Discriminating Stress From Cognitive Load Using a Wearable EDA Device

Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, *Member, IEEE*, and Ulrike Ehlert

Abstract—The inferred cost of work-related stress call for prevention strategies that aim at detecting early warning signs at the workplace. This paper goes one step towards the goal of developing a personal health system for detecting stress. We analyze the discriminative power of electrodermal activity (EDA) in distinguishing stress from cognitive load in an office environment. A collective of 33 subjects underwent a laboratory intervention that included mild cognitive load and two stress factors, which are relevant at the workplace: mental stress induced by solving arithmetic problems under time pressure and psychosocial stress induced by social-evaluative threat. During the experiments, a wearable device was used to monitor the EDA as a measure of the individual stress reaction. Analysis of the data showed that the distributions of the EDA peak height and the instantaneous peak rate carry information about the stress level of a person. Six classifiers were investigated regarding their ability to discriminate cognitive load from stress. A maximum accuracy of 82.8% was achieved for discriminating stress from cognitive load. This would allow keeping track of stressful phases during a working day by using a wearable EDA device.

Index Terms—Cognitive load, electrodermal activity (EDA), personal health systems (PHSs), stress recognition, wearable.

I. Introduction and Motivation

THE PROBLEM of work-related stress is of growing interest, especially in western countries. In 2007, the European Foundation for the Improvement of Living and Working Conditions identified stress to be the second most common work-related health problem across the European Union (EU) [1]. Work-related stress occurs when there is a poor match between job demands and the capabilities, resources or needs of the worker [2]. It is associated with various illnesses, such as cardiovascular diseases and musculoskeletal disorders (e.g., back pain) [1]. These diseases lead to absence of work resulting in high economic costs [3]. An early assessment of stress risk factors or an early detection of negative vital signs could significantly reduce this cost by early prevention. In this, we see a new opportunity for personal health systems (PHSs). If a PHS could advert us of stressful situations during office work, it could help us to create a better "work–life balance." Furthermore, such a PHS could also be used in therapies and for medical decision support.

This paper goes one step toward this goal. Since we are aiming at an application in the office environment, a suitable experiment that is close to a real-life office situation had to be found. Existing studies often use mental workload as stress-eliciting factor. This is certainly an important stress component in the office. However, there are other contributing factors such as social threat by superiors and colleagues. We have therefore chosen a stress test that includes both factors. The Montreal imaging stress task (MIST) [4] is a standardized computer-based task consisting of a stress and a control condition: The stress condition combines mental arithmetic problems under time pressure with social-evaluative threat, whereas the control condition consists of mental arithmetic with neither time pressure nor social evaluation, which is similar to relaxed working on a computer. Because an office worker is always confronted with a certain amount of cognitive load, we try to discriminate a state of mental and psychosocial stress from a state of (mild) cognitive load rather than distinguishing between a rest (doing nothing) and a stress condition.

For an "everyday life application," a minimal sensor setup is desired for comfort reasons. This paper will therefore focus on a single-sensor modality, even though multimodal signals have been recorded during our experiments. The analysis of the electrodermal activity (EDA) was chosen, because it represents an adequate measure for sympathetic activation that is related to stress. The Emotion Board—an unobtrusive, wearable device [5]—was used for recording the EDA of 33 healthy male subjects. Feature selection and classification were performed to identify EDA features that are suitable for distinguishing stress from cognitive load. This represents a necessary first step toward a future online PHS.

The following section explains the EDA and its relation to stress as well as related work in this field. Our experiment is described in Section III. Section IV explains the evaluation methods, whereas the results are presented and discussed in Sections V and VI.
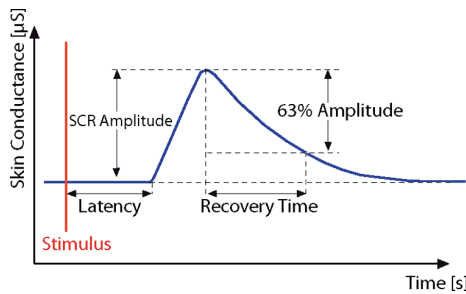
Fig. 1.   Ideal SCR with typically computed features.

II. PHYSIOLOGICAL BACKGROUND AND RELATED WORK

A. Stress, the Sympathetic Nervous System, and EDA

In the following, we will explain why the EDA is a good measure to indicate stress. If a human organism is in danger or injured, various physiological changes occur, which are denoted by the term "stress reaction" [6]. The stress reaction was first described by Selye [7]. In the short term, it helps the body to adapt to the stressor, e.g., by providing additional energy and thereby suppressing the immune system. However, if the organism has no time to recover and the stress reaction proceeds over longer time, this can have adverse effects, including the development of gastric ulcers and increased sensitivity to infections.

Two components are involved in the stress reaction [8], the hypothalamic-pituitary-adrenal (HPA) axis and the sympathetic nervous system, which forms part of the autonomic nervous system (together with the parasympathetic nervous system). The sympathetic and the parasympathetic nerves work at the same time, but in an antagonistic way. The organs are thus connected to sympathetic as well as to parasympathetic nerves. The skin, however, represents an exception: The sweat glands as well as the skin blood vessels are exclusively innervated by the sympathetic nervous system [9]. This makes the skin conductance—which is related to the number of active sweat glands—an ideal measure for sympathetic activation and therefore for the stress reaction. In contrast, other physiological measures—e.g., the heart rate—are influenced by both the sympathetic *and* the parasympathetic nervous systems.

