

Recognizing Emotions in Human Computer Interaction: Studying Stress Using Skin Conductance

Alexandros Liapis¹(✉), Christos Katsanos^{1,2}, Dimitris Sotiropoulos¹,
Michalis Xenos¹, and Nikos Karousos^{1,2}

¹ School of Science and Technology, Hellenic Open University,
Parodos Aristotelous 18, 26 335 Patra, Greece

{aliapis, ckatsanos, dgs, xenos, karousos}@eap.gr

² Technological Educational Institute of Western Greece,
M. Alexandrou 1, 26 334 Patra, Greece

Abstract. This paper reports an experiment for stress recognition in human-computer interaction. Thirty-one healthy participants performed five stressful HCI tasks and their skin conductance signals were monitored. The selected tasks were most frequently listed as stressful by 15 typical computer users who were involved in pre-experiment interviews asking them to identify stressful cases of computer interaction. The collected skin conductance signals were analyzed using seven popular machine learning classifiers. The best stress recognition accuracy was achieved by the cubic support vector machine classifier both per task (on average 90.8 %) and for all tasks (Mean = 98.8 %, SD = 0.6 %). This very high accuracy demonstrates the potentials of using physiological signals for stress recognition in the context of typical HCI tasks. In addition, the results allow us to move on a first integration of the specific stress recognition mechanism in PhysiOBS, a previously-proposed software tool that supports researchers and practitioners in user emotional experience evaluation.

Keywords: Users emotional experience evaluation · Physiological data · Skin response conductance · Physiological signal analysis

1 Introduction

Evaluation of user emotional experience (UEX) is a topic with growing significance. Beyond traditional evaluation methods (e.g. questionnaires, interviews and observation etc.), the study of physiological signals has become increasingly important in human-computer interaction. Associations among emotions and physiological signals [1] have established innovative evaluation approaches [2, 3] which offer to researchers and practitioners new insights in UEX evaluation.

So far, existing methods for emotions induction rely on intense stimuli such as scary movie clips, favorite songs, major hardware/software failures, image datasets and gaming [4–7]. Such stimuli induce intense reactions, which may be depicted in facial expressions, body postures and physiological signals, and recognized by existing

associated methods. However, recognition of emotions from subtle events [8], which are typically expected in most HCI tasks, remains challenging.

According to Lazar [9] the goal of an evaluation process is to identify system flaws which are often associated with negative emotions such as “stress” [10]. Thus, recognizing stress in typical HCI tasks is particularly important, and it is the object of this paper. Research shows [11, 12] that skin conductance, also known as Galvanic Skin Response (GSR) or Electro Dermal Response (EDR), is a reliable indicator of stress. Skin conductance is the physiological signal that was also selected and measured in this paper. To this end, 31 healthy participants performed five carefully selected stressful HCI tasks, and their skin conductance signals were monitored and analyzed using seven popular machine learning classifiers.

The purpose of this paper is twofold. First, it presents results from the first set of experiments aiming to create a publicly available dataset of physiological signals, which can be used for stress recognition in HCI. To the best of our knowledge, this is the first experimental approach in stress recognition that exclusively uses typical HCI tasks as stimuli. Second, the paper aims to investigate the performance of various algorithms in identifying stress from skin conductance. The obtained recognition results are going to guide the implementation of an automated stress identification algorithm in PhysiOBS, our previously-proposed software tool [13] aiming to support researchers and practitioners in UEX evaluation.

The rest of the paper is structured as follows. Section 2 presents the research-based approach followed for stimuli selection. In Sect. 3 the experimental general set-up and protocol, are described. Section 4 presents the used preprocessing techniques and recognition algorithms, along with their results. The paper concludes with a discussion of the implications of the presented work and directions for future research.

