Please, cite as:
Suni Lopez F., Condori-Fernandez N., Catala A. (2019) Towards Real-Time Automatic Stress Detection for Office Workplaces. In: Lossio-Ventura J., Muñante D., Alatrista-Salas H. (eds) Information Management and Big Data. SIMBig 2018. Communications in Computer and Information Science, vol 898. pp 273-288, Springer, Cham
This is a preprint. The final version is available at:
https://doi.org/10.1007/978-3-030-11680-4_27

Towards Real-time Automatic Stress Detection for Office Workplaces

Franci Suni Lopez[1 2] and Nelly Condori-Fernandez[3 4] and Alejandro Catala[5]

[1] Universidad Católica San Pablo, Arequipa, Perú
[2] Universidad Nacional de San Agustín de Arequipa, Perú
franci.suni@ucsp.edu.pe, fsunilo@unsa.edu.pe,
[3] Universidade da Coruna, Spain
[4] Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
n.condori.fernandez@udc.es, n.condori-fernandez@vu.nl
[5] Centro Singular de Investigacion en Tecnoloxias da Informacion (CiTIUS), Universidade de Santiago de Compostela, Spain
alejandro.catala@usc.es

Abstract. In recent years, several stress detection methods have been proposed, usually based on machine learning techniques relying on obstructive sensors, which could be uncomfortable or not suitable in many daily situations. Although studies on emotions are emerging and rising in Software Engineering (SE) research, stress has not been yet well investigated in the SE literature despite its negative impact on user satisfaction and stakeholder performance.
In this paper, we investigate whether we can reliably implement a stress detector in a single pipeline suitable for real-time processing following an arousal-based statistical approach. It works with physiological data gathered by the E4-wristband, which registers electrodermal activity (EDA). We have conducted an experiment to analyze the output of our stress detector with regard to the self-reported stress in similar conditions to a quiet office workplace environment when users are exposed to different emotional triggers.

Keywords: stress detection · physiological data · emotional trigger

1 Introduction

Wearable technology is gaining popularity and the interest to include these devices as useful input for diverse software applications beyond simply gathering data have awakened the interest of software industry too. Moreover, wearable sensing technology for emotion recognition is becoming less obtrusive and inexpensive, what have favored considering biosignals in different sectors such as e-health, e-commerce, wellness, e-learning, and games. Leading organizations recognize that social aspects are just as important to long-term success as economic aspects. Particular focus is given to the labor conditions for reducing risks associated with work-related stress. Physiological stress is one of the factors that are most affecting current modern industries. For instance, Graziotin *et al.* [19]

proposed a theory about the impact of the effects on programming performance. While studies on emotions are emerging and rising in Software Engineering (SE) research, stress has not been yet well investigated in the SE literature despite its negative impact not only on stakeholder performance but also on user satisfaction and acceptance (*e.g.,* [18], [11]).

Currently, most of the stress detection methods are based on Machine Learning (ML) techniques (*e.g.,* support vector machine [35]). The main disadvantage of using these methods is the need for big data sets to carry out the training stage, where the machine learns about the user behavior in relationship with predefined tasks. To address this issue regarding the training of stress detector, we use an arousal-based statistical approach for detecting stress in real-time. This introduces two main advantages for the resulting stress detector: i) reliability is independent of a training data set, in contrast to the requirement imposed by approaches based on machine learning algorithms; ii) higher flexibility is provided since the detector can be used in different user conditions. In this paper, we present and evaluate experimentally an automatic stress detector that uses wearable sensors for gathering physiological data. We targeted office employees (programmers and junior researchers) working with computers (*i.e.,* desktop, laptop), who were exposed to several types of emotional triggers.

The paper is organized as follows. Section 2 discusses related works on stress recognition, and Section 3 presents some scenarios where a real-time stress detector can be useful in SE. Section 4 presents the theory about emotional triggers. The description of the algorithms used in our stress detector is presented in Section 5. Section 6 provides the experiment design and results. Finally, conclusions and future work are discussed in Section 7.

2 Related Work

During many years, researchers in several computer science fields have paid attention to developing methods to recognize and understand human emotions. For instance, based on natural language processing (*e.g.,* [27], [38], [41]); or emotion recognition through facial expression (*e.g.,* [8], [25], [28]) or using physiological data (*e.g.,* [17], [16], [9] [20], [21], [31], [36], [15], [35], [5]).

In the strand of works dealing with stress detection relying only on physiological data, Mozos *et al.* [31] proposed to combine machine learning techniques using EDA, photoplethysmogram (PPG) and heart rate variability (HRV) signals to detect stress in social situations using The Trier social stress test (TSST) as stressful. Garcia *et al.* [15] used accelerometer (ACC) data of a mobile phone to recognize stress in real workplace environments of thirteen subjects using two classification models: naive bayes and decision trees. They obtained an accuracy of 71% and their study lasted 8 weeks.