The EDA is recorded by measuring the conductivity of the skin because the relationship between the skin conductance is proportional to the sweat secretion [10]. The EDA is usually measured at the palmar sites of the hands or the feet where the density of sweat glands is the highest ($>2000/cm^2$). The slowly changing part of the EDA signal is called the skin conductance level (SCL) and is a measure of psychophysiological activation [8]. It can vary substantially between individuals. A fast change in the EDA signal (a "peak") occurs in reaction to a single stimulus (e.g., a startle event) and is called (specific) skin conductance response (SCR). It appears between 1.5 and 6.5 s after the stimulus. Features used to describe the characteristics of an SCR include the amplitude of the SCR, the latency (between stimulus and SCR onset), and the recovery time. They are shown in Fig. 1. In contrast to the specific SCRs, the non-specific fluctuations (NS.SCRs) occur "spontaneously" without any external stimulus. The frequency and the mean amplitude of NS.SCRs are considered as measures for psychophysiological activation [9]. Like the SCL, the NS.SCRs show interindividual variation. NS.SCRs exhibit the same or a very similar shape as specific SCRs.

B. Studies in Psychophysiology

Boucsein gives an extensive summary of early EDA research in relation to stress [9]. He shows that the SCL and the NS.SCRs are sensitive and valid indicators for the course of a stress reaction, whereas other physiological measures (e.g., the heart rate) do not show equal sensitivity. Most cited studies use electrical stimuli or movies to induce a reaction of the EDA. The Lazarus Group showed that the SCL and the heart rate increased significantly during the presentation of cruel or disgusting films [11], [12]. Nomikos *et al.* [13] showed that even the expectation of an aversive event (an electrical stimulus in this case) could elicit a similar reaction in SCL as the event itself. Several studies investigating the effect of the anticipation of electrical stimulation suggest that the rising of the SCL rather reflects an increased cognitive activity related to the avoidance of aversive events than an emotional component. On the other hand, the frequency of the NS.SCRs reveals the emotional component of the stress reaction.

Further studies used experimental settings that were closer to a real-life office environment than simple electrical stimuli. Several authors [14], [15] investigated how involuntary interruptions in the work flow due to long system response times influenced the EDA. An increase of NS.SCRs for long system response times could be demonstrated. Jacobs *et al.* [16] also showed an increase in skin conductivity during mental stress.

C. Studies in Engineering Sciences

In the engineering field, we only found one study that used a single EDA-sensor setup [17]. The galvactivator is a glove-like wearable device that maps the wearer's skin conductivity to the brightness of an LED. Using a video camera, the authors investigated the aggregate brightness level emitted from 1200 such devices worn by audience members of a symposium. They found the aggregate brightness level to be correlated with stage events.

In the following, we describe studies that combine several physiological signals for detecting stress. The electrocardiogram (ECG), the electromyogram (EMG) of the trapezius (shoulder), the EDA, and the respiration were recorded during a real-life driving task in [18]. The three levels of driving stress (rest, highway, and city driving) could be classified with an accuracy of over 97% across multiple drivers and driving days.

Zhai and Barreto [19] used an interactive 'Paced Stroop test," where subjects had to select the font color of a word shown on the screen. The word itself named a color. The authors then classified the data segments with matching meaning and font color (nonstress) against the segments with mismatching meaning and font color (stress). With features from EDA, blood volume pulse, pupil diameter, and skin temperature, recorded from 32 subjects, a support vector machine (SVM) reached an

accuracy of 90.1% with 20-fold cross validation. Without the pupil diameter feature, the accuracy dropped to 61.45%.

A "Stroop color-naming task" in conjunction with a 70-dB noise stimulus was used for stress induction in [20]. Heart rate, skin temperature, and EDA were measured from 80 subjects. A classification accuracy of 96.67% is reported with a twofold cross validation. Unfortunately, not all procedures are described in detail.

Facial EMG, respiration, EDA, and ECG were used in [21] for detecting five emotional states, i.e., high stress, low stress, disappointment, euphoria, and neutral. Data were recorded from simulated car races. Data from a first race were used for training and from a second race for testing. An accuracy of 86% is reported for a single driver with an SVM classifier using radial basis functions (rbfs).

A different approach was chosen in [22]: instead of classifying discrete states, the authors tried to estimate a continuous variable (the stress level). Mental arithmetic and an alphabetic task were used to induce a stress response. Facial, physiological (heart rate, skin temperature, EDA), behavioral, and performance data (e.g., error rate) were used in a dynamic Bayesian network in order to estimate the stress level of five subjects. The workload was taken as ground-truth stress level. The correlation between the ground truth and the estimated stress level was high (between 0.79 and 0.92 for five subjects). How much the performance rate attributed to the high correlation is not reported.

The aforementioned studies mostly use a form of "mental stress." Some of them [18], [21] also include an emotional component. Aiming at an experiment that is close to a real-life office situation, we used a *combination* of mental and psychosocial stress factors for our experiments. We have not found a comparable study. Furthermore, we tried to discriminate stress from cognitive load rather than distinguishing between rest (doing nothing) and stress.