2 Research-Based Stimuli Selection (Stressors)

Eliciting emotions in a laboratory setting is challenging and needs a careful design. The appropriate stimuli should be plausible enough in order to induce a heightened level of physiological arousal. In addition, any stimuli selection method should be void of any bias introduced by researchers.

Stimuli selection process involved fifteen typical computer users (University employees, students, and colleagues) which participated in a face to face interview. Interviewees were asked to identify stressful tasks during interaction with a computer. All interviews were conducted in two phases by the same person. Each phase lasted from 15 to 20 min. First, demographics (e.g. age, skills in computer usage, profession, education etc.) were recorded. Next, participants were asked to describe at least five scenarios which stress them while interacting with a computer. Interviewees were neither informed nor participated in the stress monitoring experiment.

All the scenarios provided by the interviewees did not require any special experience or knowledge. Participants’ answers were grouped and a frequency table was created. Answers analysis did not reveal any significant differences due to demographic

parameters. Next, we pilot-tested the scenarios, starting from the most frequently mentioned. Although interaction scenarios related to financial transactions and viruses were commonly reported by interviewees, such tasks were not selected due to their requirements for being plausible enough to induce stress. For instance, a wrong charge in facilitators' credit card was not found to be stressful. In the end, the five most commonly reported scenarios were selected, excluding the aforementioned cases.

2.1 Scenario 1: Missing a File

Participants were asked to visit the website of the internal evaluation unit of the Hellenic Open University (<http://meae.eap.gr>). This website was selected because it was expected to be unfamiliar to participants. Next, they were asked to find and download a specific file from the website, save it to a network folder and log in a google email account to send the file at an email address. When participants shifted their attention from the network folder in order to create the email, experiment facilitators remotely deleted participants' downloaded file.

2.2 Scenario 2: Hardware Problems

Participants were asked to visit the website of a research group in our University (<http://quality.eap.gr>). Again, this specific website was selected in order to avoid any previous familiarity. Next, they were asked to find and copy the consortium list from one of the team's projects and then paste it in a text file. During the task, their mouse cursor speed was set in slow speed. The speed was remotely set using a custom-made software tool that had been previously installed in the testing computer.

2.3 Scenario 3: Slow Network Speed

In this scenario, participants visited a web portal that is popular in our country (<http://www.in.gr>) and were asked to find information about a specific movie. During the task, network connection was simulated at 56 Kbps in order to make interaction slower than the usual. The speed was manipulated through the Fiddler (<http://www.telerik.com/>) software.

2.4 Scenario 4: Web Advertisements (Popups)

Participants were asked to visit a popular online booking website (<http://www.booking.com>) in order to make a reservation for a predefined destination. Appropriately designed popup windows appeared in users' screen every 15 s while they were trying to complete the scenario. The popup window was relevant to both the website's content and visual appearance. The whole process was controlled remotely through a custom-made software tool that had been previously installed in the testing computer.

2.5 Scenario 5: Finding Information in Websites

Participants were instructed to visit the website of our University's library (<http://lib.eap.gr>) in order to find the authors of a specific book. In this scenario, no external action was applied. This website was chosen for this scenario because there was a plethora of complaints about its information architecture, which had been collected in a previous usability evaluation study.

3 Experiment

3.1 Setting and Equipment

The experiment was performed in our fully-equipped usability lab (<http://quality.eap.gr>). Skin conductance was recorded at 5 Hz using a Mindfield eSense sensor. Stimuli scenarios were presented randomly for each participant through the Tobii eye-tracker environment (i.e. Tobii Studio) which was also used to monitor participants' eye activity in real time (e.g. to delete participants' downloaded file in the first scenario while they were not looking at it). All scenarios were designed to require minimum typing effort in order to minimize participants' hand movements that may affect skin conductance measurements. Finally, external parameters such as testing room temperature were controlled in order to avoid noise in skin conductance recordings.

3.2 Process and Protocol

Thirty-one healthy participants (18 female), aged between 21 and 38 (Mean = 30.8, SD = 4.7) were recruited. The experiment lasted for six days.

