What's your current stress level? Detection of stress patterns from GSR sensor data

Jorn Bakker, Mykola Pechenizkiy, Natalia Sidorova
Department of Computer Science
Eindhoven University of Technology
Eindhoven, P.O. Box 513, 5600MB,The Netherlands,
Email: {j.bakker,m.pechenizkiy,n.sidorova}@tue.nl

***Abstract*—The problem of job stress is generally recognized as one of the major factors leading to a spectrum of health problems. People with certain professions, like intensive care specialists or call-center operators, and people in certain phases of their lives, like working parents with young children, are at increased risk of getting overstressed. Stress management should start far before the stress start causing illnesses. The current state of sensor technology allows to develop systems measuring physical symptoms reflecting the stress level. In this paper we (1) formulate the problem of stress identification and categorization from the sensor data stream mining perspective, (2) consider a reductionist approach for arousal identification as a drift detection task, (3) highlight the major problems of dealing with GSR data, collected from a watch-style stress measurement device in normal (i.e. in non-lab) settings, and propose simple approaches how to deal with them, and (4) discuss the lessons learnt from the conducted experimental study on real GSR data collected during the recent field study.**

I. Introduction

Stress at work has become a serious problem affecting many people of different professions, life situations, and age groups. The workplace has changed dramatically due to globalization of the economy, use of new information and communications technologies, growing diversity in the workplace, and increased mental workload. In the 2000 European Working Conditions Survey (EWCS) [12], work-related stress was found to be the second most common work-related health problem across the EU. 62% of Americans say work has a significant impact on stress levels. 54% of employees are concerned about health problems caused by stress. One in four employees has taken a mental health day off from work to cope with stress (APA Survey 2004).

Stress can contribute to illness directly, through its physiological effects, or indirectly, through maladaptive health behaviors (for example, smoking, poor eating habits or lack of sleep) [4]. It is important to motivate people to adjust their behavior and life style and start using appropriate stress coping strategies. So that they achieve a better stress balance far before increased level of stress results in serious health problems.

The avoidance of stress in the everyday working environment is impossible. Still, if people are *informed of their stress levels*, they become empowered for taking some preemptive actions in order to alleviate stress [16].

Fig. 1. *Stress@work in a nutshell: stress detection, prediction and coaching*

There are a number of factors that are likely to cause stress at work including but not limited to long work hours, work overload, time pressure, difficult, demanding or complex tasks, high responsibility, lack of breaks, conflicts, underpromotion, lack of training, job insecurity, lack of variety, and poor physical work conditions (limited space, inconvenient temperature, limited or inappropriate lighting conditions) [10].

In [1] we proposed the conceptual framework for managing stress at work. One very important step in the process of stress management is making the worker aware of the past, current or expected stress. We aim at the automation of the identification of the stress causes of an employee in question, as well as the identification of the common causes of stress for employees within an organisation. Figure 1 shows the main ideas of our approach: We aim at making stress and stressors visible by (1) keeping track of the calendar events and daily routine of the worker, (2) measuring stress-related physiological signs from the sensor data, (3) annotating these events with the sensor data and the results of automated analysis of additional information sources, such as sentiment classification of the incoming and outgoing e-mails or social media messages [18] and explicit

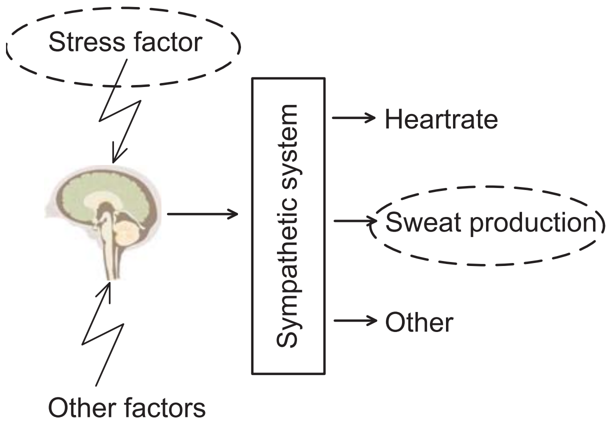


Fig. 2. *The reaction to stress factors is governed by the autonomous nervous system. This path is shared with a lot of other mechanisms.*

user feedback, (4) extracting the relationship between event data and sensor data, i.e. relations between the increases and decreases in the stress level with the characteristics of the events of daily lives (what, where, when, with whom, etc.), and (5) using extracted knowledge about this relationship for personalized coaching.

In order to find this relationship, a number of subtasks need to be done. One of the main subtasks is *detecting stress from the sensor data*. Due to modern ICT and sensor technologies, objective measuring of the stress level in non-lab settings becomes possible. Such symptoms as voice, heart rate, galvanic skin response (GSR) and facial expressions are known to be highly correlated with the level of stress a person experiences [3], [5], [7]. In this paper we focus on the use of the GSR data (reflecting sweating) measured by a prototype device worn at a wrist.

The direct use of the GSR measurements obtained is not that straightforward. Partly this is caused by noise and inaccuracies in the collected sensor data, but what is more crucial – the reaction to various stress factors is governed by the autonomous nervous system and this "path" to the symptomatic system is shared with a lot of other mechanisms, such as the mechanism of adaption to the outside temperature and humidity (Figure 2).