III. EXPERIMENT

The MIST, [4] was created in order to evaluate effects of psychosocial stress on physiology and brain activation by functional MRI (fMRI). It has shown to induce a moderate stress response when referring to the HPA axis. For our purposes, the test was slightly modified for a use outside of an fMRI environment by psychologist professionals in agreement with the inventors of the MIST. The procedure was authorized by the ethics commission. For the study, 33 male, healthy subjects (mean age: 24.06; mean body mass index (BMI): 23.63) were recruited. They were paid 80 Swiss francs for participating in two sessions of 2 h. The second session took place two weeks after the first. As a cover story, the subjects were told that they were taking part in an experiment investigating the relationship between cognitive performance and physiological characteristics. In reality, they were confronted with mental stress and social evaluative threat during one session (the *stress condition*), and with mild cognitive load during another session (the *cognitive load condition*). In order to eliminate possible habituation effects, half of the subjects were exposed to the stress condition during the first session and to the cognitive load condition during the second
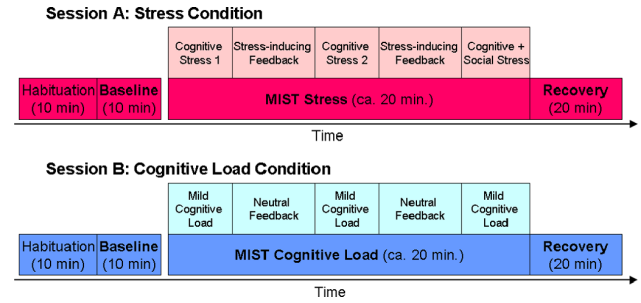

Fig. 2. Experiment procedure for stress and cognitive load condition. All subjects participated in both conditions during two sessions in random order with a time distance of two weeks.
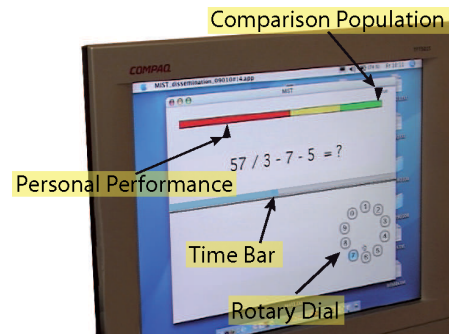

Fig. 3. MIST screen during the stress condition.

session, whereas for the other half, the sequence was *vice versa*. Both conditions consisted of three experiment phases denoted by "baseline," "MIST," and "recovery" (see Fig. 2); the two conditions were the same except for the MIST phase as explained in the following.

During the *stress condition*, the subjects performed mental arithmetic tasks on a computer. The program adapted the difficulty and the time limit, such that the subjects could only solve 45%–50% of the arithmetic problems correctly. This corresponds to a stressful office situation where the worker is presented with work requirements that do not match his capabilities.

Fig. 3 shows the screen display during the stress condition. In addition to the arithmetic task and the rotary dial for response submission, a time bar showed the remaining time for solving the task, and "time out" was displayed when the time had elapsed. When subjects gave an answer before the end of the given time, the feedback "right" or "wrong" was displayed. A color bar showed a comparison between the individual performance and the performance of a simulated, representative comparison population. The subjects were told that the test would not work if their performance did not reach the green range of the bar and the study leader would monitor the whole experiment remotely. This represented a social evaluative threat. Additionally, after each block of mental arithmetic, the subjects received feedback from the experiment leaders regarding their performance. This represents a similar situation as in real life when the boss complains about one's working productivity. During the first feedback phase, only mild social stress was induced: the experiment leader told the subject that his performance was not as good as expected, and asked questions regarding

methodological aspects of the task and the computer (e.g., "Is there a problem with the response submission on the keyboard?"), thus attributing the "problems" to the task itself and not to the capabilities of the subject. During the second feedback phase, strong social stress was induced: The experiment leader called his supervisor, the study leader, who immediately entered the laboratory and asked questions referring to personal problematic characteristics. These included: "Did you sleep badly?," "Did you ever have math problems at school?," and "Did you drink alcohol or take drugs?." The study leader also pointed out the high cost involved in such studies and requested the subjects to make more effort. He stayed in the room behind the subject and observed him during the next mental arithmetic block. By comments like "Are you stressed?," he tried to increase the stress level even more. After two-third of the time, the study leader acted resigned, told the experiment leader to continue the mental arithmetic, and left the room. The following list shows the experimental schedule (see also Fig. 2).

1) Instructions and signing of consent form.
2) Habituation and baseline period (20 min): Questionnaires and reading magazines.
3) Cognitive stress I (4 min): Mental arithmetic under time pressure and with performance evaluation (as described before).
4) First feedback phase inducing mild social stress: Methodological questions.
5) Cognitive stress II (4 min): Like cognitive stress I, but with additional pressure not to fail again.
6) Second feedback phase inducing strong social stress: Personal questions.
7) Combination of strong social and cognitive stress (4 min): Like cognitive stress II, but with study leader observing the subject.
8) Recovery period (1 h): Questionnaires and reading magazines.