Sanno and Picard [35] implemented different machine learning classifiers to detect stress: Support Vector Machine (SVM) with linear kernel, SVM with Radial basis function (RBF) kernel, k-nearest neighbors, Principal component analysis (PCA) and SVM with RBF kernel and k-nearest neighbors. Their work is focused on comparing the performance the implemented algorithms using the collected data of the subjects (skin conductance, ACC and mobile phone usage) of five days. Kocielnik *et al.* [23] described a framework to detect stress in the context of a person's activities. They use a min-max algorithm and and ACC as source data. Bogomolov *et al.* [5] collected mobile phone activity (i.e. call log, SMS log, Bluetooth interactions) of 117 subjects to recognize stress during common daily activities. They applied different classifiers: SVM, artificial neural networks, ensemble of tree classifiers based on a Breiman's Random Forest (RF) and Friedmans Generalized Boosted Model (GBM). Similarly, using a range of machine learning techniques, some other examples can be found in the existing literature (*e.g.,* [34], [3], [10], [36], [21], [37]).

Table 1. Comparative chart of the most representative related works of stress recognition.

Author	**Classification algorithm**	**Source data**	**Evaluation tools**	**Context**
Mozos *et al.* [31]	SVM, AdaBoost, and k-nearest neighbor.	EDA, PPG and HRV.	Accuracy of 89.75%, precision of 89.5% and recall of 95%.	Social situations using the TSST.
Garcia-Ceja *et al.* [15]	Naive bayes and decision trees.	ACC.	Accuracy of 71%.	Real working environments.
Sano and Picard [35]	SVM, RBF, k-nearest neighbors, PCA, SVM and PCA.	SC, ACC and mobile phone usage.	Accuracy of 75%.	Stress detection that subjects are able to perceive and report.
Kocielnik *et al.* [23]	Min-max algorithm.	SC and ACC.	No reported	Subject's activities.
Bogomolov *et al.* [5]	SVM, ANNs, tree classifiers based on RF and GBM.	Mobile phone activity	Accuracy of 72.39%.	Common daily activities.

Table 1 summarizes the most representative related work, illustrating the diversity of used algorithms to recognize stress. Most of these works use a machine learning method to implement the classifier; as we indicated previously, there can be some issues concerning the large training datasets required and their request to gather data and train new classifiers for every single task/context of use. In contrast, we use an arousal-based statistical approach that does not need a big dataset to learn a model for recognizing stress and can work in different tasks.

Next we envision prospective scenarios of use in the context of SE in which a real-time stress detector could be useful.

3 Scenarios of Use in the Context of SE

A real-time stress detector could enhance/contribute to the **emotional labor in SE**, which refers to the process of managing feelings and expressions to fulfill the emotional requirements of a software engineer. For instance, in tasks that demand the collaboration of stakeholders with different perspectives, such as reviews-based requirements validation [13], analysts could get awareness on their stress level, which can be helpful not only for regulating their negative emotions but also for having a better performance in validating requirements.

Another potential scenario is in the development of large and complex software projects that require a continuous evolution, and maintenance, where the history of stress could help human resources managers in their decision-making processes. For instance, identifying members of a development team, who could be experiencing long-term stress that might be affecting their productivity. These members suffering stress could become potential deserters from the company.
Two general worth exploring scenarios where a real-time stress detector can be useful are:

Usability and software testing based on emotions for quality assurance and user experience: To detect interaction pitfalls or defects in the user interface and/or software functionality that could cause certain level of stress on end-users, which can be used to enhance the software quality [11]. It has a diagnosis purpose of the software developed, and therefore as an additional input to assure quality in future software versions. The real-time feature is relevant to determine which variations in the stress detection are associated to the use of specific parts of the software (*e.g.,* elements of the user interface [32], awareness elements of a software game [39]) which can facilitate us to discover new quality requirements from actual user needs (*e.g.,* [12]).

Development of self-adaptive software systems guided by emotion: As a kind of context-aware system, in which part of the user context is provided by the emotional states over time while interacting with software systems, self-adaptation could be guided by emotions ([30],[29]).

4 Emotional Triggers

An emotion is just a response we give to a stimulus or event, whether it is external, or even internal, such as a memory or an idea [14]. Additionally, in experimental settings, researchers can generate emotions on users intentionally, by using specific emotional triggers determined by the emotions to be induced.

In this respect, an emotional trigger is any stimulus that generates a negative or positive emotion (*e.g.,* uncomfortable or comfortable temperature, environmental noise, *etc.*). According to Kanjo *et al.* [22], emotional triggers can be classified into seven types: environment, physical movements, memories, perception, interacting with others, accomplishments and failure.

Nowadays, different kinds of emotional triggers exist. We briefly introduce existing stress triggers that will be used to evaluate our stress detector: Westman and Walters [40] and Passchier-Vermeer and Passchier [33] considered an environmental trigger, where participants are exposed by five minutes to listen fire alarm sounds. The Sing-a-Song Stress Test [7] is a social trigger, where participants are asked to sing a song aloud for 30 seconds with their arms still.An example of a cognitive trigger is the Stroop Task [24], where participants have to pay attention and react to the color of a word while ignoring the word itself. A reduced version of this trigger works with 4 colors, using the words "green", "red", "yellow", "blue" written in all four different colors. Words are presented randomly for each participant.