We have conducted a pilot case study aimed at the identification of likely challenges we need to address to make our approach work in practice. In this paper, we focus only on the problem of detecting changes in the stress level from the GSR sensor data alone. We study the peculiarities of noise and disturbances in the signal and argue the need of the related contextual data for improving the quality of stress detection.

The rest of this paper is organized as follows. In Section II, we formulate the problem of stress identification and categorization from the sensor data stream mining perspective. We focus on a subproblem of arousal identification in online settings, which we formulate as a drift detection task. We highlight the major problems of dealing with GSR data, collected from a watch-style stress measurement device in normal (i.e. in non-lab) settings, and propose simple approaches how to deal with them. In Section III we present the results and lessons learnt from the conducted experimental study on real GSR data collected during the recent pilot field study. Finally in Section IV we give conclusions and discuss directions for further work.

II. Acute stress identification

Stress comes in three flavors:

1) *Acute*: stress caused by an acute short-term stress factor.
2) *Episodic acute*: acute stress that occurs more frequently and/or periodically.
3) *Chronic*: stress caused by long-term stress factors and can be very harmful in long run.

Most people experience acute stress during their everyday life. It is a primal flight-or-fight response to immediate stress factors and is not considered harmful. When the frequency of these occurrences increase, physiological symptoms might occur. This type of stress is associated with a very busy and chaotic life and can be considered to be harmful when it occurs over prolonged periods of time. The last type of stress, chronic, is considered to be the most harmful. Prolonged periods of stress could be caused by personal circumstances or other long-term factors.

In our work, we want to prevent people from transferring to the chronic category and therefore, we target the acute and episodic acute stress. Particularly, in this paper we focus on the *identification of acute stress* in order to facilitate coaching of the episodic acute stress.

Acute stress is a mechanism that brings the body into a state of alertness. As shown in Figure 2, it is controlled by the autonomous nervous system. This system maintains a constant equilibrium (also known as homeostasis). A change in this equilibrium results in different changes in the bodily functions (e.g. activity of digestion system).

Stress can be seen as a state of emergency that is preceded by arousal due to an external stimulus, see Figure 3. After the factor causing stress (the stressor) disappears, the body relaxes and returns to a normal state.

Figure 4 shows the general case with more relationships between the four states depicting the inner process of stress.

The problem of stress identification can be formulated in different ways, e.g. as a traditional classification task, as one-class classification, as event identification, and as time series subsequence classification to name a few main options.

It should be also noticed that acute stress can also be positive (e.g. caused by an excitement or an intrinsic motivation or an engagement in the working process), and, consequently, staying in a normal state for too a long period without any acute stress can be a sign of monotone uninteresting work or poor motivation of the employee. Therefore, we would like to perform a more detailed classification of the states in the future.

In this paper we consider a simplified setting assuming that a person is either in the *normal* state or in a *stressed* state. The change between the two states can be sudden or incremental; typically, arousal is more rapid and relaxation

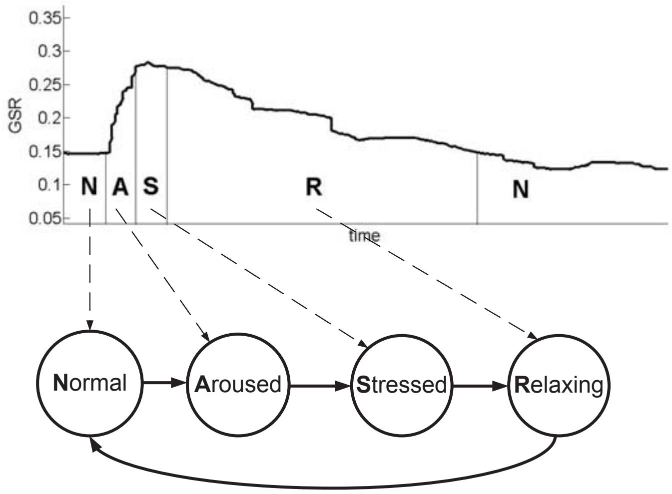


Fig. 3. An example of acute stress pattern observed from GSR data and how it can be mapped to the symbolic (time-stamped) representation of person's stress.

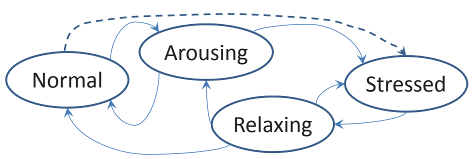


Fig. 4. Four states depicting the inner process of stress.

takes considerably longer. As we will show, different change patterns can be observed.

A. Arousal as change detection

The principal task is to detect whether a person is stressed at a particular moment in time or not. In other words, the detector assigns a label "stressed" or "not stressed" based on the observed historic data.

Detecting changes in GSR data is not as straightforward as someone might think looking at the example in Figure 3. Different types of noise in the data and changes in GSR data due to other factors than stressors make it a non-trivial task. In this section, we give illustrative examples of noise and other factors affecting the GSR signal.

