Smart Technologies for Long-Term Stress Monitoring at Work

Rafal Kocielnik
TU Eindhoven
r.d.kocielnik@tue.nl

Natalia Sidorova
TU Eindhoven
n.sidorova@tue.nl

Fabrizio Maria Maggi
University of Tartu
f.m.maggi@ut.ee

Martin Ouwerkerk
Philips Research
martin.ouwerkerk@philips.com

Joyce H.D.M. Westerink
Philips Research
joyce.westerink@philips.com

Abstract

Due to the growing pace of life, stress became one of the major factors causing health problems. We have developed a framework for measuring stress in real-life conditions continuously and unobtrusively. In order to provide meaningful, useful and actionable information, we present stress information, derived from sensor measurements, in the context of person's activities. In this paper, we describe our framework, discuss how we address arising challenges and evaluate our approach on basis of the field studies we have conducted. The main results of the evaluation are that the results of long-term measurements of stress reveal people information about their behavioral patterns that they perceive as meaningful and useful, and trigger their ideas about behavioral changes necessary to achieve a better stress balance.
Keywords: *stress monitoring; health and wellbeing; self-awareness; sensor technologies; skin conductance*

1. Introduction

Recent statistics indicate the great impact of stress-related problems on individual lives and on the economies of different countries, making stress a pressing issue with a need for practical solutions. In the 2000 European Working Conditions Survey (EWCS) [11], work-related stress was found to be the second most common work-related health problem across the EU. In the Netherlands, 1 out of 7 disabled gets his condition because of stress at work (TNO Survey 2006).

Stress is experienced by people every day and is intrinsically related to the interplay between the environment and the person [8]. The appreciation of being over-stressed often comes too late, when health problems already manifest themselves: people's ability to recall, recognize and understand their stress may be hampered by their life style, with multiple tasks and responsibilities encountered everyday. Therefore, we aim at creating a *framework allowing people to discover their stress reaction patterns using unobtrusive monitoring technologies*.

Current approaches to stress monitoring are mostly based on applying questionnaires or carrying out individual/group meeting with psychologists. Questionnaires can be experienced as quite obtrusive and they rely mainly on people's ability to accurately recall their experiences. In practice, however, our memory does not reflect our experiences equally and is biased towards most recent events and abnormalities. Individual/group meetings with a psychologist can be very effective, but are very costly both in terms of money and time.

The physiological signals of stress, as reflected by changes in blood pressure, heart rate, pupil dilation, sweat gland activity, reflected in skin conductivity, can be objectively measured in unobtrusive ways using modern sensor technology. The rich body of research in this area indicates that these measurements can be effectively used for stress detection (see [14] for an overview). Continuous, sensor-based monitoring registering stress-related signals resolves the problem of memory bias and provides an objective record of one's stress levels. The context information about the life events of the person is nowadays omnipresent due to the use of digital calendars, social media, emails, logs of phone calls, etc. *Our hypothesis is that collecting sensor data for prolonged periods and presenting it in relation to digital life data enables the user to discover personal stress and relief patterns*.

To test our hypothesis we built a framework collecting and interpreting a variety of data to provide the user with information about the stress experienced in relation to the context in which it was experienced. We conducted field studies with user groups at two organizations to evaluate the feasibility and efficiency of our approach. In our studies we

978-1-4799-1053-3/13/$31.00

used DTI-2, a sensor wristband developed by Philips Research [15] for continuous unobtrusive monitoring of physiological signals and environmental conditions (skin conductivity, skin temperature, ambient temperature and lightings, 3D acceleration), allowing to estimate experienced stress level. We generated visual representations relating the stress levels of a user to different aspects of his/her daily activities, as specified in the digital calendar, based on the data from several weeks of study. The evaluation with users provides us with evidence in favor of our hypothesis, since users were able to discover stress patterns they were not aware of but the validity of which they estimate as very likely; they want to test it by taking actions which, they expect, will result in breaking these patterns and achieving a better stress balance.

Unlike our work, most of the works in the area of stress detection were performed in lab conditions, where the environment is strictly controlled, the stressful conditions are administered artificially and, as a consequence, do not necessarily conform to the ones experienced in real life. The measurements in lab settings are short-term and mostly static; almost all the personal, natural context of experiencing stress is removed.