For the *cognitive load condition*, the schedule was analogous to the one described before. However, during the mental arithmetic phases, there was no time limitation for solving the tasks and no social evaluation. The color and time bars were thus not displayed. This corresponds to mild cognitive load that occurs when working on a computer during a quiet office day. During the feedback phases, the experiment leader asked friendly, neutral questions, e.g., "How did you come to participate in this study?" and "What are you studying?." Fig. 2 illustrates the schedules for the stress and the cognitive load conditions.

A. Measured Signals

Beside the EDA, other signals such as ECG, breathing, and movement have been recorded. As explained in Section II-A, in this paper, we focused on the EDA, because it indicates sympathetic activation and can be measured unobtrusively. The Emotion Board was used to measure EDA, see Fig. 4. It employs an exosomatic quasi-constant voltage method for measuring the skin conductance. Hereby, a constant voltage (500 mV) is applied to a voltage divider composed of the skin resistance and a reference resistor (100 kΩ). Measuring the voltage at
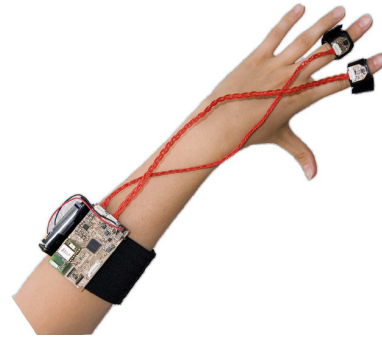

Fig. 4. Emotion Board: unobtrusive, wearable device for measuring EDA. Size: 41 mm × 67 mm. Number of electronic parts: 74. Power consumption: 182 mW.

the reference resistor with an operation amplifier allows us to determine the skin conductance. To eliminate high-frequency noise, a second-order low-pass filter with a cutoff frequency of $f_c = 5$ Hz is applied before A/D conversion of the measured signal (referred to as "level" in the following). Applying an additional high-pass filter (second-order, $f_c = 0.05$ Hz) yields the phasic part of the EDA signal. For further noise reduction, this signal is once more low-pass filtered (second-order, $f_c = 5$ Hz), amplified, and fed to the A/D converter. A Bluetooth wireless link is used to transfer the EDA data at 16 Hz. The finger straps of the Emotion Board incorporate reusable dry Ag/AgCl electrodes from Brainclinics [23], which were attached to the middle phalanges of the left index and middle fingers, see Fig. 4. In case of dry skin, conductive gel was applied, whereas for wet skin, a version of the Emotion Board with reduced gain was used. See [5] for more information about the Emotion Board.

IV. EVALUATION METHODS

This section describes the employed evaluation methods. They represent a first step toward a future online PHS, which is capable of recognizing stress. After preprocessing the data and detecting the peaks in the high-pass-filtered EDA signal, we calculated descriptive statistics to identify 16 potentially meaningful features for the three experiment phases, "baseline," "MIST," and "recovery," of the two experiment conditions (see Fig. 2). In a second step, the features were automatically grouped into the conditions "stress" and "cognitive load" using different classifiers with a leave-one-person-out cross validation. For each classifier, the best feature combinations were determined.

A. Preprocessing and Peak Detection

After correcting few errors caused by Bluetooth communication, the signals were smoothed to reduce the noise. An example is shown in Fig. 5. Afterward, the peaks in the high-pass-filtered signal were detected by applying an adaptive threshold, as shown in Fig. 6. Beside the height of the peaks, the time interval between consecutive peaks was also calculated. In the following, we will refer to the reciprocal of this value by the term "instantaneous peak rate." From these measures, various features were calculated.
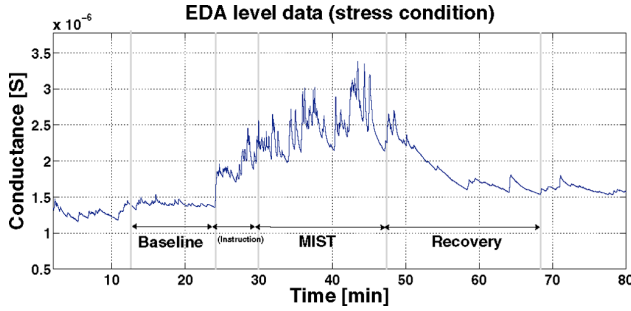
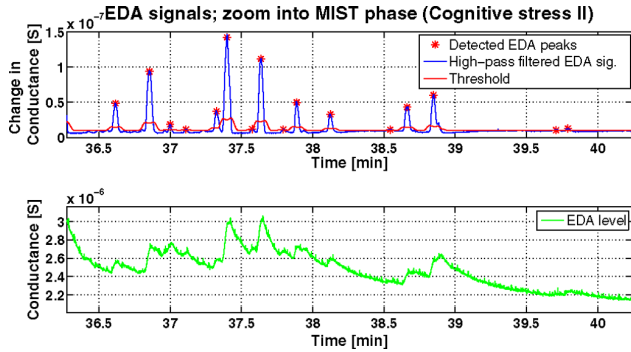Fig. 5. Corrected, smoothed EDA-level signal recorded during a stress condition.