Types of noise. The quality of the GSR signal depends primarily on the continuity of the contact between the device and the skin of the test person. The skin conductance is measured by two electrodes that require skin contact in order to produce a reliable signal. However, this contact is not the same for every person. For some people, the device fits less well (e.g. because they dislike wearing it tight enough to guarantee good contact, or because they have very dry skin); due to a poor fit, we get noise in the signal (see Figure 6). A person might also accidently touch the device (or do this periodically in case of having such a habit), thus increasing the pressure and influencing the GSR measurement; this also

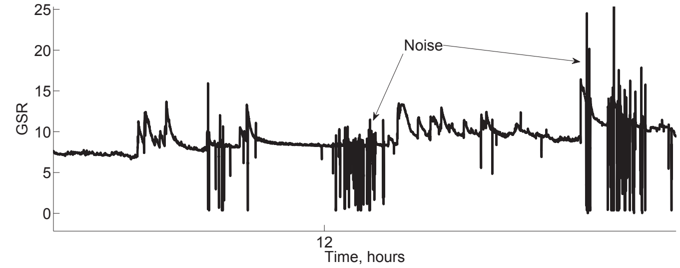


Fig. 5. The GSR signal contains two-sided local noise peaks that are probably caused by a physical disturbance of the contact between the skin and the sensors, e.g. if someone has a habit to touch from time to time the watch or the stress meter in this case.

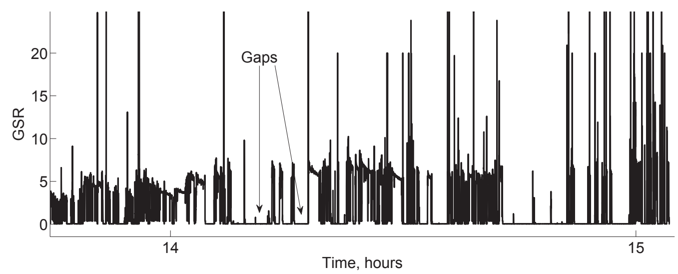


Fig. 6. When the fit between the skin and the sensors is not tight enough, the contact is continuously broken. A characteristic of this behavior is the high amount of gaps (ground value of sensor) in the signal.

creates noise in the signal in the form of gaps (see Figure 5).

Note that the skin in contact with the device contains slightly more sweat than the skin next to the device, and when the device is shifted on the skin, there is a resettling period of about 15 minutes during which the skin that came in contact with the device gets about the same level of sweat as the skin that was in contact with the device before the shift, thus resulting in about the same GSR (under assumption that no change in the stress level happens in this period).

Importance of context. There are a lot of different factors that influence the internal state of a person. Rising GSR levels might be related to a rise in temperature or to heavy physical work or exercises. In other words, the GSR change patterns can be related to contexts that are mostly hidden.

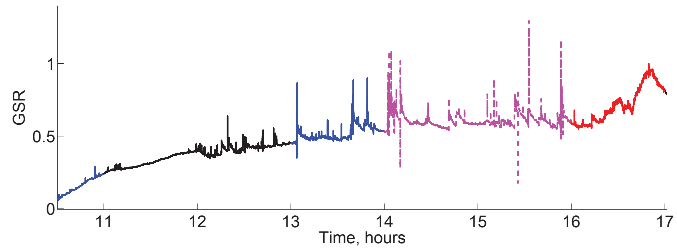


Fig. 7. Prior to a stressful event (red-lined peak), the GSR level is gradually rising. Is this rise caused by an external factor or is it due to anticipation of the event?

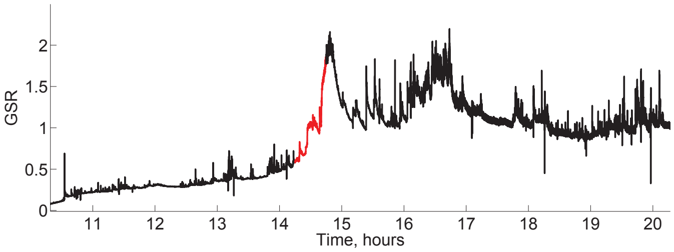


Fig. 8. After a stressful event (red-lined peak) the GSR level does not return to the level it had prior to the event. This might indicate that there is no relaxation process.

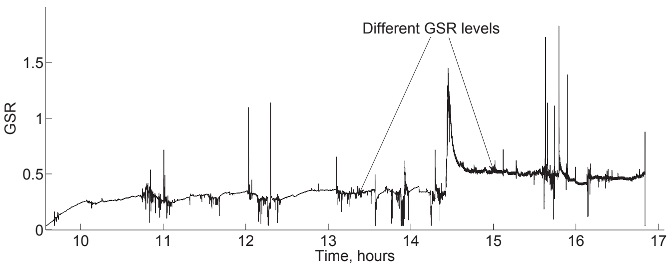


Fig. 9. After a suspected stressful event the GSR level does not return to the level it had prior to the event. This might be an indicate that there is no relaxation process or what is more like in this case - the baseline level of GSR corresponding to normal unstressed state changed.

