Annotation and Prediction of Stress and Workload from Physiological and Inertial Signals.

Arindam Ghosh[1], Morena Danieli[1], Giuseppe Riccardi[1]

Abstract— **Continuous daily stress and high workload can have negative effects on individuals' physical and mental well-being. It has been shown that physiological signals may support the prediction of stress and workload. However, previous research is limited by the low diversity of signals concurring to such predictive tasks and controlled experimental design. In this paper we present 1) a pipeline for continuous and real-life acquisition of physiological and inertial signals 2) a mobile agent application for on-the-go event annotation and 3) an end-to-end signal processing and classification system for stress and workload from diverse signal streams. We study physiological signals such as Galvanic Skin Response (GSR), Skin Temperature (ST), Inter Beat Interval (IBI) and Blood Volume Pulse (BVP) collected using a non-invasive wearable device; and inertial signals collected from accelerometer and gyroscope sensors. We combine them with subjects' inputs (e.g. event tagging) acquired using the agent application, and their emotion regulation scores. In our experiments we explore signal combination and selection techniques for stress and workload prediction from subjects whose signals have been recorded continuously during their daily life. The end-to-end classification system is described for feature extraction, signal artifact removal, and classification. We show that a combination of physiological, inertial and user event signals provides accurate prediction of stress for real-life users and signals.**

I. INTRODUCTION

Continuous high stress and workload can have negative effects on a person's physical and mental well-being. It has been strongly linked to numerous chronic health risks, such as cardiovascular disease, diabetes mellitus, obesity, hypertension, and coronary artery disease. It is also a contributory cause for unsuitable human behaviour, failure, and psychological breakdown among people from different age groups, professions, and culture, and has become a growing concern in workplaces around the world. A high level of stress due to heavy workload has been shown to increase the level of fatigue [1] among employees and even increase the risk factors of cancer [2].

While it is easy to identify and understand the source of physical stress or severe acute psychological stress, subtle and chronic stress due to continuous workload is more difficult to detect. Because of this most people are unaware of the level of stress in their lives. By the time people decide to seek medical help they are already in the advanced stages of stress induced exhaustion or are suffering from some noticeable ailment. To prevent this, there is a need for early identification and understanding of stress.

Since the early 1980s psychologists have used validated questionnaires for detecting stress and workload. Psychosomatic medicine often relies on questionnaire-based assessment of perceived stress [3]. Self assessment questionnaires are also widely used for evaluating stress coping strategies. Recently, researchers in the field of affective computing have shown that mobile phones and wearable devices are also capable of monitoring and measuring the levels of stress and workload. Mobile phones, along with wearable devices have turned into guardian angels who can keep watch over our health and wellbeing by unobtrusively monitoring our physiological signals. Physiological signals such as galvanic skin response (GSR), and heart rate variability (HRV) have been used to recognize stress and workload under different experimental settings ranging from driving scenarios [4] to working in an office [5] or a call center [6]. By watching out for risk factors such as sudden blood pressure drops [7] or abnormal heart rate [8] they can provide early life saving warnings.

However, while these physiological measures have shown promising results in controlled settings, continuous ambulatory monitoring in naturalistic settings suffers from a few challenges: 1) It is important to obtain stress and workload related annotations for training such a predictive system. This is difficult under real life situation since stress perception of an event changes or is often forgotten after a stretch of time; 2) Physiological signals are highly susceptible to noise from motion and local artifacts; 3) Combining the covert and overt signals is a complex task.

In this paper we address these problems using a structured approach for continuous stress and workload monitoring and feedback elicitation from users. Furthermore, we combine covert physiological signals with overt information provided by the users to improve the prediction of stress and workload. This can be used to create smart agents which can proactively warn users about their stress and workload states.

In Section II we describe the experimental design, the platform and protocol for data collection. In Section III we discuss the data preprocessing, feature extraction and supervised learning for stress and workload recognition.

