pärkkä, J. Merilahti, E. M. Mattila, E. Malm, K. Antila, M. T. Tuomisto, A. V. Saarinen, M. van Gils, and I. Korhonen, "Relationship of psychological and physiological variables in long-term self-monitored data during work ability rehabilitation program," *IEEE Trans. Inf. Technol. B.*, vol. 13, no. 2, pp. 141–151, Mar. 2009.

[9] Y. Shi, M. H. Nguyen, P. Blitz, B. French, and S. Fish, "Personalized stress detection from physiological measurements," presented at the Int. Symp. Quality Life Tech., Las Vegas, NV, USA, 2010.

[10] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[11] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.

[12] B. Cinaz, B. Arnrich, R. La Marca, and G. Tröster, "Monitoring of mental workload levels during an everyday life office-work scenario," *Pers. Ubiquit. Comput.*, vol. 17, no. 2, pp. 229–239, 2013.

[13] C. D. Spielberger, *State-Trait Anxiety Inventory for Adults*, 2nd ed. Palo Alto, CA, USA: Consulting Psychologists Press, 1989.

[14] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and evaluation of an ambulatory stress monitor based on wearable sensors," *IEEE Trans. Inf. Technol. B.*, vol. 16, no. 2, pp. 279–286, Mar. 2012.

[15] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable eda device," *IEEE Trans. Inf. Technol. B.*, vol. 14, no. 2, pp. 410–417, Mar. 2010.

[16] E. Jovanov, A. O. Lords, D. Raskovic, P. G. Cox, R. Adhami, and F. Andrasik, "Stress monitoring using a distributed wireless intelligent sensor system,," *IEEE Eng. Med. Biol.*, vol. 22, no. 3, pp. 49–55, May/Jun. 2003.

[17] B. Arnrich, C. Setz, R. La Marca, G. Tröster, and U. Ehlert, "What does your chair know about your stress level?," *IEEE Trans Inf. Technol. B.*, vol. 14, no. 2, pp. 207–214, Mar. 2010.

[18] P. Rani, J. Sims, R. Brackin, and N. Sarkar, "Online stress detection using psychophysiological signal for implicit human-robot cooperation," *Robotica*, vol. 20, no. 6, pp. 673–686, 2002.

[19] L. Salahuddin and D. Kim, "Detection of acute stress by heart rate variability using a prototype mobile ecg sensor," in *Proc. 2006 Int. Conf. Hybrid Inf. Tech.*, 2006, pp. 453–459.

[20] L. M. Hirshfield, E. T. Solovey, A. Girouard, A. Kebinger, R. J. K. Jacob, A. Sassaroli, and S. Fantini, "Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy," in *Proc. 27th Int. Conf. Human Factors Comput. Syst.*, Boston, MA, USA, 2009, pp. 2185–2194.

[21] M.-Z. Poh, N. C. Swenson, and R. W. Picard, "A wearable sensor for unobtrusive, long-term assessment of electrodermal activity," *IEEE Trans. Inf. Technol. B.*, vol. 57, no. 5, pp. 1243–1252, May 2010.

[22] U. Lundberg, R. Kadefors, B. Melin, G. Palmerud, P. Hassmén, M. Engström, and I. Dohns, "Psychophysilogical stress and emg activity of the trapezius muscle," *Int. J. Behav. Med.*, vol. 1, no. 4, pp. 354–370, 1994.

[23] R. L. Mandryk and M. S. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies," *Int. J. Human–Comput. Studies*, vol. 65, no. 4, pp. 329–347, 2007.

[24] V. Krishnaveni, S. Jayaraman, S. Aravind, V. Hariharasudhan, and K. Ramadoss, "Automatic identification and removal of ocular artifacts from EEG using wavelet transform," *Meas. Sci. Rev.*, vol. 6.2, no. 4, pp. 45–57, 2006.

Qianli Xu (M'07) received the B.E. and M.E. degrees from Tianjin University, Tianjin, China, in 1999 and 2002, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2007.

He is a scientist with the Visual Computing Department, Institute for Infocomm Research, Singapore. His research interests include design for human factors, healthcare systems, intelligent products and services. His publications appear in IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A:, *ACM SIGCHI*, *Design Studies*, etc.

Tin Lay Nwe (M'13) received the B.E. degree in electronics from the Yangon Institute of Technology, Myanmar, in 1994, and the Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2004.

She is currently a Scientist with Human Language Technology Department, Institute for Infocomm Research, Singapore. Her research interests include robust speech recognition, paralinguistic information processing of speech, music information retrieval, and rich transcription.

Cuntai Guan (S'91–M'92–SM'03) received the Ph.D. degree in electrical and electronic engineering from Southeast University, Nanjing, China, in 1993.

He is the Principal Scientist and Department Head at the Institute for Infocomm Research, A∗STAR, Singapore. He is the A∗STAR MedTech Programme Leader of neuro-technology. His research interests include neural and biomedical signal processing, machine learning and pattern recognition, neural and cognitive process and its clinical applications, and brain–computer interfaces.

Dr. Guan is an Associate Editor of IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING (2010–2013), IEEE ACCESS (2012–2015), *Journal of Brain Computer Interface*, *Australasian Medical Journal*, *Frontiers in Neuroprosthetics*, and *A∗STAR Research Publication*.