One of these patterns is a steady increase of the GSR level (see Figure 7). This might be an indication of changing environmental factors (e.g. temperature), but it might also be a genuine stress response. For instance, once a certain event has been scheduled, the person might get stressed in anticipation of the event. This is an interesting pattern for the stress detection task.

The same holds for the patterns in Figures 8 and 9. In these time series there is a suspected stress peak: in Figure 8 the red part corresponds to an event tagged by the user as being stressful, in Figure 9 there is an untagged short-term increase in the GSR level. In both cases, the GSR level does not return to the original baseline after passing the peaks. The question is whether this is due to continuous stress (because of the user being still busy with what has happened) or some other factors.

For some series we learnt from the users' feedback that certain patterns were caused by environmental factors or user activity context. In Figure 10 the person is exercising between 12:00hr and 13:00hr. The effect of the exercises is clearly visible in the GSR time series. Moreover, due to the form and the intensity of the picks, we can discriminate those from genuine stress.

These context-dependent patterns will be important in the overall stress detection task. Knowing whether a person relaxes after a stressful event or whether he or she experiences anticipating stress is very important. Here we do not handle these contexts explicitly.

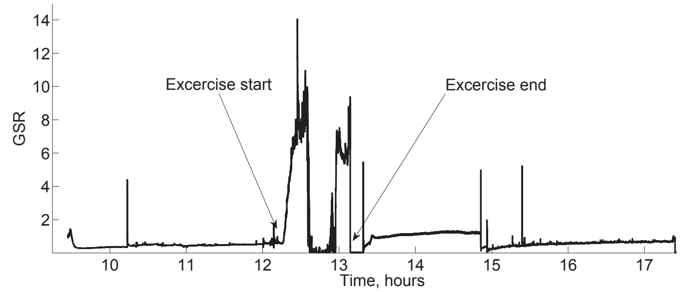


Fig. 10. Doing physical exercises results in a high GSR level, yet is not related to the emotional stress.

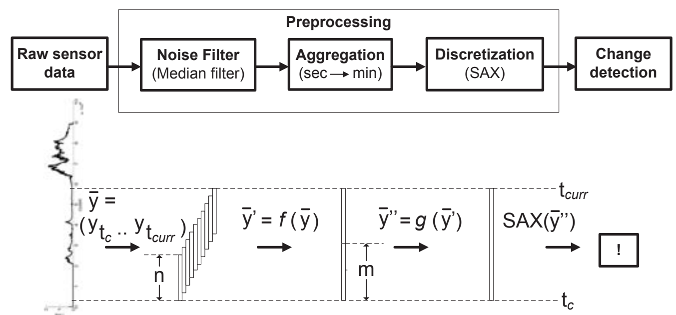


Fig. 11. Arousal detection approach: the GSR data is first (1) filtered, (2) aggregated, and (3) discretized in the preprocessing phase and then passed to a change detection technique. Each step is applied to a window of data that is kept until a change has been detected.

B. Approach

The main task is to determine whether the observed portion of the signal contains a change that corresponds to an arousal. Formulating this problem as a change detection task on univariate time series, we consider a four step approach for arousal detection as shown in Figure 11. In the preprocessing phase we take the raw GSR sensor data and according to the operational settings (i.e. offline vs. online) perform its filtering, aggregation and discretisation. The processed data is served to a change detection technique.

The purpose of arousal detection can be twofold. The first is to obtain labels for the supervised learning process aimed at finding relationships between stress occurrences and external events of factors causing stress. In this case we can perform change detection in offline settings, i.e. the complete data series can be used in preprocessing and detection steps. The second purpose is to use an online detection mechanism in online or semi-online settings as an alarm for making the user aware of stress (and possibly asking for feedback that can be related back to the subjective labeling process, i.e. the user can confirm or reject the alert). Although we do not fix the purpose of the task in this paper, we only describe an online method that detects arousal for the point in time that might be as much as a minute in the past.

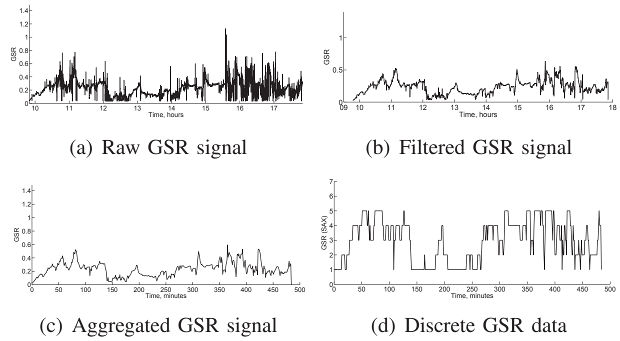
Preprocessing. The three preprocessing steps that we use are shown in Figure 11 and exemplified for the illustrative purposes in Figure 12. The main objective of the preprocessing phase is to remove noise from the GSR time series. The first type of noise is due to poor contact between the sensors and the skin (see Figure 6). If the contact is not sufficient, the sensor will not measure anything. The second type of noise is a local disturbance of the signal (see Figure 5). These local disturbances are caused by mechanical movements (e.g. user bumps device onto something) and should not be considered to be actual measurements.