II. EXPERIMENTAL DESIGN

This preliminary study was conducted with five healthy subjects (three males and two females, between ages of 30 and 45) holding regular desk jobs. The subjects were selected after an initial prescreening interview with a psychologist

[1] Signals and Interactive Systems Lab, Department of Information Engineering and Computer Science University of Trento, Italy arindam.ghosh@unitn.it, morena.danieli@unitn.it, giuseppe.riccardi@unitn.it

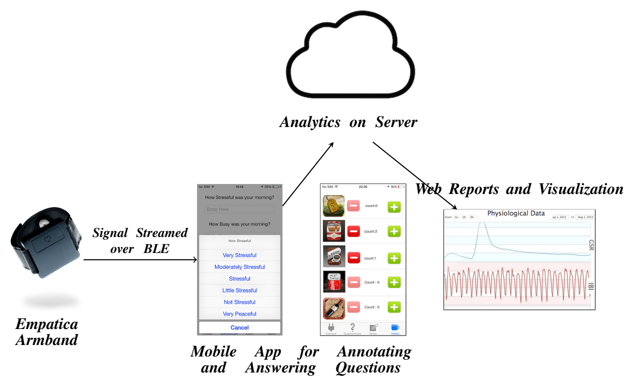

Fig. 1: StressTracker experimental platform including the wristband, mobile, and web components.

to eliminate the possibility of underlying mental condition such as hidden hypertension which might confound the study. The protocol for this research study was approved by the ethics committee of the Università degli Studi di Trento. The subjects were provided with the Italian version of the emotion regulation questionnaire by Balzarotti et al [9]. It is a ten-item self-report questionnaire which uses the cognitive reappraisal and expression suppression scales to evaluate an individual's tendency to regulate his or her emotion and respond to stress.

Emotion regulation refers to the process that individuals use to feel, express, and control the emotions they experience in their daily lives [10]. Different individuals use different emotion regulation strategies which can affect the way they experience and deal with stress. It has been shown that individuals who use *antecedent-focussed* strategies such as *Cognitive Reappraisal*, experience and cope with stress differently than individuals who use *response-focussed* strategies like *Expression Suppression*.

Subsequently, each subject was provided access to the StressTracker data annotation and acquisition platform. The platform and protocol is described in the following sections.

A. StressTracker Platform Description

The StressTracker data acquisition and annotation platform has been developed for continuously tracking, monitoring, visualizing and annotating stress-related signals. The platform consists of the following components:

Empatica E3 wristband: The Empatica E3[1] wristband is an unobtrusive, wearable, lightweight, wireless, multi-sensory signal acquisition device. It has four inbuilt sensors for continuously reporting Galvanic Skin Response (GSR), Photoplethysmograph (PPG) data, Skin Temperature (ST), and Tri-Axial Acceleration (ACC). It also reports Inter-Beat Interval (IBI) at discrete intervals.

StressTracker iPhone Agent: The StressTracker iPhone Agent was developed as a companion application for the Empatica E3 wristband. The *StressTracker* agent is capable of storing and streaming the physiological signals from the wristband and the motion sensors from the phone. It also elicits regular voice and text annotations from the subjects. The subjects also answered regular questions regarding their perceived state of stress and workload. Using the Agent Application, they also reported activities such as smoking, consumption of alcoholic and caffeinated beverages during the day.

StressTracker Web Interface: The StressTracker web interface can be used by the psychologist or the subjects to gain either a realtime or periodic view of both the acquired signals and the annotations. The web interface has different access rights depending on the role of the user (psychologist or subject). A subject is free to edit/update/delete his or her data or annotation at any time. The physiological signals and their variations during the day are depicted on a timeline along with their notes and annotations. These signals and notes can assist a subject in self reflection to gain a better understanding of their mental health.

B. Protocol and Data Collection

Each subject wore the Empatica E3 wristband for a period of seven days (five working days and a weekend) for 8-10 hours every day from morning, till evening. Thrice during the day (once in the morning, at lunchtime, and at the end of the day) the subjects used the StressTracker Agent Application to report their perceived stress and workload states. The answers were selected from a six-point Likert scale which ranged from "Very Peaceful" to "Very Stressful" for perceived stress; and "Completely free" to "Very Busy" for perceived workload. The subjects were also asked to take regular text and voice notes annotating activities, and events such as consumption of alcohol, nicotine and any other caffeinated beverage during the day. At the end of the day, they noted a brief textual or verbal description of their day, and using the online platform, reviewed, added or edited any information provided.