Fig. 6. Zoom into MIST phase "Cognitive stress II" shown in Fig. 5: (bottom) EDA-level signal and (top) EDA high-pass-filtered signal. The red stars indicate the automatically detected peaks.
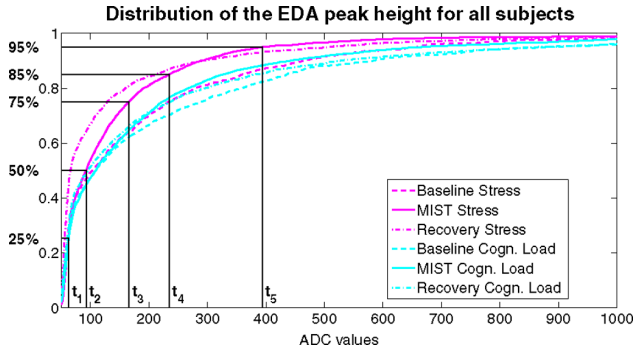

Fig. 7. Distribution of the peak height for the three experiment phases (baseline, MIST, and recovery) during stress and cognitive load for all subjects. The labels t_1-t_5 indicate the quantile thresholds for 25%, 50%, 75%, 85%, and 95% respectively, for the MIST phase of the stress condition.

B. Descriptive Statistics for Finding Meaningful Features

As described in Section II-A, the standard EDA features are only defined for specific SCRs. Since the occurrence of SCR-eliciting stimuli, such as startle events, are not known in our experiment, specific SCRs and NS.SCRs cannot be distinguished, and the standard EDA features are thus not directly applicable. For this reason and because statistical EDA analyses during close-to-life stress scenarios are scarce, the EDA data were first explored in order to find potentially useful features to distinguish stress from cognitive load.

The distribution of the peak height during the different experimental phases is shown in Fig. 7 for all subjects. The curves

show which percentage of the peaks is smaller than a certain threshold. The curves for the MIST and the recovery phase in the stress recording (red continuous and dash dotted lines) are "above" all the other curves. This means that there were more small peaks occurring during the MIST and the recovery period of the stress condition than during all the other periods. The distributions of the peak height during the different experimental phases thus carry information to distinguish these phases.

Quantile thresholds are a suitable type of features to describe the observed differences in the distributions. The quantile thresholds for the peak height have been calculated for each experiment phase, condition, and subject individually in such a way that 25%, 50%, 75%, 85%, and 95% of the peaks were smaller than the respective thresholds. An example is shown in Fig. 7. The five thresholds for "MIST stress" are indicated by t_1-t_5. In order to reflect the part of the distributions, where the experimental phases show the largest differences, i.e., in the upper left corner, a higher resolution was employed between 75% and 100%. Accordingly, the same thresholds were calculated for the instantaneous peak rate too.

C. Feature Calculation

As already mentioned, the recording of each subject and each experiment condition (cognitive load and stress) was divided into the experiment phases "baseline," "MIST," and "recovery" (see Fig. 2). The following 16 features were calculated for each of the three experiment phases, as well as for the concatenation of "MIST" and "recovery":

1) Mean, maximum and minimum EDA level.
2) Slope of the EDA level (calculated by a linear regression).
3) Mean EDA peak height.
4) Mean EDA peak rate in peaks/min. (i.e., number of peaks/duration of experiment phase).
5) Quantile thresholds at 25%, 50%, 75%, 85%, and 95% for the EDA peak height and the instantaneous peak rate.

We denote the aforementioned features as "nonrelative features."

1) Baseline-Related Features: In order to investigate to what extent a compensation of interindividual or daily differences increases the system performance, we also calculated the features for "MIST," "recovery," and "MIST+recovery" in relation to the baseline period. For all the level features, except for the slope, we subtracted the mean value of the level during the baseline from the features calculated during the other time periods. For all the remaining features, we divided the features by the corresponding baseline features. In the following, we denote these features as "relative features."

D. Feature Selection and Classification

For classification, we chose the accuracy of a "leave-one-person-out cross validation" as a performance index. This means that the classifier is trained with the data of all subjects, except one, and tested on the data of the subject that was left out for training. This procedure is repeated for all subjects and the accuracy is calculated by dividing the number of correctly classified test data points by the total number of test data points. The
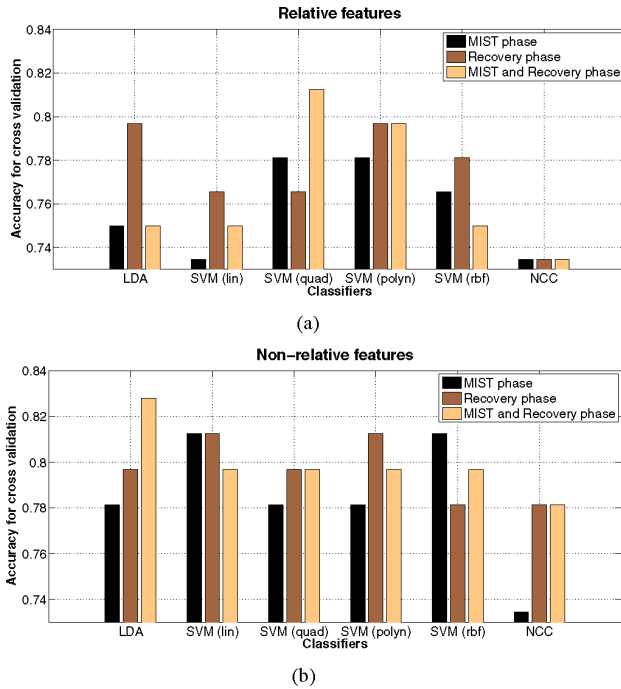
Fig. 8.   (a) Cross validation results with best *relative* features. The maximum accuracy of 81.3% is achieved by the SVM classifier with quadratic kernel when using the concatenation of the "MIST" and the "recovery" phases. (b) Cross validation results with best *nonrelative* features. The maximum accuracy of 82.8% is achieved by the LDA when using the concatenation of the "MIST" and the "recovery" phases.