Noise caused by contact loss is problematic, since we cannot be sure whether the signal can be trusted in these areas. In these cases the frequency of the ground value (i.e. when the sensors are not measuring anything) is a lot higher than in a normal time series. When such periods occur in the GSR signal, we alarm the problem and do not consider it further in the arousal detection task. More specifically, we count the number of occurrences of these faulty measurements and exclude the time series if this number exceeds the number of other points.

Noise caused by local disturbances must be filtered out because they might be mistaken for genuine peaks. As shown in Figures 3, one of the important parts of the arousal detection task is to catch the transition from normal GSR levels to aroused levels. This transition is characterized (for a typical stress pattern) by a sudden peak in the GSR level. The filter should filter out local disturbances while maintaining the typical peaks. Therefore, the noise is filtered out by using a median filter [14]. This is a filter that is used in image processing, and it preserves edges (opposed to e.g. a moving average) while filtering out noise.

The preprocessing step is applied to windows within the window of kept data. Let $\bar{y} = (y_{t_c}, \ldots, y_{t_{curr}})$ be the portion of kept data from either the start or the last change point (y_{t_c}) until the most recent sample ($y_{t_{curr}}$). The filter computes the filtered values $\bar{y}' = f(\bar{y})$ over a moving window of size n ($n = 100$ in the experiments) from $\bar{y}_1$ until $\bar{y}_k$, where f is the filter function and $k = t_{curr} - t_c$. Each consecutive block of m samples in $\bar{y}'$ is aggregated to one value, $\bar{y}'' = g(\bar{y}')$. In the last preprocessing step, this data is discretised using SAX [8] into a discrete time series from 1 to 5, SAX($\bar{y}''$). The levels can be interpreted as being levels of stress (1: completely relaxed and 5: maximum arousal). However, they should not be interpreted as absolute levels of arousal, but rather as a local relative measure of arousal. Please, notice that discretisation of the time series does not lead to an easy identification of the change points (see Figure 13 for an illustrative example. However, the dicretisation can help the change detector to be more accurate.

The signals are measured with a sampling frequency of 4 Herz, yet it does not make sense to expect the stress detection to have timing requirements in the order of tenths of seconds. For this reason, we aggregate the data to the order of minutes. We use $m = 240$ in the experiments, thus after the aggregation step 1 sample point $\bar{y}''_i$ corresponds to 1 minute. In the experiments, we took $\bar{y}''_i = max(\bar{y}'_{block_i})$. As said, in



(a) Raw GSR signal (b) Filtered GSR signal

(c) Aggregated GSR signal (d) Discrete GSR data

Fig. 12. An example of GSR signal in its original form and after each of the three individual steps in the data preprocessing: the raw GSR signal shown in (a) is filtered using a median filter (b), then the values are aggregated to the minute level (c), and finally they are discretised using SAX encoding (d) to be used as an input for a change detection technique.

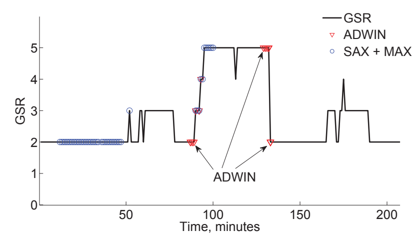


Fig. 13. An illustration that discretising the data with SAX does not immediately give us information about a change in arousal, e.g. by taking a maximal value of the current window. The blue circles indicate the changes alerted by such an approach. The red triangles indicate the change points alerted by the ADWIN change detection method taking the SAXified time-series as an input. ADWIN is considered below.

this setup it is important that the aggregation step is applied after the filter in order to avoid the influence of local noise.

The discretisation using SAX is done online in a progressive way. That is, the SAX representation is recomputed over the historic data as new instances come within the training window.

Change detection. Change detection in time series has been a topic of interests in different domains. Existing approaches can be divided into two broad groups of techniques. Techniques from the first group are based on monitoring the evolution of performance indicators like classification model accuracy or some property of the data. Cumulative Sum (CUSUM), introduced in [11] and recently used in [17] is one of the statistical process monitoring mechanisms. This method monitors the mean of the input data (that can be also any filter residual) and gives an alarm when it is significantly different from zero, i.e. deviates from the normal process behaviour. Other methods rely on time series forecasting techniques such as Neural Networks and Auto Regression functions [15] that estimate parameter changes online based on an offline mapping.

Techniques from the second group are based on monitoring distributions on two different time-windows: a reference window summarizing past information and a window over the most recent examples. Statistical tests based on Chernoff bound, which decide whether samples drawn from two probability distributions are different, were studied in [6]. ADaptive WINdowing (ADWIN) [2] that we use in our experimental study keeps a variable-length window of recently seen data points. It tries to keep the window of the maximal length that is still statistically consistent with the hypothesis that there has been no change in the mean signal value inside the window.

Thus, we consider two different approaches for change detection. Both approaches are aimed at finding statistically significant changes in data. The first approach that we call here *Fit* is based on monitoring the model error, and the second approach ADWIN is based on monitoring the data signal itself. Both approaches were recently used for change detection in the task of online prediction of the fuel mass flow in a boiler [13].