A total of 206 hours of sensor data was collected. The subjects annotated 61 instances of reported stress, and 60 instances of reported workload using the app.

III. EXPERIMENTS AND RESULTS

In most stress-related studies, workload is taken as the cognitive demand of the task. In such controlled experiments, the mental workload is increased by varying task complexity, and its effect on the stress response of the subject is observed. Our goal is however, to study and predict stress and workload individually under naturalistic settings. For our experiments we observed a low correlation (pearson = 0.58 p-value $<$ 0.05) between perceived stress and perceived workload.

In this section we discuss the initial preprocessing of the collected data to minimize noise. Then we discuss the feature extraction and classification experiments.

To detect the stress state and workload of the subject we need to extract useful information-bearing features from the different signal streams. We extract and combine features from the different physiological and inertial sensors on the Empatica Wristband and iPhone with the ***personal features*** of the subject. The personal features of the subject are the

[1] www.empatica.com

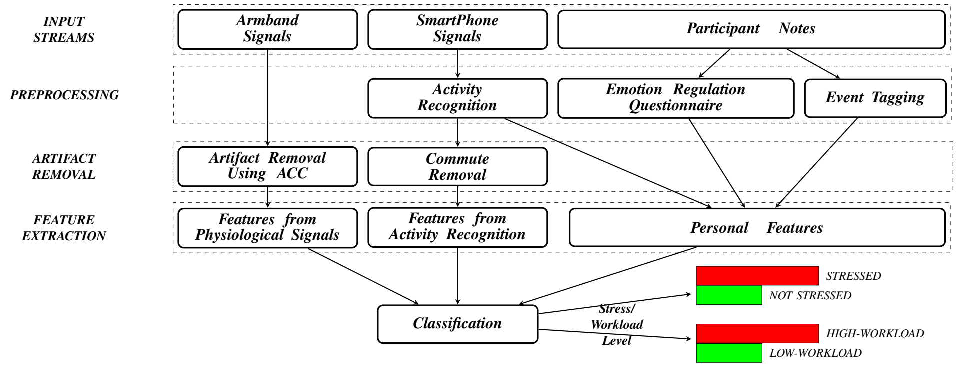

Fig. 2: End-to-end signal processing pipeline & classification system which shows the steps starting from the signal acquisition to the classification.

features extracted from the motion profile of the subject, the emotion regulation questionnaire and the daily subject feedback.

A. Data Preprocessing

Wrist based physiological sensing devices are affected by artifacts due to local or gross motion, so they need to be preprocessed before feature extraction. As a first step we recognize the activity profile of the subject, then use this for artifact removal and signal estimation.

Activity Recognition and Artifact Removal - The activity recognition model is learned using the approach, data and features of [11]. The authors have shown that by combining the features of the onboard inertial sensors of a smartphone, it is possible to make accurate predictions about a subject's activity even at low sampling rate. We identify six activities - *Walking*, *Standing*, *Sitting*, *Driving*, *Travelling by bus*, and *Travelling by Train* and segment our data with continuous labels. For every 3 minutes, we take the majority of the recognized activities and label the 3 minute segments as belonging to that activity class. We combine the three activity classes *Driving*, *Travelling by bus*, and *Travelling by Train* into a single class *"commute"*. Using this algorithm, for the collected user data, we recognized 88.5 hours of sitting, 16.6 hours of walking and 20.3 hours of standing and 79.3 hours of commute. We consider the periods for which the subject was either walking, sitting or standing for our stress and workload recognition tasks. We use the duration of commute to create a daily **Activity Profile** of the subject. We use these features later as the **personal features**. This reduces the number of motion artifacts arising out of gross body motions which occur during commute (specially in a bus or train).