Recently, there were several attempts to detect stress in real-life settings as well. In [1], we described the ideas for a framework for real-life stress monitoring, aiming at stress prediction and coaching, and the results from a pilot field study, backed by the findings on stress pattern discovery [2] in the sensor measurements performed in real life conditions. From the other approaches we are aware of, *AffectiveHealth*[1] and *AffectAura* [9] are the ones with the goals most close to ours. *AffectiveHealth* is based on a live recording of a number of physiological signals and providing the user with a visualization of the estimated stress level in the form of a dynamic spiral [13]. This approach does not provide the user with the context in which stress has been experienced, relies on the user's memory and interprets time as a linear dimension only. This limits the possibilities of discovering the relation between stress and other dimensions, like e.g. topics of meetings, people involved in those meetings, days of the week, time of the day. The *AffectAura* project also aims at emotion detection in real life environments. In this setting, although real-life, the amount and form of sensors makes the solution applicable only to office workers spending most of their time in front of their computer, and not applicable for occupations like teachers, managers, doctors, nurses, etc. Finally, there are some works focusing on one specific occupation, like drivers (see e.g. [3, 12]). Unlike occupation-specific settings, we aim at a solution applicable for a broad range of occupations.

The rest of this paper is organized as follows. In Section 2, we describe our approach. Section 3 provides the procedure used for the interpretation of the stress data. In Section 4, we discuss the conducted studies and the conclusions drawn from the evaluation of the study results by the users. Finally, in Section 5, we summarize our approach and its future potential.

2. Approach

Our approach is based on continuous unobtrusive measurement of stress for a period of time that would be long enough to reveal useful patterns related to life events of the user. Continuous measurement allows for capturing reactions to different events, avoiding the pitfalls of approaches based on taking the "snapshots" of stress picture. Since we aim at an unobtrusive solution applicable in most of the practical settings, we have chosen for three means of data acquisition: 1) a wristband producing continuous sensor measurements; 2) a calendar application for collecting information about activities in time; and 3) a short questionnaire for collecting subjective feedback (modified Self Assessment Manikin questionnaire) (see Figure 1).

The wristband we use – *The Discrete Tension Indicator (DTI-2)*, developed by Philips Research – is an unobtrusive, wearable device that combines multiple sensors measuring skin conductance, 3D acceleration, band temperature, skin temperature and ambient light (see Figure 1). Skin conductance, as measured by DTI-2, turns out to reflect stress reactions well enough to use it as basis for estimations of the stress levels of a person [15] (note that we are currently interested not in precise up-to-seconds timing of stress reaction but rather in general patterns of stress and relaxation related to different facets of a person's life). Acceleration measurements on a wrist allow for activity recognition, thus supplying context information. We ask the users to wear the device on their dominant hand, since that allows for a better recognition of certain activities, like writing. A wrist-worn device is perceived by the majority of our users as more comfortable than solutions such as chest belts for heart-rate monitoring, and not violating privacy, which is often the case with cameras used for facial expression analysis. The measurements are stored by the device on an internal SD card, and they can also be streamed live to a receiving station via a Bluetooth wireless link, which makes it appropriate for run-time feedback to the user, when desirable.

Since we are interested in patterns of stress/relaxation related, initially, to events at work (what, when, where, with whom, etc.), we make use of *calendar* information, developing further the ideas we proposed in [1]. In that work we used calendar information from *MS Outlook* supplied with the information about subjective stress levels that the users entered using categories of events (a standard option in MS Outlook, allowing the user to define his/her own cate-