Fit: Performance monitoring-based change detection with the non-parametric test. In this study we assume that the general pattern of arousal resembles the curve as shown in Figure 3. We also assume that there is no global model that predicts the general GSR signal for a person. Instead of using a global model in combination with statistical change detection methods, we opt for a method that computes local models.

If we assume that the stress level of a person is stable in between changes, the changes can be detected by monitoring the error of a locally fitted model. Given historic (preprocessed) data, the objective is to fit a simple regression model.Based on the observed Mean Squared Error for the incoming points, we can apply a statistic measure (e.g. Mann Whitney U test [9]) to determine whether a significant change in the prediction error has occurred.

Every time a new point arrives, the data is split into two sets. The first set is a reference set that excludes the new point. The second set is a test set that includes the new point. For each of the two sets a model is trained while iteratively leaving out one of the points. When there is an overall significant difference between the two sets, it is considered to be a change point and a cut is made.

ADWIN: Change detection based on raw data using adaptive windowing. ADWIN method works as follows: given a sequence of signals it checks whether there are statistically significant differences between the means of each possible split of the sequence. If a statistically significant difference is found, the oldest portion of the data backwards from the detected point is dropped and the splitting procedure is repeated until there are no significant differences in any possible split of the sequence. More formally, given the GSR data stream, suppose a_1 and a_2 are the means of the two subsequences as a result of a split. Then the criterion for a change detection is $|a_1 - a_2| > \epsilon_{cut}$, where

$$\epsilon_{cut} = \sqrt{\frac{1}{2a} \log \frac{4k}{\delta}}, a = \frac{1}{\frac{1}{k_1} + \frac{1}{k_2}}, \tag{1}$$

here k is total size of the sequence, while k_1 and k_2 are sizes of the subsequences respectively.

TABLE I
DATA SET SUMMARY.

Number of users	5
Number of time series	72
Time series per user (mean)	14
Mean length (samples)	98721
Number of change points overall	368
Mean change points per series	6.5

III. EXPERIMENTAL STUDY

In this section we present the results from the conducted experimental study on real GSR data collected during the recent pilot field study. First, we give a concise description of the constructed dataset and experiment setup, and then provide a summary of the quantitative evaluation and some highlights of the qualitative analysis of interesting cases.

Dataset description. Table I summarizes the main characteristics of the data set. The data consists of the GSR data measured on five persons in the course of the four weeks. The data was collected from a watch-like device worn by the persons during working hours. Since the sampling rate is 4 Hz and the typical working day is roughly 8 hours, the average length of the raw time series is 98721. All together the data set contained 72 time series. 26 time series were excluded from the experiments for either of the two reasons: the GSR level showed very low variation or the contact of the sensors was not sufficient to yield a usable signal (these were detected automatically by a filter and then verified by the visual inspection).

For each of the remaining 56 time series we annotated the change points based on the visual inspection. Overall the set of time series contains 368 change points with an average of around 6.5 change points per time series.

The users participated in the study were instructed to annotate any meeting in their agenda (MS Outlook Calendar) with information about their feeling towards the meeting ("nice", "exciting", "neutral", "annoying", or "tense"). Although this information was available, it was not used in this investigation. The reason for this is that the primary objective in this work is to detect GSR peaks; however, a lot of the peaks do not correspond to any meeting recorded in the agenda. Moreover, the actual stress related to a meeting does not necessarily shows up at the time of the meeting. It might precede the event (see Figure 7) or continue to influence the person afterwards (see Figure 8). In the ideal case, these labels reflect the state transitions as shown in Figure 4, but in reality it is hard to discern the separate state changes.

Therefore, instead of using the working agenda annotations provided by the users, we used manually added labels based on the visual inspection of the GSR time series. In the experimental study presented in this paper, we labeled only the change points, i.e. from the problem formulation perspective, each point is labeled to be either a change point or not – that our arousal detection approach will try to detect based on the already observed GSR values.

TABLE II
TP AND FP RATES OF DETECTING THE CHANGE POINTS. THE MEAN μ VALUES ARE PERCENTAGES WITH RESPECT TO PERFECT DETECTION.

	$\mu(\frac{TP}{P})$	$\sigma(\frac{TP}{P})$	$\mu(\frac{FP}{TP+FP})$	$\sigma(\frac{FP}{TP+FP})$
Fit	0.66	0.16	1.66	0.16
ADWIN	0.08	0.01	1.01	0.1

TABLE III
THE DISTANCE BETWEEN THE TIME OF THE ACTUAL CHANGE (t_a) AND THE TIME OF THE DETECTION (t_d).

	$\mu(\|t_a - t_d\|)$	$\sigma(\|t_a - t_d\|)$
Fit	2.8	0.54
ADWIN	2.5	1.2

Experiment setup and evaluation. On each of 56 time series we perform three steps: preprocess the data as discussed in the previous section (see Figures 11 and 12), apply each of the change detection methods, compare the labels to the changes signalled by the method.

The techniques are applies on each time series in a progressive way. That means that we assume that the data arrives as a stream (one point at the time). Historic data is kept until a change point is suspected. After that a new window is created from the change point onwards.