Local fidgeting of the wristband, and motion artifacts arising out of posture and activity are common sources of artifacts experienced in a wearable monitoring system. Local artifacts can be identified by looking for abrupt changes in the accelerometer which are concurrent with the abrupt changes in all the physiological signals. When we identify such segments, we remove the next 30 seconds of the data from the time of the artifact initiation.

B. Feature Extraction and Machine Learning

Physiological Features - We use the GSR, PPG data, IBI and skin temperature signals from the Empatica E3 wristband for our stress detection algorithm. The Galvanic Skin Response (GSR) which refers to the variation in the electrical properties of the skin due to the action of the sweat glands, is arguably one of the most useful indicators of the sympathetic arousal which controls the human "fight-or-flight response" [12]. The time series of GSR comprises of two basic components: the ***Skin Conductance Level (SCL)*** which is a slow moving, tonic, habituating background signal component and the ***Skin Conductance Response (SCR)*** which is a rapid short-lasting change in the GSR caused mainly due to Neuronal Activity. After removing per subject baselines and normalizing the GSR values in the [0,1] range, we extract a total of 24 features from the SCL and SCR.

We smooth the PPG data provided by the Empatica E3 wristband to derive the Blood Volume Pulse (BVP) signal, and extract 8 statistical features from this BVP signal stream. The Empatica E3 wristband reports the Inter Beat Interval (IBI) at a discrete rate. The IBI or the N-N interval, is the time difference between two normal sinus beats. The heart rate can be calculated from this signal as $HR = 60/IBI$.

Heart Rate Variability (HRV) is an important measure of stress and mental activity of a person. We extract well explored HRV features such as SDNN (Standard Deviation of the N-N interval), pNN50 (percentage of consecutive N-N intervals which differ by more than 50 milliseconds), RMSSD (root mean square of the successive difference of N-N interval) [13], [14], [5] among others, extracting a total of 10 features from this IBI signal stream. We extract a further 8 statistical features from the Skin Temperature (ST) signals.

Inertial Sensor Features - The Empatica wristband has an accelerometer to calculate the tri-axial acceleration. We also have at our disposal the accelerometer and gyroscope channels from the iPhone. From each of these channels these we extract 10 statistical features (including mean, SD, min, max, number of peaks per minute).

Personal Features - The personal features for each subject comprise of three sets of features. The first source is the activity profile of the user. The second set of features is derived from the Emotion Regulation Questionnaire which the

Signal Streams	Avg F-measure
BVP	0.74
GSR	0.79
IBI	0.69
ST	0.44
Inertial	0.56
BVP + GSR	0.82
IBI + BVP + GSR	0.89
IBI + BVP + GSR + Inertial	0.82
IBI + BVP + GSR + Personal Features	**0.91**

TABLE I: Classification results for individual and best combinations of features for **perceived stress** using Random Forest Algorithm with LOSO evaluation.

Signal Streams	Avg F-measure
BVP	0.63
GSR	0.44
IBI	**0.73**
ST	0.37
IBI+BVP	0.69
Inertial	0.72
IBI+Inertial	**0.78**
IBI + BVP + Inertial	0.71
IBI + Inertial + Personal Features	0.75

TABLE II: Classification results for individual and best combinations of features for **perceived workload** using Random Forest algorithm with LOSO evaluation.

subjects had filled in. We calculate the Emotion Suppression (ES) and Cognitive Reappraisal (CR) scores for each subject. The third source is the event tagging done by the subjects. We consider the counts of the beverage, caffeine and alcohol intake which was reported by the subjects using the agent application.

Machine Learning - To perform classification, we formulate our daily stress and workload recognition as two independent binary classification tasks. Stress is classified into two classes as "Stressed" and "Not Stressed", and workload as "High Workload" and "Low Workload". We use a "Leave One Subject Out" (LOSO) cross validation scheme for all classification tasks. We perform classification on both individual and combined signal streams. For signal stream combination we perform a feature level fusion of the physiological features with the features from the activity profile and the personal features of the subjects.