[1] www.sics.se/ah

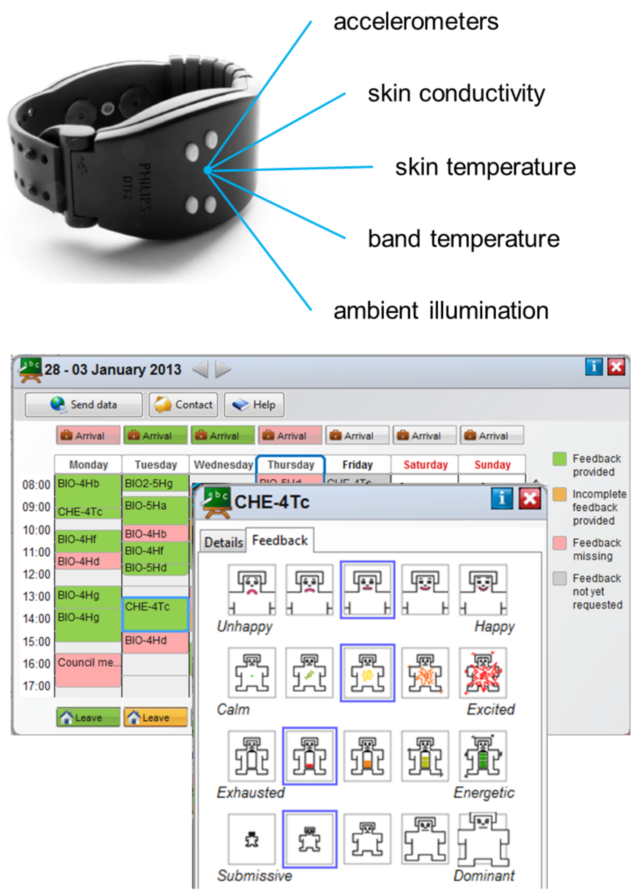

Figure 1. Our information sources: DTI-2 sensor wristband, calendar, questionnaire

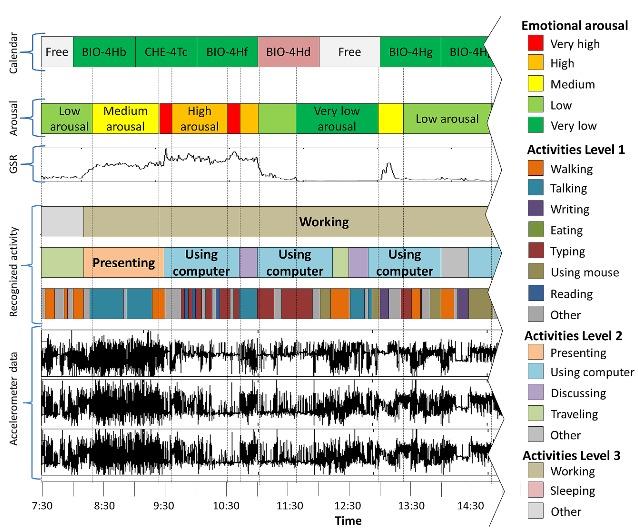

Figure 2. Raw sensor signal along with labels resulting from analysis.

gories). However, this approach triggered privacy concerns for a number of users, as calendars are shared within organizations. Therefore, we now provide a custom-made calendar application as part of our solution. Our calendar application can be used as a stand-alone calendar for the users who do not use other calendar applications, and it can be integrated with other calendars. Currently, we support automatic extraction of schedule information from *MS Outlook*. Similar plug-ins can be easily created for other applications (e.g. Google calendar, MacOSX iCal). The advantages of a separate calendar application lie in a better protection of users' privacy, seamless integration with the measurement device, and in possibilities for richer communication with the user through a customized GUI.

To collect subjective user feedback, we supply our application with an option for the user to indicate his/her state in relation to a calendar event on four dimensions – *Valence*, *Arousal*, *Energy* and *Dominance* – on a five-point scale, and give additional comments in a free form. We added the *energy dimension* to the three dimensions of the Self Assessment Manikin questionnaire [5] in order to allow for identification of personal *fatigue*-related patterns.

In order transform the raw data to the form which is meaningful to the user, we align the data along the time line and process it to obtain useful labels (Figure 2); the processed data form then the input for our visualization tool allowing us to generate views relating stress to different aspects of events occurred during the time period under observation (see Figure 4).

3 Data Analysis

Sensor measurements obtained from field trials suffer from artifacts caused by occasional movements of the device on the wrist, excessive motion, electrical artifacts and other sources [4]. Although most artifacts can be prevented in signals recorded in controlled experimental conditions, real-life signals need proper filtering. In order to avoid misinterpretation, we use a conservative strategy, rather disregarding the signals suffering from any artifacts than trying to retrieve interpretations that might be incorrect.

The skin conductance measurements can be decomposed into a phasic, fast changing component called *Electrodermal Response (EDR)* and a tonic, slow changing component called *Electrodermal Level (EDL)*. Due to the long-term setting of our studies and the battery life time, decreasing with the increase of sampling frequency of the device, the recordings were made with 2Hz sampling frequency, which is sufficient for estimating *EDL*, but may not be enough for separating the tonic (*EDL*) and phasic (*EDR*) components of the signal [4]. Therefore, we focused mainly on estimating the *EDL* level as an indicator of emotional arousal.