method is similar to "leave-one-out cross validation" [24]. Since some of the features are expected to be correlated, we employed a feature selection using a wrapper approach. With 16 features, a complete search for the best combination of features is still computationally acceptable. The cross validation was thus performed for all possible feature combinations, and the combination that yielded the best accuracy was chosen for each classifier. The following classification methods have been applied:

1) linear discriminant analysis (LDA);
2) SVM with linear, quadratic, polynomial, and rbf kernels;
3) nearest class center (NCC) algorithm.

All evaluations were performed in MATLAB.

V. RESULTS

In this section, we present the classification results for the six classifiers. We also compare the classification performance for distinguishing between the "stress" and the "cognitive load" condition when taking features from different experiment phases as classifier input. These phases are "MIST," "recovery," and the concatenation of the "MIST" and the "recovery" phases. The results are reported for both relative and nonrelative features. Data from 32 subjects were evaluated, since one subject had to be excluded from the analysis due to strong artifacts in the EDA signals. The artifacts were caused by a poor fixation of the Emotion Board hardware to the arm, which resulted in a loose connection when the subject was moving.

Fig. 8(a) and (b) shows the results for the relative and the nonrelative features, respectively. The maximally achieved

accuracy, i.e., the accuracy of the best feature combination determined by leave-one-person-out cross validation, is presented in the plots. It can be observed that all classification methods perform in a similar range. The maximum accuracy of 82.8% is achieved by the LDA when using nonrelative features calculated from the concatenation of the "MIST" and the "recovery" phases as classifier input. The NCC algorithm shows the poorest performance, especially when using relative features. The more complex SVM classifiers perform better or at least equally as the linear SVM classifier when using relative features. Over all, the classifiers perform poorer with relative features than when using nonrelative features.

Table I shows the feature combinations for the classifiers that achieved the maximum accuracy for the phases "MIST," "recovery," and "MIST+recovery" when using relative as well as nonrelative features. The selected features are marked by a cross. The number in the last column indicates how many times a certain feature was chosen. The feature selected most often is the 50% quantile threshold for the peak height. The 85% quantile threshold for the instantaneous peak rate represents the second most chosen feature. The third rank is achieved by three features where two of them are quantile features. The mean EDA peak rate as well as the 50% quantile threshold of the peak rate were never chosen.

The feature combination that reached the overall performance maximum (with LDA and "MIST+recovery") includes the mean EDA peak height, the 50%, 75%, and 85% quantile thresholds for the peak height, the 85% and 95% quantile thresholds for the instantaneous peak rate, and the EDA slope.

VI. DISCUSSION, CONCLUSION, AND OUTLOOK

A. Discussion of the Results

As presented in the previous section, all investigated classifier methods performed in a similar range. The maximum accuracy of 82.8% was achieved by the LDA when using the concatenation of the "MIST" and the "recovery" phases for feature calculation. Generally, taking only the "MIST" yields slightly smaller accuracies than taking the "recovery" phase or the concatenation of both. The "recovery" phase thus seems to contain useful information for distinguishing stress from mild cognitive load.

Furthermore, the classifiers perform better with nonrelative features than when using relative features. This means that the calculation of the baseline features is not necessary, and therefore, no calibration procedure would be needed for a practical system. However, since we did not expect the nonrelative features to perform better, it has to be tested whether this observation may not be due to the complete search feature selection that selects the features optimally for the 32 subjects (even though a cross validation scheme was used).

When using nonrelative features, linear and nonlinear classification methods performed similarly, which leads to the assumption that the problem is linearly separable. On the other hand, when using relative features, the more complex SVM classifiers performed better or at least equally as the linear SVM classifier. Transforming the features thus might decrease linear

TABLE I

RESULT OF THE COMPLETE SEARCH: FEATURE COMBINATIONS FOR THE CLASSIFIERS THAT ACHIEVED THE MAXIMUM ACCURACY FOR THE PHASES "MIST," "RECOVERY," AND "MIST+RECOVERY" WHEN USING RELATIVE AS WELL AS NONRELATIVE FEATURES

| Experiment Phase | MIST | | MIST | | Recovery | | Recovery | | MIST + Recovery | | Number |
| Type of Features | relative | | nonrel. | | relative | | nonrel. | | relative | nonrel. | of |
Classifiers	SVM quadratic	SVM polynom.	SVM linear	SVM rbf	LDA	SVM polynom.	SVM linear	SVM polynom.	SVM quadratic	LDA	times selected
Mean EDA level	x					x	x		x		4
Mean EDA p. rate											0
Mean EDA p. height		x	x		x	x			x	x	6
25% th. p. height					x	x					2
50% th. p. height	x	x	x	x	x	x	x		x	x	9
75% th. p. height					x					x	2
85% th. p. height		x	x	x	x	x				x	6
95% th. p. height	x										1
25% th. inst. p. rate	x	x						x			3
50% th. inst. p. rate											0
75% th. inst. p. rate				x				x			2
85% th. inst. p. rate	x	x	x	x			x	x		x	7
95% th. inst. p. rate	x		x	x				x	x	x	6
Max EDA level						x	x				2
Min EDA level							x				1
Slope of EDA level		x			x		x	x		x	5

The selected features are marked by a cross. If multiple combinations achieved the same result, the combination with the smallest number of features was chosen. th. stands for threshold, p. for peak, and inst. for instantaneous.

separability. For a practical system, where computational complexity is critical, we therefore suggest the use of nonrelative features combined with a linear classification method.