5 Automatic Stress Detector

Figure 1 shows the stress detection process of our approach that has been automated to detect stress of individuals (*e.g.,* programmers, testers) in real time. We use wearable sensors (*i.e.,* E4-wristband[6]) to collect physiological data. In this first version, we focus only on sensing electrodermal activity (EDA) as a main input for the implementation of the stress detector. A transient increase on the EDA signal is proportional to sweat secretion and it is related to stress [2], [6]. The main functionality of the stress detector is to determine whether

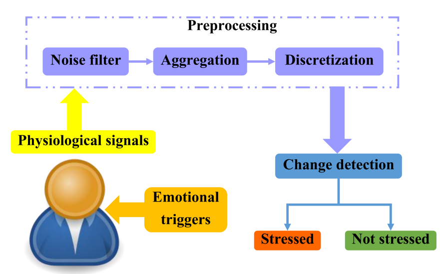


Fig. 1. Overview of a stress detection process.

the user is stressed or not. The detector will mark a label of "*stressed*" or "*not stressed*". We have implemented the preprocessing steps proposed by Bakker *et al.* [2] for arousal detection in an integrated pipeline to enable real-time processing (see Figure 1 for the involved preprocessing steps). Next, we explain the methods/algorithms that were used in the stress detection process.

[6] https://www.empatica.com/e4-wristband

5.1 Noise filter

We use *EDA* to recognize stress changes, then the first stage in the pipeline of the stress detector is to collect raw signals by the Empatica's E4-wristband, Figure 2 (part a) presents a common sample of *EDA* signals, which is measured in microSiemens (μS), a unit of electric conductance. Usually for measuring EDA is required two electrodes that need skin contact to produce a reliable signal, therefore the quality of the collected *EDA* signals depends on the continuity of the contact between user skin and the device's sensors. However, the contact is not the same in all users and noise could be introduced in the signal. Hence, noise filtering is needed to mitigate these issues in the input (*e.g.,* in Figure 2 (part a), we can find some gaps as a consequence of weak skin contact). Before analyzing *EDA* signals, it is important to clean raw data, because noise might be mistaken as genuine peaks. Therefore, the first step of the preprocessing is to apply a median filter over a moving window of size $n = 100$ EDA samples, as suggested in [2]. Figure 2 (b) shows the noise filtering of the collected raw data.

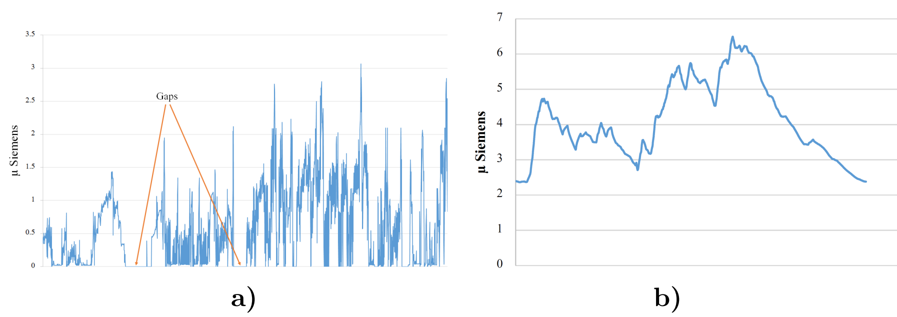


Fig. 2. a) Gaps occur when the contact between the user skin and the sensors is not tight. **b)** Filtered raw *EDA* signals.

5.2 Aggregation

The EDA signal acquired by the E4-wristband is sampled at 4Hz (*i.e.,* the device provides 4 samples or readings per second, which means 240 samples per minute). Based on [2], we apply an aggregation step of each minute over the filtered input signal: given y' is a moving window of size $m = 240$ (the EDA samples of one minute), where $y_1, ..., y_m$ is aggregated to a single value y'' where $y'' = max(y')$. For instance, Figure 3 (part a) shows the aggregation of approximately 7200 filtered EDA samples to 30 representative points (collected signals of 30 minutes).

5.3 Discretization

In this step, the data is discretized using the symbolic aggregate approximation (SAX) method [26]. It is a means for very efficient local discretization of time

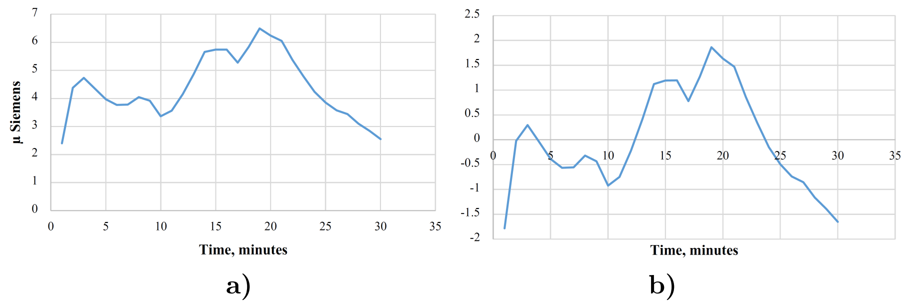


Fig. 3. a) Aggregated process over previous filtered data. **b)** Z-normalization of aggregated data.

series subsequence from 1 to 5 that can be interpreted as levels of stress variation (1: completely relaxed to 5: maximum arousal). Those levels should not be understood as absolute levels of arousal, but rather as a local relative measure of arousal.