The change points are evaluated by measuring the distance between the point identified by a detection algorithm as a change point and the closest actual change point within a preset boundary threshold. The reason for doing this is that there is no strict requirement that a change point should be detected at exactly the point where it occurs. We should allow for some leniency with respect to the actual time where it is detected. Therefore, we measure the True Positive rate within a window of 5 minutes around the actual change point. Instead of the False Positive rate, the False Discovery rate is reported, since the amount of True Negatives is very large with respect to the True Positives.

Results. The results of the experiments are shown in Tables II and III. As can be seen from Table II none of the methods was able to catch all of the change points. The *fit* method detected more change points than *ADWIN*, but at a cost of more False Positives. The positioning of the change points is better handled by *ADWIN*.

In Figure 14 and 15 there are a lot of False Positives in the beginning of the time series. This is probably due to the online encoding. In the beginning, if the signal is flat, small fluctuations are blown-up by the discretization step. This might lead to more False Positives. Yet the *fit* method shows this behavior along the whole length of the time series.

There are two reasons why the True Positive rate is low for *ADWIN*. The first is that it does not detect small peaks. The second is that it also does not detect the change in cases where the signal is slowly rising or falling (like in Figure 17).

Although we did not study thoroughly the effect of the preprocessing techniques on the performance of the change detection methods, some examples indicate that when time-

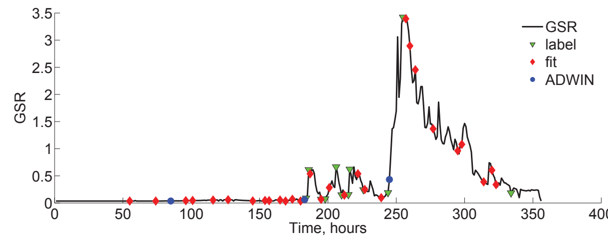


Fig. 14. A flat signal followed by a high peak. On the down-curve of the high peak there are many smaller peaks that are more difficult to detect.

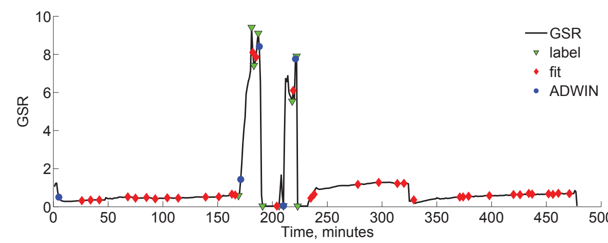


Fig. 15. One of the stress time series and the change points. Green triangles depict the ground truth, red diamonds depict the detection of the *fit*-method, and the blue circles depict the detection of *ADWIN*.

series is filtered and then aggregated, but not discretised with SAX, change detection may become less accurate (e.g. in Figure 18 ADWIN missed two change points; cf. ADWIN in Figure 16).

Discussion. The main difficulty of the stress detection task is that arousal comes in many different forms. Since the experiments were done in uncontrolled settings, it is difficult

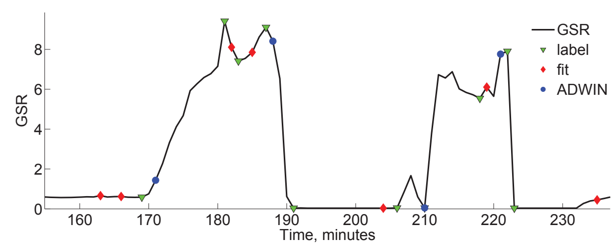


Fig. 16. Closeup of the time series in Figure 15. *ADWIN* clearly detects the high peaks, whereas the *fit* method is more sensitive to small local changes.

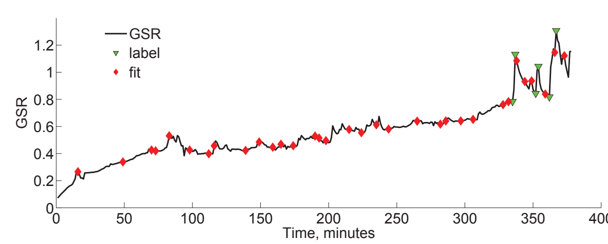


Fig. 17. Steadily increasing signal is not detected by *ADWIN*, yet there are a lot of False Positives from the *fit* method.

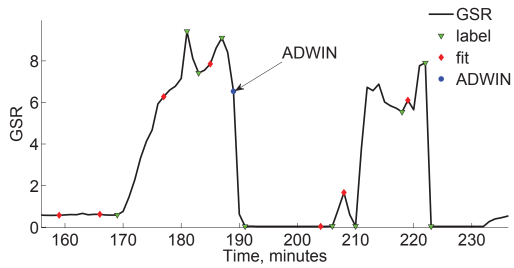


Fig. 18. Detection results of *ADWIN* and *fit* on the same time series as in Figure 16, but without SAX discretisation in preprocessing.

to interpret the patterns in the data. The manual labels are not arbitrary, but their interpretation in terms of real arousal is difficult.