First, participants were informed that they would interact with some websites in order to perform some tasks. Subsequently, they completed an appropriate consent form along with a questionnaire about demographic information. Next, the skin conductance sensor was placed on participants' non dominant hand in the middle and ring finger respectively. A short time of approximately five minutes was given to participants in order to familiarize with the sensor, while signals' transmission quality was checked. In addition, participants' body posture in front of the eye-tracker was also checked. During this short time, the facilitators were available to answer in any of the participants' question.

The experimental process started with a 1:30 min baseline recording [6, 14], during which participants were asked to relax. Subsequently, the five stress-inducing scenarios were presented to participants in a random order. At the end of each scenario, participants were asked to provide subjective ratings of their emotional experience both on a valence-arousal [15] and on a 1–7 rating scale; however analysis of these ratings is beyond the scope of this paper. Each session lasted approximately 40 min per participant including short breaks between scenarios. Skin conductance was not monitored during the breaks or the self-assessment process.

4 Analysis and Results

In this section, signal preprocessing and classification results are presented. All in all, 182 skin conductance signals were recorded from 31 participants involved in five interaction tasks and a baseline condition. In four cases (once in task 1, once in task 2 and twice in task 4), signal was not recorded successfully due to sensor malfunction or experimenter error.

The collected signals were smoothed using hanning window function. Smoothing window width for each signal was determined by experimentally adjusting the following root mean square error function:

$$\text{Error} = \text{SQRT}(\sum (X_i - X_{i-1})^2) / (2 * N), \tag{1}$$

where Σ calculates the sum of first difference between sample values (X_i and X_{i-1}), and N is the total number of samples. This error value represents the signal’s variability due to sampling rate frequency.

The smoothing process involved the following steps. First, an initial error value was calculated for each raw signal. Next, raw signals were smoothed using a five-point width hanning window, and the error value was recalculated. While the error correction value between raw and smoothed signal was below 76 %, the width of the hanning window was increased by five points and the raw signal was smoothed again. Some signals had to reach a window width value of 100 points or more to meet this error correction percentage, resulting in substantial signal degeneration. Thus, they were set to be auto-excluded from the feature extraction.

The smoothing window step was selected to be equal to the sampling rate (5 Hz). The error correction threshold was set to 76 % based on two criteria: (a) keep signals’ crucial information, such as lows and peaks; Fig. 1 illustrates an instance of 200 samples (40 s) from a participant’s skin conductance signal for 76 % and 90 % error correction, and (b) use the signals’ majority in feature extraction; Fig. 2 illustrates that as the correction error gets higher than 76 %, significantly more signals are auto-excluded from the feature extraction process due to signal degeneration.

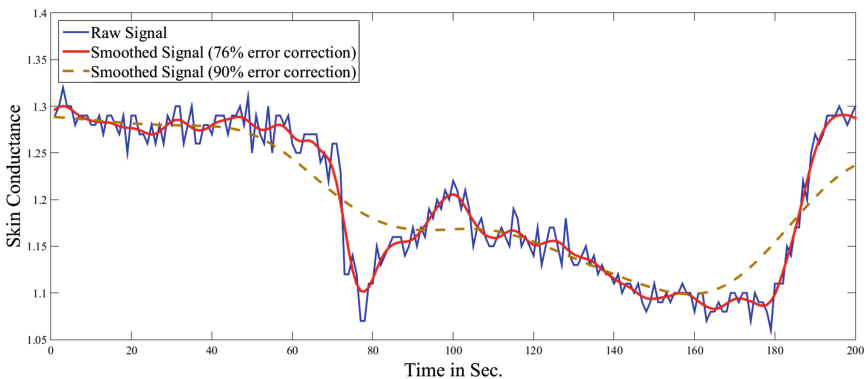


Fig. 1. Raw vs smoothed signal for 76 % and 90 % error correction.