The input for SAX is a time-series X of length n and the output is a string of length w, where $w < n$ typically, the output string is normalized to an alphabet of size > 2. The algorithm consists of the following two stages:

- Transformation of original time-series into a Piecewise Aggregate Approximation (PAA) representation. To do this, first it is necessary a Z-normalization (see Equation 1), where the mean is around 0 and the standard deviation is close to 1, using the following formula:

$$x'_i = \frac{x_i - \mu}{\rho} \quad (1) \qquad\qquad \overline{\text{x}}_i = \frac{M}{n} \sum_{j=\frac{n}{M}(i-1)+1}^{(\frac{n}{M})i} x_j \quad (2)$$

  Where μ is the mean of the time series and ρ is the standard deviation. After the Z-normalization, we can apply PPA transform, which approximates the time-series into vector $\overline{\text{X}} = (\overline{\text{x}}_1, ..., \overline{\text{x}}_M)$ of length $M \leq n$ (See Figure 3 (b)). Where each $\overline{\text{x}}_i$ is calculated with the Equation 2. With the objective to reduce the dimensionality from n to M, first we divide the time-series to n/M equally sized samples and calculate the mean for each sample (See Figure 4 (a)).
- Transformation of the PAA data into a string. The method use a breakpoint or cuts $B = \beta_1, \beta_2, ..., \beta_{\alpha-1}$ such that $\beta_{i-1} < \beta_i$ and $\beta_0 = -\infty, \beta_\alpha = \infty$ divides the total area in equal subareas. Additionally, it assigns a symbol $alpha_j$ to each interval $[\beta_{j-1}, \beta_j)$, and the final conversion from PAA coefficients $\bar{C}$ into a SAX string $\hat{C}$ is with the Equation 3. Figure 4 (b) shows SAX transformation of the previous preprocessing signals.

$$\hat{c} * i = alpha * j, iif, \bar{c} * i \in [\beta_{j-1}, \beta_j) \quad (3)$$

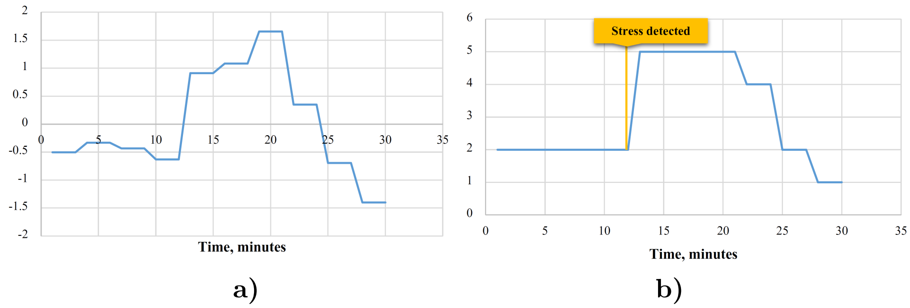


Fig. 4. a) PAA representation of the preprocessing data. **b)** *SAX* representation and stress detection using ADWIN algorithm.

5.4 Change detection

We use a change detection algorithm based on ADaptive WINdowing (ADWIN) method [4]. ADWIN computes the mean for each split of a sequence of signals and analyzes the statistically significant difference between two consecutive splits. When a statistically significant difference is detected at point p_i, ADWIN drops the data backwards from p_i, after it repeats the splitting procedure until no significant differences be found in the sequence. For instance, given ϕ_1 and ϕ_2 as the means of two splits of a sequence of EDA signals, then $|\phi_1 - \phi_2| > \in_{cut}$ is the condition for a change detection that is computed with the Equation 4.

$$\in_{cut} = \sqrt{\frac{2}{m} . \sigma_W^2 . ln \frac{2}{\delta'}} + \frac{2}{3m} ln \frac{2}{\delta'} \tag{4}$$

where σ_W^2 is the variance of the elements of W. δ is the desired confidence and $\delta' = \delta/(ln\ n)$ [2]. Figure 4 (b) shows the output the algorithm detecting a stress change.

6 Experiment

In order to validate the stress detector implemented, we designed an experiment where participants experienced different stressful situations caused by emotional triggers introduced in Section 4.

The goal of the experiment is to evaluate the performance of our stress detector in terms of its accuracy. This evaluation was performed from the viewpoint of office workers in the context of performing certain tasks that cause stress (emotion trigger). From this goal, the following research question is derived:

RQ$_1$: *How accurately is the stress detector able to recognize subjects stress under different types of emotional triggers?*

Based on the defined research question, we have the **independent variables** *emotional trigger*, originally with 3 levels (environmental: fire alarm; cognitive: Stroop Task; and social: Sing-a-Song Stress Test). After running a pilot study, we decided to remove the social trigger to reduce the length of the experiment to thirty minutes. This is further explained in the data collection section. Figure 5 (c) shows a screenshot of the instructions for the Stroop Task.