We apply the following procedure to produce estimations

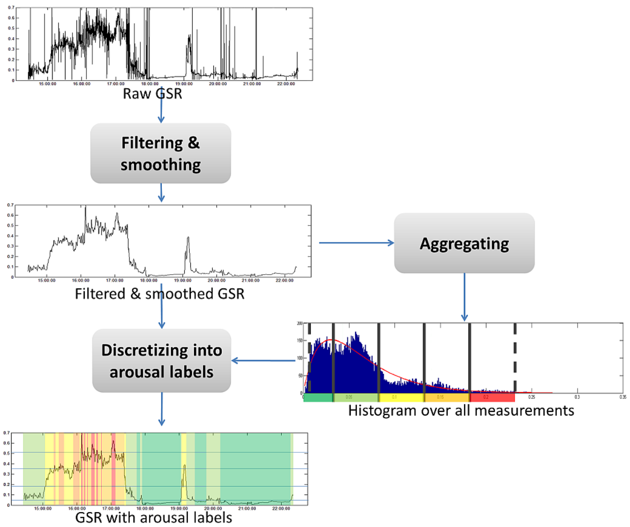

Figure 3. Translating the raw GSR signal into discrete arousal categories

of the stress level from sensor data (see Figure 3):

1. *Remove the first 15 minutes and the last 10 seconds of the series of skin conductance measurements*, since the sweat level in the skin gets adjusted to the new situation (the device put on) at the beginning of the measurement and there are, likely, movement artifacts just before the removal of the device from the wrist.

2. *Remove signal affected by losing contact with skin*: For each 5 seconds window we calculate the *ratio of lost/overall signal*. As lost signal, we consider registered values of skin conductance below $0.001\mu S$, which is below realistic values. In case this ratio is above 0.9, we remove the signal for the whole window, based on the results from [10].

3. *Identify and remove anomalies*: We use the shape characteristics of a genuine skin conductance response described in [4] to identify spikes in the signal that cannot correspond to genuine changes in skin conductance values. Based on experimental results, we set the maximal possible increase of the signal value to 20% per second and the maximum decrease to 10% per second and eliminate those (sequences of) samples that do not meet these criteria, going in the forward and the backward directions. Based on visual inspection of the signal, we established that this approach accurately removes positive and negative spikes that result from artifacts.

4. *Smooth the signal*: The slow changing component of the signal (*EDL*) can already be determined from time windows starting from 10–30 s [4]. We smooth the signal using a moving median filter for a window size of 1 minute, hence being a bit more conservative than the minimum recommended time interval to limit the influence of possible remaining artifacts. Since part of the signal within the window could have been removed as a result of filtering, we only take into consideration windows with at last 40% of the signal remaining.

5. *Define a slicing of skin stress estimation values into five arousal categories in order to ease the interpretation by the user*: To take into account differences in skin conductance between people, we define this slicing on the basis of the personal histogram of smoothed *EDL* values for the whole period of monitoring. We assume that, in the monitoring period, the person had at least one period of being calm for at least 5 minutes. We use the *min-max algorithm* for overlapping 5 minutes windows to find this most calm period and the maximal *EDL* value in it, denoted as ℓ_0. Based on the 300 bins histogram, we define the ℓ_5 value as the value before the first empty bin above ℓ_0. Taking into account that the fluctuations of skin conductance in the very calm states are low, we define δ as $(\ell_5 - \ell_0)/4.5$, and $\ell_1 = \ell_0 + 0.5\delta$, $\ell_2 = \ell_1 + \delta$, $\ell_3 = \ell_2 + \delta$ and $\ell_4 = \ell_3 + \delta$. We classify the values below ℓ_1 as "very low arousal", the values from $[\ell_1; \ell_2)$ as "low arousal", $\ldots$, and the values above ℓ_4 as very high arousal.

Note that the reported influence of the ambient temperature on *EDR* (Skin resistance level) is at a level of $3\%/^0C$ [6], and in case the fluctuations in the temperature are significant enough, a temperature correction should be applied before deriving the arousal levels.

4. Evaluation

To evaluate our approach, we organized a field trial with staff members of a university. 10 users were asked to wear the DTI-2 wristband at least during working hours for a period of 4 weeks. Calendar data was imported from their Outlook calendar to our calendar application, as described in Section 2. The users uploaded measurements from DTI-2 to our server remotely once a day. The participants were not given any view on their data before the end of the study. At the end of the study, we provided them with views on their data and performed a qualitative evaluation in form of a semi-structured interview. The main purpose was to determine if the generated views are 1) meaningful, 2) useful for the user, and 3) triggering actionable conclusions on improving stress balance.