Four out of five features selected most often by the classifiers that achieved the highest accuracies are quantile threshold features. Interestingly, the mean EDA peak rate, which would probably be one of the first features one would extract, was never selected. Instead, the quantile features of the instantaneous EDA peak rate (at 85% and 95%) form part of the five best features. This confirms our hypothesis that the quantile thresholds are suitable features to distinguish stress from cognitive load. Moreover, these features are independent of the number of peaks occurring during the respective time period, which is an advantage because individuals differ substantially in their average number of spontaneous peaks.

B. Conclusion

We can conclude that, in our experiment, the monitoring of EDA allows a discrimination between cognitive load and stress with an accuracy larger than 80% with leave-one-person-out cross validation. We have found that the EDA peak height and the instantaneous peak rate carry information about the stress level of a person. Quantile thresholds are suitable features to describe these distributions. For the classification, not only the stress phase itself, but also the recovery phase should be taken into account.

C. Outlook

So far, we have achieved an accuracy larger than 80% in discriminating stress from cognitive load using EDA in an experimental setting. Since the MIST experiment represents a close-to-real-life stress scenario, we expect similar results in future long-term office experiments. A PHS that discriminates stress from cognitive load during office work could add up the stressful time periods during the day and give respective feedback to the user. This would help the user to employ breaks and holidays more efficiently. For the Emotion Board to serve as a standalone PHS for stress prevention, further development is needed. This includes online signal processing as well as appropriate feedback to the user.

REFERENCES

[1] European Foundation for the Improvement of Living and Working Conditions. Work-related stress. (2007). [Online]. Available: http://www.eurofound.europa.eu/ewco/reports/TN0502TR01/TN0502TR01.pdf
[2] U.S. National Institute for Occupational Safety and Health. Stress at work. (1999). [Online]. Available: http://www.cdc.gov/niosh/docs/99-101/
[3] State Secretariat for Economic Affairs. Die Kosten des Stresses in der Schweiz. (2003). [Online]. Available: http//www.seco.admin.ch/dokumentation/publikation/00008/00022/01511/
[4] K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, and J. C. Pruessner, "The Montreal imaging stress task: Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain," *J. Psychiatry Neurosci.*, vol. 30, no. 5, pp. 319–325, 2005.
[5] J. Schumm, M. Bächlin, C. Setz, B. Arnrich, D. Roggen, and G. Tröster, "Effect of movements on the electrodermal response after a startle event," *Methods Inf. Med.*, vol. 47, no. 3, pp. 186–191, 2008.
[6] J. P. J. Pinel, *Biopsychologie*, 2nd ed. Heidelberg, Germany: Spektrum, 2001.
[7] H. Selye, "A syndrome produced by diverse nocuous agents," *Nature*, vol. 138, p. 32, 1936.
[8] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, Eds., *Handbook of Psychophysiology*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[9] W. Boucsein, *Electrodermal Activity*. New York: Plenum, 1992.
[10] C. Darrow, "The rationale for treating the change in galvanic skin response as a change in conductance," *Psychophysiology*, vol. 1, pp. 31–38, 1964.
[11] R. S. Lazarus, *Psychological Stress and the Coping Process*. New York: McGraw-Hill, 1966.

[12] R. S. Lazarus and E. M. Opton, "The study of psychological stress: A summary of theoretical formulations empirical findings," in *Anxiety and behavior*, C. D. Spielberger, Ed. New York: Academic, 1966.

[13] M. S. Nomikos, E. Opton, J. R. Averill, and R. S. Lazarus, "Surprise versus suspense in the production of stress reaction," *J. Pers. Soc. Psychol.*, vol. 8, pp. 204–208, 1968.

[14] W. Boucsein, "Psychophysiological investigation of stress induced by temporal factors in human-computer interaction," in *Psychological Issues of Human-Computer Interaction in the Work Place*, M. Frese, E. Ulich, and W. Dzida, Eds. Amsterdam, The Netherlands: North Holland, 1987, pp. 163–181.

[15] W. Kuhmann, W. Boucsein, F. Schaefer, and J. Alexander, "Experimental investigation of psychophysiological stress-reactions induced by different system response times in human-computer interaction," *Ergonomics*, vol. 30, pp. 933–943, 1987.

[16] S. C. Jacobs, R. Friedman, J. D. Parker, G. H. Tofler, A. H. Jimenez, J. E. Muller, H. Benson, and P. H. Stone, "Use of skin conductance changes during mental stress testing as an index of autonomic arousal in cardiovascular research," *Amer. Heart J.*, vol. 128, pp. 1170–1177, 1994.

[17] R. W. Picard and J. Scheirer, "The galvactivator: A glove that senses and communicates skin conductivity," in *Proc. 9th Int. Conf. Human-Comput. Interact.*, New Orleans, LA, Aug. 2001, pp. 1538–1542.