As **dependent variables:** *subject stress status*, which is measured in a nominal scale (stressed or not stressed); and *perceived stress* measured by means of a self-response questionnaire.

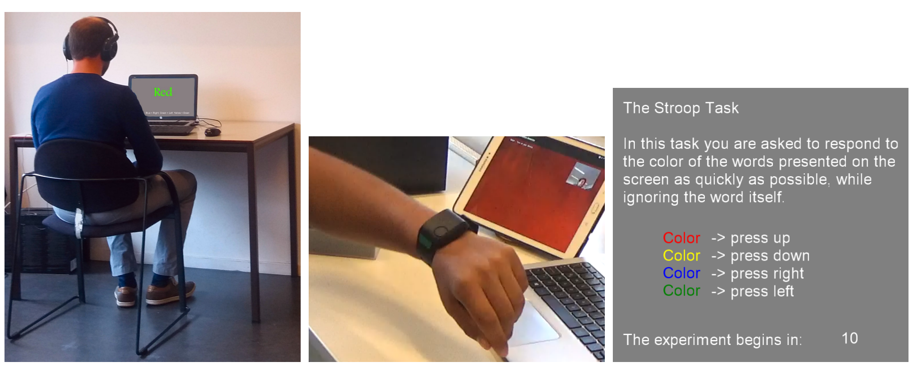
Our hypothesis is that *when different types of emotional triggers are delivered, the stress detector is able to recognize stress with a similar accuracy.* Accuracy refers to the closeness of a measured value to a ”true value”. In our study, the true value of perceived stress was determined by the subjects of the experiment.

6.1 Subjects

Twelve subjects from University of Twente (The Netherlands), involved in research in computing areas (*i.e.,* Master students, PhD candidates), participated voluntarily in the experiment, whose ages ranged between 21 and 32 years old. Seven are women and five men.

6.2 Instrumentation and procedure

The experiment was carried out in a quiet room equipped with a table and a chair as shown in Figure 5(a). Subjects interacted with a laptop where the Stroop Task (cognitive trigger implemented with Psychopy[7]) was installed. Also, subjects wore the E4-wristband and headphones to interact with the environmental trigger. Figure 5 (b) shows the correct position of the E4-wristband on the non-dominant hand of the subject.



a) **b)** **c)**

Fig. 5. a) A subject in the experiment room interacting with emotional triggers. b) E4-wristband placed on the non-dominant hand. c) Instructions for interacting with cognitive trigger (stroop task).

[7] http://www.psychopy.org/

The evaluation followed a within-subjects design, where all subjects were exposed individually to both cognitive and environmental triggers (treatments). The order in which the subjects interacted with the treatments were assigned randomly. Figure 6 shows the procedure of the experiment that consists of two phases:

Phase 1. Firstly, the subjects were asked to read and sign the informed consent form, which described the purpose and structure of the experiment. Subjects were informed beforehand about the sensing device and the possibility of experiencing some stress during the experiment. Furthermore, they were informed that they could pass on the task at any time if they considered stress unbearable. After signing the consent form, each subject got put on and adjusted the E4-Wristband to enable the gathering of physiological data. Then subjects were asked to complete a demographic questionnaire, and press a button to start the experimental tasks when they were ready. This phase lasts around five minutes.

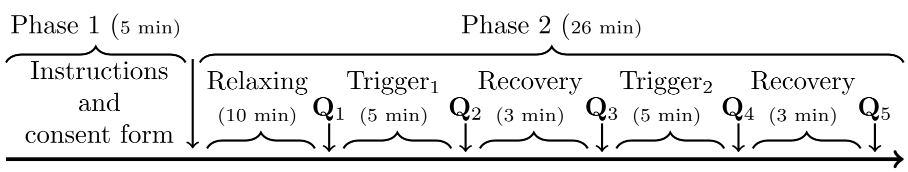


Fig. 6. Experiment procedure and timeline.

Phase 2. Each subject was asked to sit on her/his own chair in a comfortable position for 10 minutes. We asked them to stay quiet and relaxed during this period. Then subjects interacted with the corresponding treatments (five minutes each). Also, subjects had three minutes of recovery after each emotional trigger. Participants self-reported their stress status before, during the delivery of the corresponding emotion trigger and after the last trigger, the questions were answered progressively. The closed questions that were formulated during the experiment were in a 7-point-ordinal-scale (presented on-screen). For instance, delivering first an environmental trigger and then a cognitive trigger, the sequence of questions were as follows (see Figure 6):

- $\mathbf{Q_1}$**:** How stressed are you at this moment?
- $\mathbf{Q_2}$**:** How stressed were you WHILE listening the noise?
- $\mathbf{Q_3}$**:** How stressed are you at this moment?
- $\mathbf{Q_4}$**:** How stressed were you WHILE doing the color task?
- $\mathbf{Q_5}$**:** How stressed are you at this moment?

6.3 Data collection

The twelve subjects S01-S12 interacted with two emotional triggers successfully (i.e. environmental and cognitive). The experiment obtained an ethical approval from the Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science of the University of Twente. Raw data and questionnaires answers were encrypted (WinZip AES encryption: 128-bit AES) and stored in a secured remote location for later analysis.