We used LifelogExplorer [7] to generate views on the collected data. Figure 4 shows a view in which circles ("group" chosen in the menu) represent subjects of the

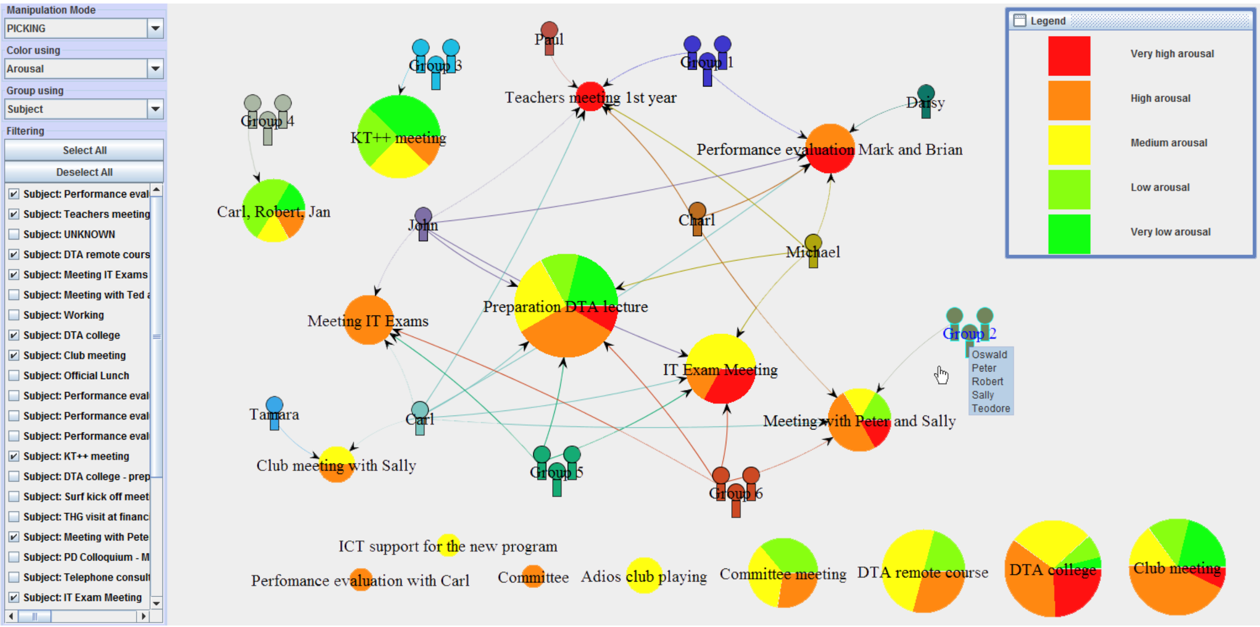

Figure 4. An example of a view with aggregation on subjects of calendar events and filtering used

meetings in the user's calendar, with the size of a circle reflecting the total amount of time spent in meetings on this subject. The pie chart inside the circle shows the time distribution of estimated arousal levels. The human-like figurines represent the people present at the specific activity (the view is based on real data, but, for privacy reasons, all subjects and names have been changed). Also other dimensions, like location, can be chosen instead of the subject dimension.

Note that information about users lives is inherently incomplete, e.g. information about meeting participants might be missing (resulting in isolated circles). Therefore, the user's knowledge remains indispensable for interpreting the views. Moreover, the views can only suggest correlations, but they cannot reveal causalities. It is up to the user to judge whether there can be a causal dependency between certain things and check it by adapting his behaviour/schedule/....

To determine whether the data indeed reflects the perception of a person, we presented a number of views (with different choices of dimensions for circles and figurines) and asked to comment on the value of these views.

Data is meaningful: In most cases the data has proven to reflect participants perception about their stress levels: *"It is immediately visible that in the week ... there is less stress, less activity than in other weeks."*
"The data that I see about the performance evaluations. The stress level in these job performance evaluations says much more about what the performance evaluation really was like. Much more than what is written on the paper. (...) It does not lie! (Laugh!) The report may lie, but the stress level does not lie."

In order to quantitatively evaluate the extent to which the presented information confirms user perception, we picked, at random, some of the individual aggregations presented in each view. We asked the user to relate to these aggregations one-by-one. Due to extent of the data and time constraints during the interview, we referred only to $\sim 58\%$ of all the aggregations present in the views. Out of these $\sim 91\%$ were correctly reflecting the perception of the user, i.e. "green" was described as calm or relaxing, and "red" as stressful or engaging.

Provides new information: In many cases, the data not only confirms the perception of our participants, but also provides them with new information: *"For instance, about the ... meetings. I really learned that the experience I have there is indeed reflected in the stress levels. It is an eye opener for me."*
"I learned that teaching through video conferencing is really different from teaching in class."