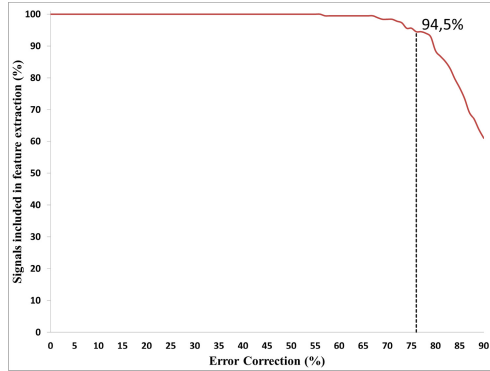


Fig. 2. Signals included in feature selection as a function of error correction.

After signal smoothing, 21 statistical features (e.g., mean, median, min, max, standard deviation, minRatio and maxRatio) [11] were extracted. The same statistics were extracted from the first and the second differences of signal.

The extracted features were used to train seven classifiers offered in the MATLAB R2015a Statistics and Machine Learning Toolbox v10.0: (a) Linear Discriminant Analysis (LDA), (b) Quadratic Discriminant Analysis (QDA), (c) Simple Decision Tree (S-Tree), (d) Linear Support Vector Machine (L-SVM), (e) Quadratic Support Vector Machine (Q-SVM), (f) Cubic Support Vector Machine (C-SVM), and (g) k-Nearest Neighbors (k-NN).

Table 1 presents classifier accuracies (%) for stress identification per task and for all tasks, using 100-times 10-fold cross validation for all tasks. C-SVM classifier had the best stress recognition accuracy both per task (Min = 89.6 %, Max = 91.6 %) and for all tasks (Mean = 98.8 %, SD = 0.6 %).

Table 1. Classifier accuracies (%) for stress identification per task and for all tasks. The last column presents results for the aggregated dataset of tasks, and not the cross-task mean.

	Task 1 Mean ± SD	Task 2 Mean ± SD	Task 3 Mean ± SD	Task 4 Mean ± SD	Task 5 Mean ± SD	All tasks Mean ± SD
LDA	90.5 ± 2.4	88.8 ± 2.7	88.8 ± 2.7	87.7 ± 2.0	88.0 ± 2.2	95.5 ± 0.4
QDA	83.8 ± 3.1	85.5 ± 3.6	83.0 ± 3.3	82.7 ± 3.9	82.5 ± 3.6	95.6 ± 1.0
S-Tree	88.6 ± 0.9	86.8 ± 1.7	87.9 ± 1.4	88.6 ± 1.5	86.4 ± 3.2	95.3 ± 0.9
L-SVM	88.4 ± 1.9	88.4 ± 1.3	88.2 ± 1.9	88.0 ± 1.2	89.3 ± 2.2	96.9 ± 0.4
Q-SVM	84.8 ± 3.0	84.5 ± 2.1	83.8 ± 3.6	83.6 ± 1.8	84.8 ± 1.5	95.2 ± 0.9
C-SVM	91.6 ± 2.2	89.6 ± 2.2	91.1 ± 1.7	90.5 ± 2.2	91.1 ± 2.2	98.8 ± 0.6
k-NN	90.2 ± 2.3	88.9 ± 2.5	88.9 ± 2.8	87.9 ± 3.1	89.8 ± 1.5	97.3 ± 0.8

5 Conclusion and Future Work

In this work a physiological dataset from 31 healthy participants involved in five stressful tasks and a baseline relax condition was created. A research-based approach was followed to produce the selected tasks. First, 15 typical computer users, not involved in the stress monitoring experiment, were asked to describe at least five stressful interaction experiences. Then, the ones mentioned most frequently were pilot-tested and five were selected for the stress monitoring experiment. The collected skin conductance signals were first preprocessed and then used to train seven popular machine learning classifiers to automatically detect the two emotional classes (stress – no stress) from skin conductance.