We validated the experimental design with a pilot study involving two participants (who did not take part in the final evaluation), to ensure the task descriptions were fully understandable, its implementation error-free, and to check the time and any further issue regarding the experimental design. The initial experiment was originally designed with three emotional triggers (cognitive, social and environmental), and approximately it lasted forty minutes. The feedback collected in the pilot suggested that the experiment was going to last too long and that the social trigger was not causing stress as expected given the time available. Hence, we changed our design to exclude the social-emotional trigger in order to prevent issues and reduce the experiment time approximately to thirty minutes (around 6480 EDA samples).

Table 2. Labeled results of the questionnaires and stress detector.

Subject	**Trigger**	**Reported stress**	**Stress detector**	**Trigger**	**Reported stress**	**Stress detector**
S01	Cognitive	Stressed	Stressed	Environmental	Not stressed	Not stressed
S02		Not stressed	Not stressed		Not stressed	Not stressed
S03		Not stressed	Not stressed		Stressed	Stressed
S04		Not stressed	Not stressed		Not stressed	Not stressed
S05		Stressed	Not stressed		Stressed	Not stressed
S06		Stressed	Not stressed		Not stressed	Not stressed
S07		Not stressed	Not stressed		Not stressed	Not stressed
S08		Not stressed	Not stressed		Not stressed	Not stressed
S09		Not stressed	Not stressed		Not stressed	Not stressed
S10		Stressed	Stressed		Not stressed	Not stressed
S11		Not stressed	Not stressed		Not stressed	Stressed
S12		Not stressed	Not stressed		Not stressed	Stressed

6.4 Threats to validity

Internal validity. As our main objective is to evaluate the performance of our stress detector, in this study, we decided to use three well-known emotional triggers from the psychology community. However, given that these triggers had not been used previously in Software Engineering, we acknowledge the fact that the selected triggers could not always generate stress on the subjects(programmers and researchers in computer science) due to different other factors (*e.g.,* greater resilience) that were not investigated in this study.

Given this interaction with different emotional triggers (treatments), with the purpose of avoiding that the the first emotional trigger does not affect on the next one, we set out relaxing and recovery periods.

Another possible threat is the effect of the instrumentation used during the experiment (*i.e.,* E4-Wristband), which could also have been causing any stress level. In order to know whether this instrument could be considered as additional potential triggers, we asked participants to complete a post-questionnaire regarding this issue for further investigation.

External validity. Given the low number of subjects and the fact they were researchers working in computing-related areas but not fully working as software engineers, one potential threat to external validity is regarding the generalization of our results. Moreover, as our controlled experiment was conducted in a lab setting, involving real practitioners would have been harder. We think that having this lab-setting still allowed us to evaluate the stress detector without interruption of external factors (*i.e.,* meetings, calls) as a first necessary step to validate and continue developing the detector.

Construct validity. The use of a single device to measure physiological stress (a construct) could be considered as main threat to construct validity of this study.

6.5 Results

The self-reported stress was rated in 7-point ordinal-scale questions that were gathered before and during the experiment (1 = "*not stressed*" to 7 ="*extremely*"). We labeled the overall self-reported score as "*stressed*" when the difference led to an increase equal o higher than 3 (threshold) in the perceived value of self-reported stress; otherwise, it was labeled as "*not stressed*" according to [5]. Table 2 summarizes the labels for each subject to assess the accuracy or *trueness* of our detector, by comparing the computed label with the final self-reported label; red cells indicate the cases where the stress detector missed [8].

For answering our research question, we use the following (well-known) metrics regarding: precision (Equation 5), recall (Equation 6) and accuracy (Equation 7):

$$precision = \frac{TP}{TP + FP} \quad (5) \qquad recall = \frac{TP}{TP + FN} \quad (6)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

[8] Raw data and details of each subject can be found at https://goo.gl/eQ4KC2

where TP indicates true positives, TN true negatives, FP false positive and FN false negatives. In our case, we consider true positive cases the examples where reported stress and stress detector are labeled as *stressed* (see Table 2).

We obtained an accuracy of 79.17%, a precision of 60% and a recall of 50%. By comparing our results with other machine-learning based recognition methods, we consider that our method has a good accuracy because it oscillates between 70% and 85% [15] [1], values reported in the literature of stress recognition using physiological data. It is also important to remark that most of these existing recognition methods do not report precision and recall measures. Overall, a method that gets a high recall value it is considered as a good detector. In our study, despite we obtained a 50% of recall, we consider that it can be due to the threats that were identified after the results analysis.

7 Conclusions and Future Work

In this paper, we present the stress detection process of the arousal-based statistical method. The different algorithms and techniques used for supporting such detection process were implemented and integrated as part of our real-time automatic stress detector, in a single processing pipeline. In order to evaluate its accuracy, an experiment was conducted with 12 subjects using the E4-wristband device to gather physiological data. Comparing the outcome of our stress detector with the reported by each subject (perceive stress), the detector obtained an accuracy of 79.17%.