Triggers actionable advices: Further, the self-reflection in many cases triggers people to rethink their behavior: *"I learned which things on ... are stressful and I perhaps should relax and do something about that. That is very good to see." "I'd love to have more data, do this over a number of weeks and see whether this pattern re-emerges."*

5. Conclusion

The increase in battery life, storage capacity and miniaturization of wearable sensors has made real time recording of various bodily responses feasible. Many solutions are already on the market, ranging from simple accelerometer-based ones, like *Jawbone UP*[2] and *DirectLife*[3] offering activity intensity recognition, to solutions that measure skin conductance, heart rate and body temperature, like *Affectiva Q sensor*[4], *Basis*[5], or *BodyMedia armband*[6]. Solutions such as *Noldus Face Reader*[7] record individual's facial expressions in order to perform automatic emotion recognition. At the same time, pure software solutions can capture user activities related to computer use, like *Spector Pro*[8] and *PC Pandora*[9], recording the use of keyboard, mouse and different software tools. In the coming years, we expect these technologies to make a step from the "quantified-self" segment of the market to a general consumer market, with sensors being embedded into consumer products like watches, shoes, etc. Therefore, it is the right time to develop methods for transforming data produced by these sensors to information the user can action upon.

In our approach, we monitor stress in relation to the events of the everyday life. The results of such long-term monitoring can be used by the person directly or can serve as input for consultations with a psychologist. Confronting people with quantitative information about their life can trigger self-coaching. Moreover, the same monitoring technologies can be used to measure the short-term and long-term effects of different stress coping strategies on the person in question and thus help with choosing appropriate ones.

Our field study shows that even using a limited number of sensors we obtain data to provide users with information about their behavioral patterns that they perceive as meaningful and useful, and trigger their ideas about behavioral changes necessary to achieve a better stress balance.

Acknowledgements

We thank EIT ICT Labs for the financial support of this work, the occupational health organization HumanCapitalCare for facilitating the organization of the studies, and the staff members of TU/e for participating in our studies.

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 *Proceedings of the 2nd ACM SIGHIT International health informatics symposium*, pages 673–678. ACM, 2012.

[2] J. Bakker, M. Pechenizkiy, and N. Sidorova. What's your current stress level? detection of stress patterns from GSR sensor data. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on Data Mining*, pages 573–580. IEEE, 2011.

[3] H. Bořil, P. Boyraz, and J. Hansen. Towards multimodal drivers stress detection. *Digital signal processing for in-vehicle systems and safety*, pages 3–19, 2012.

[4] W. Boucsein. *Electrodermal activity*. Springer, 2011.

[5] M. Bradley and P. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[6] W. W. Grings. Recording of electrodermal phenomena. *Methods in physilogical psychology (Bioelectric recording techniques, Part C: Receptor and effector processes*, 1(1):273–296, 1974.

[7] R. Kocielnik, F. M. Maggi, and N. Sidorova. Enabling self-reflection with lifelogexplorer: Generating simple views from complex data. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth 2013, Venice, Italy, May 5-8, 2013*. IEEE, 2013.

[8] R. Lazarus and S. Folkman. *Stress, appraisal, and coping*. Springer Publishing Company, 1984.

[9] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski. AffectAura: an intelligent system for emotional memory. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 849–858. ACM, 2012.

[10] J. O'Gorman and C. Horneman. Consistency of individual differences in non-specific electrodermal activity. *Biological Psychology*, 9(1):13–21, 1979.

[11] 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.

[12] G. Rigas, Y. Goletsis, and D. Fotiadis. Real-time driver's stress event detection. *Intelligent Transportation Systems, IEEE Transactions on*, 13(1):221–234, 2012.

[13] P. Sanches, K. Höök, E. 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 *Proceedings of the 8th ACM conference on designing interactive systems*, pages 47–56. ACM, 2010.

[14] N. Sharma and T. Gedeon. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine*, 108(3):1287–1301, 2012.

[15] J. Westerink, M. Ouwerkerk, G. de Vries, S. de Waele, J. van den Eerenbeemd, and M. van Boven. Emotion measurement platform for daily life situations. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

[2] jawbone.com/up
[3] www.directlife.philips.com/
[4] www.affectiva.com/q-sensor/
[5] https://mybasis.com/
[6] www.bodymedia.com/
[7] http://www.noldus.com/facereader
[8] http://www.spectorsoft.com
[9] www.pcpandora.com/