[18] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[19] J. Zhai and A. Barreto, "Stress detection in computer users based on digital signal processing of noninvasive physiological variables," in *Proc. 28th IEEE EMBS Annu. Int. Conf.*, New York, 2006, pp. 1355–1358.

[20] M.-H. Lee, G. Yang, H.-K. Lee, and S. Bang, "Development stress monitoring system based on personal digital assistant (PDA)," in *Proc. 26th IEEE EMBS Annu. Int. Conf.*, Sep. 2004, pp. 2364–2367.

[21] C. D. Katsis, G. Ganiatsas, and D. I. Fotiadis, "An integrated telemedicine platform for the assessment of affective physiological states," *Diagn. Pathol.*, vol. 1, pp. 16–19, 2006.

[22] W. Liao, W. Zhang, Z. Zhu, and Q. Ji, "A real-time human stress monitoring system using dynamic Bayesian network," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR2005)—Workshops*, Washington, DC, 2005, p. 70.

[23] Brainclinics. (2010). [Online]. Available: http://www.brainclinics-products.com

[24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley–Interscience, 2000.

Cornelia Setz received the Master's degree in electrical engineering and information technology in 2006 from the Swiss Federal Institute of Technology, Zurich, Switzerland, where she is currently working toward the Ph.D. degree in the Wearable Computing Laboratory.

During her studies, she has specialized in signal processing and control while gaining work experience at Phonak AG for two years. At the Wearable Computing Laboratory, she is engaged in the European SEAT Project that tries to enhance airplane passenger's comfort by incorporating unobtrusive sensors into the seat. Her current research interests include multidisciplinary research bringing new technologies into the field of psychology.

Bert Arnrich received the Dipl.-Inform. (M.Sc.) and Dr.-Ing. (Ph.D.) degrees from the University of Bielefeld, Bielefeld, Germany, in 2001 and 2006, respectively.

He was involve in medical data mining, machine learning, and statistical modeling in close cooperation with a heart institute. He developed a data mart system in order to discover patterns of heart disease, patient's response to treatment, and real-time risk stratification. Since 2006, his research on pervasive computing technologies is directed toward paving the way for a pervasive, user-centered, and preventive healthcare model. He is currently a Senior Researcher in the Wearable Computing Laboratory, Swiss Federal Institute of Technology, Zurich, Switzerland. Within national and European research projects, he is engaged in personal health systems.

Johannes Schumm received the Dipl.-Ing. degree in electrical engineering from the Rheinisch-Westfaelische Technische Hochschule Aachen University, Aachen, Germany, in 2006.

He is currently a Ph.D. student in the Wearable Computing Laboratory, Department of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology, Zurich, Switzerland. His research interests include pervasive healthcare and reliability of physiological monitoring systems.

Roberto La Marca received the M.Sc. and Ph.D. degrees in psychology from the University of Zurich, Zurich, Switzerland, in 2005 and 2009, respectively.

He is currently a Teacher and a Research Associate in the Department of Clinical Psychology and Psychotherapy, University of Zurich, and in the Wearable Computing Laboratory, Swiss Federal Institute of Technology, Zurich, Switzerland. He is involved in several projects about biopsychological stress responses. He has focused on salivary and cardiorespiratory markers of the hypothalamus-pituitary-adrenal axis, and the sympathetic and parasympathetic nervous system, with a special interest on cardiovagal activity. His research interests include the interaction of several physiological systems in healthy subjects under laboratory and real-life conditions.

Gerhard Tröster (M'93) received the Diploma in electrical engineering in 1979 from the Technical University (TU) Karlsruhe, Karlsruhe, Germany, the Doctorate degree from the TU Darmstadt, Darmstadt, Germany, in 1984.

Since 1993, he has been a Full Professor of electronics at the Swiss Federal Institute of Technology, Zurich, Switzerland, where he heads the Electronics Laboratory, established the Multichip Module Electronic Packaging Group, and founded the Wearable Computing Laboratory in 2000, focusing on interdisciplinary approach combining IT, signal processing, electronic platforms, wireless sensor networks, smart textiles, and human computer interaction. He was with Telefunken (Atmel), Heilbronn, for eight years, where he was engaged in design methods of analog/digital systems in CMOS and BiCMOS technology.

Ulrike Ehlert has been a Full Professor of clinical psychology at the Institute of Psychology, Clinical Psychology and Psychotherapy, University of Zurich, Zurich, Switzerland, since 1999. In 2000, she set up a workgroup on human behavioral neurobiology at the Neuroscience Center, University of Zurich, and the Swiss Federal Institute of Technology, Zurich, Switzerland. Because of her former work as the Head of a psychological consultation-liaison service of two general hospitals and her current work as the Head of the Outpatient Clinic and Research Unit for Behavioral Medicine, University of Zurich, she has a broad range of experience in diagnosis and psychotherapeutic treatment of stress related disorders. In addition, she was engaged in the field of psychobiological research into psychiatric and psychosomatic disorders for many years. Her research interests include psychobiological mechanisms underlying stress-related disorders. She has published various textbooks and reviews on diagnosis and treatment in behavioral medicine, and has assessed many different studies related to the etiology of stress-related disorders.