An interesting observation is that although some subjects did not feel stress when an emotional trigger was delivered, the outcome of the detector was consistent with the corresponding perceived stress value. However, from this observation, we can also see that the emotional trigger was not very effective for generating stress in all cases (subjects). A possible explanation for this might be due to the different resilience extent of our participants or the need to exposing them longer to the stimuli. As a future empirical work, beyond checking the accuracy of the detector, more research is needed regarding the role of emotional triggers and resilience of people working in office workplaces (*e.g.,* developer facing stress in unexpected situations). To do this, we plan to conduct a field experiment with practitioners from a Spanish SME involved in software projects with multi-tenancy characteristics.

Acknowledgments

Authors would like to thank to Dirk Heylen, head of HMI Lab of University of Twente, for facilitating us the HMI Lab to conduct the experiments and his early feedback. Also, We thank all the participants who took part in our research. This work has been supported by grant 234-2015-FONDECYT (Master Program) from Cienciactiva of the National Council for Science, Technology and Technological Innovation (CONCYTEC-PERU). Moreover, this work has received financial support from the Spanish Ministry of Economy, Industry and

Competitiveness with the Project: TIN2016-78011-C4-1-R; Council of Culture, Education and University Planning with the project ED431G/08, the European Regional Development Fund (ERDF).

References

1. Alberdi, A., Aztiria, A., Basarab, A.: Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. Journal of Biomedical Informatics **59**, 49–75 (feb 2016)
2. Bakker, J., Pechenizkiy, M., Sidorova, N.: What's your current stress level? detection of stress patterns from gsr sensor data. In: Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops. pp. 573–580. ICDMW '11, IEEE Computer Society, Washington, DC, USA (2011)
3. Bauer, G., Lukowicz, P.: Can smartphones detect stress-related changes in the behaviour of individuals? In: 2012 IEEE International Conference on Pervasive Computing and Communications Workshops. pp. 423–426 (March 2012)
4. Bifet, A., Gavaldà, R.: Learning from time-changing data with adaptive windowing. In: Proceedings of the 2007 SIAM International Conference on Data Mining, pp. 443–448. Society for Industrial and Applied Mathematics (apr 2007)
5. Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., Pentland, A.S.: Pervasive stress recognition for sustainable living. In: 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS). pp. 345–350 (March 2014)
6. Boucsein, W.: Electrodermal Activity. Springer US (2012)
7. Brouwer, A.M., Hogervorst, M.A.: A new paradigm to induce mental stress: the sing-a-song stress test (SSST). Frontiers in Neuroscience **8** (jul 2014)
8. Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan, S.: Analysis of emotion recognition using facial expressions, speech and multimodal information. In: Proceedings of the 6th International Conference on Multimodal Interfaces. pp. 205–211. ICMI '04, ACM, New York, NY, USA (2004)
9. Canento, F., Fred, A., Silva, H., Gamboa, H., Lourenço, A.: Multimodal biosignal sensor data handling for emotion recognition. In: 2011 IEEE SENSORS Proceedings. pp. 647–650 (Oct 2011)
10. Carneiro, D., Castillo, J.C., Novais, P., Fernández-Caballero, A., Neves, J.: Multimodal behavioral analysis for non-invasive stress detection. Expert Systems with Applications **39**(18), 13376–13389 (dec 2012)
11. Condori-Fernandez, N.: Happyness: An emotion-aware QoS assurance framework for enhancing user experience. In: Proceedings of the 39th International Conference on Software Engineering Companion. pp. 235–237. ICSE-C '17, IEEE Press, Piscataway, NJ, USA (2017)
12. Condori-Fernandez, N., Catala, A., Lago, P.: Discovering requirements of behaviour change software systems from negative user experience. In: Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings. pp. 222–223. ICSE '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3183440.3195072, http://doi.acm.org/10.1145/3183440.3195072
13. Condori-Fernandez, N., España, S., Sikkel, K., Daneva, M., González, A.: Analyzing the effect of the collaborative interactions on performance of requirements

validation. In: Salinesi, C., van de Weerd, I. (eds.) Requirements Engineering: Foundation for Software Quality. pp. 216–231. Springer International Publishing, Cham (2014)