Many examples suggest us that interpretation of the GSR data can be rather ambiguous and deciding whether a particular observed pattern corresponds to stress or something else (like a physical exercise) is a non-trivial task even for a human expert. (We asked the domain expert to analyze GSR curves like the ones presented in the paper and he had confirmed that they were ambiguous and additional information was required to make a confident judgement whether the peaks correspond to genuine stress or they are results of other factors). Therefore, even "ideal" noise-free GSR data may be insufficient for accurate determining the level of stress. This suggests that the reliable translation of physiological data gathered by using sensor technology into the "stress level rates" is only possible when additional sources of information are available. For example, apart from the GSR measurements, we can also use measurements of acceleration in three dimensions. Exploring the potential of accelerometer data for detecting the activity context (e.g. physical exercises, walking, active discussion etc) is an interesting direction for further research.

Other sources of additional data may include subjective user feedback collected via questionnaires, annotation of the events/signal, etc., as well as various external data extracted e.g. from the social media, e-mail correspondence or electronic agendas. Having access to such additional data facilitates the use of pattern mining for finding relations between the increases and decreases in the stress level with the characteristics of the events of daily lives (what, where, when, with whom, etc.).

IV. Conclusions and Future Work

The detection of stressful events is a challenging task. The information coming from sensor measurements is highly ambiguous and dependent on hidden contexts. The detection of separate stress peaks in the GSR data is also challenging due to the varieties of patterns in the data. Moreover, it is not clear without additional information whether certain peaks correspond to a significant physiological process and how to categorize them if they do.

In the further work, we plan to mine different sources of data for stress detection and categorization. This includes the statistics from the calendar, e-mail correspondence and social media [18].

An additional source of information is the similarities or differences between persons. Each person will handle stress in a different way, but some might share characteristics when it comes to anticipation, relaxing, or the general impact of stress on observable variables. Using these sources of data collected under more controlled settings we hope to be able to get more reliable and more fine-grained categorization of stress patterns.

Acknowledgement

This research has been partly supported by EIT ICT Labs, Health & Wellbeing thematic line (http://eit.ictlabs.eu) and NWO HaCDAIS Project.

References

[1] J. Bakker, L. Holenderski, R. Kocielnik, M. Pechenizkiy, and N. Sidorova. Stress@work: From measuring stress to its understanding, prediction and handling with personalized coaching. In *Proc. of the 2nd ACM SIGHIT International Health Informatics Symposium, IHI'12*. ACM Press, 2012.

[2] A. Bifet and R. Gavaldà. Learning from time-changing data with adaptive windowing. In *Proc. of the 7th SIAM Int. Conference on Data Mining, SDM'07*, 2007.

[3] W. Boucsein. *Electrodermal activity*. New York and London: Plenum Press, 1992.

[4] K. Glanz and M. Schwartz. Stress, coping, and health behavior. *Health behavior and health education: Theory, research, and practice*, pages 211–236, 2008.

[5] H. S. Hayre and J. C. Holland. Cross-correlation of voice and heart rate as stress measures. *Applied Acoustics*, 13(1):57 – 62, 1980.

[6] D. Kifer, S. Ben-David, and J. Gehrke. Detecting change in data streams. In *Proceedings of the International Conference on Very Large Data Bases*, pages 180–191, Toronto, Canada, 2004. Morgan Kaufmann.

[7] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, 1993.

[8] J. Lin, E. J. Keogh, L. Wei, and S. Lonardi. Experiencing SAX: a novel symbolic representation of time series. *Data Min. Knowl. Discov.*, 15(2):107–144, 2007.

[9] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Math. Statistics*, 18:50–60, 1947.

[10] S. Michie. Causes and management of stress at work. *Occupational and Environmental Medicine*, 59(1):67, 2002.

[11] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

[12] P. Paoli, D. Merllié, and F. per a la Millora. *Third European survey on working conditions 2000*. European Foundation for the Improvement of Living and Working Conditions, 2001.

[13] M. Pechenizkiy, J. Bakker, I. Žliobaitė, A. Ivannikov, and T. Kärkkäinen. Online mass flow prediction in cfb boilers with explicit detection of sudden concept drift. *SIGKDD Explor. Newsl.*, 11:109–116, May 2010.

[14] W. K. Pratt. *Digital Image Processing*. John Wiley & Sons, 1978.

[15] N. D. Ramirez-Beltran and J. A. Montes. Neural networks for on-line parameter change detections in time series models. *Computers & Industrial Engineering*, 33(1-2):337 – 340, 1997. Proc. of the 21st Int. Conference on Computers and Industrial Engineering.

[16] P. Sanches, K. Höök, E. K. Vaara, C. Weymann, M. Bylund, P. Ferreira, N. Peira, and M. Sjölinder. Mind the body!: Designing a mobile stress management application encouraging personal reflection. In *Conference on Designing Interactive Systems*, pages 47–56, 2010.

[17] M. Severo and J. Gama. Change detection with Kalman Filter and CUSUM. In *Ubiquitous Knowledge Discovery*, LNCS 6202, pages 148–162. Springer Berlin / Heidelberg, 2010.

[18] E. Tromp and M. Pechenizkiy. Senticorr: Multilingual sentiment analysis of personal correspondence. In *Proc. of IEEE ICDM 2011 Workshops*. IEEE Press, 2011.