We use the WEKA implementation of the Random Forests algorithm for all classification tasks. The Random Forests algorithm, which was introduced by Breiman in [15], is an ensemble learning method and is a conglomeration of tree-based classifiers. The results of classification of the level of stress and workload are reported in Tables I and II respectively. We report the classification results for individual signal streams and the results for the best combinations. We observe that a combination of physiological and personal signals give the highest F-measure for the stress classification task. Combining the physiological features with person specific features, we arrive at a high value of F-measure of 0.91. However, adding inertial features leads to a drop in performance. From Table II we observe that individually, features extracted from the IBI stream are the best indicators of perceived workload (0.73). Combining them with the features from the inertial sensors provide an improvement (0.78) in classification performances. The personal features which were indicative of stress, do not provide any improvement in the workload classification task.

IV. Conclusion and Future Work

In this paper we demonstrate a method to continuously track and measure stress and workload in naturalistic settings which can be deployed for on-the-go acquisition and monitoring of subjects. We show that the performance of the combinations of weak signal streams is greater than that of individual signals for predicting stress and workload.

References

[1] J. Lim, W.-c. Wu, J. Wang, J. A. Detre, D. F. Dinges, and H. Rao, "Imaging brain fatigue from sustained mental workload: an asl perfusion study of the time-on-task effect," *Neuroimage*, vol. 49, no. 4, pp. 3426–3435, 2010.

[2] M. Irie, S. Asami, S. Nagata, M. Miyata, and H. Kasai, "Relationships between perceived workload, stress and oxidative dna damage," *International archives of occupational and environmental health*, vol. 74, no. 2, pp. 153–157, 2001.

[3] H. Fliege, M. Rose, P. Arck, O. B. Walter, R.-D. Kocalevent, C. Weber, and B. F. Klapp, "The perceived stress questionnaire (psq) reconsidered: validation and reference values from different clinical and healthy adult samples," *Psychosomatic medicine*, vol. 67, no. 1, pp. 78–88, 2005.

[4] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 6, no. 2, pp. 156–166, 2005.

[5] B. Cinaz, R. La Marca, B. Arnrich, and G. Tröster, "Towards continuous monitoring of mental workload," in *ACM UbiComp*, 2010.

[6] J. Hernandez, R. R. Morris, and R. W. Picard, "Call center stress recognition with person-specific models," in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 125–134.

[7] B. Filipová, M. Penhaker, O. Vlcek, and M. Cerny, "Mobile phone application in biotelemetry," in *Proceedings of International Conference on Software and Computer Applications (ICSCA 2011)*, 2011.

[8] A. Sirikham, "Abnormal heart rate detection device warning via mobile phone network," in *Multimedia Computing and Systems (ICMCS), 2011 International Conference on*. IEEE, 2011, pp. 1–6.

[9] S. Balzarotti, O. P. John, and J. J. Gross, "An italian adaptation of the emotion regulation questionnaire," *European Journal of Psychological Assessment*, vol. 26, no. 1, pp. 61–67, 2010.

[10] J. J. Gross, "The emerging field of emotion regulation: An integrative review." *Review of general psychology*, vol. 2, no. 3, p. 271, 1998.

[11] A. Ghosh and G. Riccardi, "Recognizing human activities from smartphone sensor signals," in *Proceedings of the 22nd ACM International Conference on Multimedia*, 2014.

[12] A. S. Jansen, X. Van Nguyen, V. Karpitskiy, T. C. Mettenleiter, and A. D. Loewy, "Central command neurons of the sympathetic nervous system: basis of the fight-or-flight response," *Science*, vol. 270, no. 5236, pp. 644–646, 1995.

[13] N. Meshkati, "Heart rate variability and mental workload assessment," *Advances in Psychology*, vol. 52, pp. 101–115, 1988.

[14] P. Nickel and F. Nachreiner, "Sensitivity and diagnosticity of the 0.1-hz component of heart rate variability as an indicator of mental workload," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 45, no. 4, pp. 575–590, 2003.

[15] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.