14. Ekman, P.: Basic emotions. In: Handbook of Cognition and Emotion, pp. 45–60. John Wiley & Sons, Ltd (jan 1999)
15. Garcia-Ceja, E., Osmani, V., Mayora, O.: Automatic stress detection in working environments from smartphones x2019; accelerometer data: A first step. IEEE Journal of Biomedical and Health Informatics **20**(4), 1053–1060 (July 2016)
16. Girardi, D., Lanubile, F., Novielli, N.: Emotion detection using noninvasive low cost sensors. In: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE (oct 2017)
17. Gouizi, K., Maaoui, C., Reguig, F.B.: Negative emotion detection using emg signal. In: 2014 International Conference on Control, Decision and Information Technologies (CoDIT). pp. 690–695 (Nov 2014)
18. Graziotin, D., Wang, X., Abrahamsson, P.: Software developers, moods, emotions, and performance. IEEE Software **31**(4), 24–27 (July 2014)
19. Graziotin, D., Wang, X., Abrahamsson, P.: How do you feel, developer? an explanatory theory of the impact of affects on programming performance. PeerJ Computer Science **1**, e18 (aug 2015)
20. Guendil, Z., Lachiri, Z., Maaoui, C., Pruski, A.: Emotion recognition from physiological signals using fusion of wavelet based features. In: 2015 7th International Conference on Modelling, Identification and Control (ICMIC). pp. 1–6 (Dec 2015)
21. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. Trans. Intell. Transport. Sys. **6**(2), 156–166 (Jun 2005)
22. Kanjo, E., Al-Husain, L., Chamberlain, A.: Emotions in context: examining pervasive affective sensing systems, applications, and analyses. Personal and Ubiquitous Computing **19**(7), 1197–1212 (Oct 2015)
23. Kocielnik, R., Sidorova, N., Maggi, F.M., Ouwerkerk, M., Westerink, J.H.D.M.: Smart technologies for long-term stress monitoring at work. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. pp. 53–58 (June 2013)
24. Lattimore, P.: Stress-induced eating: an alternative method for inducing ego-threatening stress. Appetite **36**(2), 187–188 (apr 2001)
25. Le, H.T., Vea, L.A.: A customer emotion recognition through facial expression using kinect sensors v1 and v2: A comparative analysis. In: Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication. pp. 80:1–80:7. IMCOM '16, ACM, New York, NY, USA (2016)
26. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. pp. 2–11. DMKD '03, ACM, New York, NY, USA (2003)
27. Liu, Y., Yu, X., Chen, Z., Liu, B.: Sentiment analysis of sentences with modalities. In: Proceedings of the 2013 International Workshop on Mining Unstructured Big Data Using Natural Language Processing. pp. 39–44. UnstructureNLP '13, ACM, New York, NY, USA (2013)
28. Menne, I.M., Lugrin, B.: In the face of emotion: A behavioral study on emotions towards a robot using the facial action coding system. In: Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. pp. 205–206. HRI '17, ACM, New York, NY, USA (2017)

29. Mocholi, J., Jaen, J., Krynicki, K., Catala, A., Picón, A., Cadenas, A.: Learning semantically-annotated routes for context-aware recommendations on map navigation systems. Applied Soft Computing Journal **12**(9), 3088–3098 (2012). https://doi.org/10.1016/j.asoc.2012.05.010
30. Mocholí, J.A., Martínez, J.J., Catalá, A.: Towards affection integration on context-aware recommendation of semantically annotated routes. In: Workshop Proceedings of the 7th International Conference on Intelligent Environments, IE 2011, Nottingham, United Kingdom, July 25-28, 2011. pp. 51–62 (2011). https://doi.org/10.3233/978-1-60750-795-6-51, https://doi.org/10.3233/978-1-60750-795-6-51
31. Mozos, O.M., Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Dobrescu, R., Ferrandez, J.M.: Stress detection using wearable physiological and sociometric sensors. International Journal of Neural Systems **27**(02), 1650041 (mar 2017)
32. Ormeño, Y.I., Panach, J.I., Condori-Fernandez, N., Pastor, O.: Towards a proposal to capture usability requirements through guidelines. In: IEEE 7th International Conference on Research Challenges in Information Science (RCIS). pp. 1–12 (May 2013). https://doi.org/10.1109/RCIS.2013.6577677
33. Passchier-Vermeer W, P.W.: Noise exposure and public health. Environ Health Perspect. p. 108(suppl 1): 123–31 (2000)
34. Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Mozos, O.M.: Stress detection using wearable physiological sensors. In: Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Toledo-Moreo, F.J., Adeli, H. (eds.) Artificial Computation in Biology and Medicine. pp. 526–532. Springer International Publishing, Cham (2015)
35. Sano, A., Picard, R.W.: Stress recognition using wearable sensors and mobile phones. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. pp. 671–676 (Sept 2013)
36. Sriramprakash, S., Prasanna, V.D., Murthy, O.R.: Stress detection in working people. Procedia Computer Science **115**, 359 – 366 (2017), 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India
37. Sun, F.T., Kuo, C., Cheng, H.T., Buthpitiya, S., Collins, P., Griss, M.: Activity-aware mental stress detection using physiological sensors. In: Gris, M., Yang, G. (eds.) Mobile Computing, Applications, and Services. pp. 211–230. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
38. Tang, D.: Sentiment-specific representation learning for document-level sentiment analysis. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. pp. 447–452. WSDM '15, ACM, New York, NY, USA (2015)
39. Teruel, M.A., Condori-Fernandez, N., Navarro, E., González, P., Lago, P.: Assessing the impact of the awareness level on a co-operative game. Information and Software Technology **98**, 89 – 116 (2018). https://doi.org/https://doi.org/10.1016/j.infsof.2018.02.008, http://www.sciencedirect.com/science/article/pii/S0950584918300314
40. Westman, J.C., Walters, J.R.: Noise and stress: a comprehensive approach. Environmental Health Perspectives **41**, 291–309 (oct 1981)
41. You, Q.: Sentiment and emotion analysis for social multimedia: Methodologies and applications. In: Proceedings of the 2016 ACM on Multimedia Conference. pp. 1445–1449. MM '16, ACM, New York, NY, USA (2016)