Results showed high identification accuracies, with the best being the one achieved by the Cubic Support Vector Machine (C-SVM) both per task (on average 90.8 %) and for all tasks (Mean = 98.8 %, SD = 0.6 %). This is an important finding that demonstrates the potentials of physiological signals in the study of subtle interaction events, which are typically expected in most HCI tasks, such as finding information in complex websites or being distracted by web advertisements while making an online booking. Our work makes a contribution towards this direction. In addition, the results allow us to move on a first integration of the specific automated stress recognition mechanism in PhysiOBS, our previously-proposed software tool [13] that supports continuous and multiple emotional states analysis by user experience practitioners.

One of our future aims is to replicate our findings by performing additional experiments following the same methodology using more peripheral physiological signals, such as blood volume pressure, respiration and temperature. In this way, we will also extend our emotionally-labeled dataset for stress recognition in typical HCI tasks, which we plan to make freely-available to the research community. Future work also includes investigating the effect (if any) of users' characteristics, such as gender, age and computer self-efficacy, on the stress recognition accuracy.

Acknowledgments. This paper has been co-financed by the European Union (European Social Fund – ESF) and Greek National funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) (Funding Program: “Hellenic Open University”).

References

1. Chanel, G., Rebetez, C., Bétrancourt, M., Pun, T.: Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Trans. Syst. Man Cybern. Part Syst. Hum.* **41**, 1052–1063 (2011)
2. Vermeeren, A.P.O.S., Law, E.L.-C., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K.: User experience evaluation methods: current state and development needs. In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. pp. 521–530. ACM, New York (2010)
3. Ganglbauer, E., Schrammel, J., Tscheligi, M.: Possibilities of psychophysiological methods for measuring emotional aspects in mobile contexts. In: *Proceedings of Mobile HCI (2009)*

4. Mandryk, R.L., Atkins, M.S., Inkpen, K.M.: A continuous and objective evaluation of emotional experience with interactive play environments. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 1027–1036. ACM, New York (2006)
5. Scheirer, J., Fernandez, R., Klein, J., Picard, R.W.: Frustrating the user on purpose a step toward building an affective computer. *Interact. Comput.* **14**(2), 93–118 (2001)
6. Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: DEAP: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **3**, 18–31 (2012)
7. Lin, T., Omata, M., Hu, W., Imamiya, A.: Do physiological data relate to traditional usability indexes? In: *Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*. pp. 1–10. Computer-Human Interaction Special Interest Group (CHISIG) of Australia, Narrabundah, Australia (2005)
8. Ward, R.D., Marsden, P.H.: Affective computing: problems, reactions and intentions. *Interact. Comput.* **16**, 707–713 (2004)
9. Lazar, D.J., Feng, D.J.H., Hochheiser, D.H.: *Research Methods in Human-Computer Interaction*. Wiley, New York (2010)
10. Hernandez, J., Paredes, P., Roseway, A., Czerwinski, M.: Under pressure: sensing stress of computer users. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 51–60. ACM, New York (2014)
11. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* **6**, 156–166 (2005)
12. Lunn, D., Harper, S.: Using Galvanic skin response measures to identify areas of frustration for older web 2.0 users. In: *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, pp. 34:1–34:10. ACM, New York (2010)
13. Liapis, A., Karousos, N., Katsanos, C., Xenos, M.: Evaluating user’s emotional experience in HCI: the physiOBS approach. In: Kurosu, M. (ed.) *HCI 2014, Part II. LNCS*, vol. 8511, pp. 758–767. Springer, Heidelberg (2014)
14. Drachen, A., Nacke, L.E., Yannakakis, G., Pedersen, A.L.: Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In: *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, pp. 49–54. ACM, New York (2010)
15. Russell, J.A., Weiss, A., Mendelsohn, G.A.: Affect Grid: a single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* **57**, 493–502 (1989)