Recognition of Time Pressure via Physiological Sensors: Is the User's Motion a Help or a Hindrance? *

Margeritta von Wilamowitz-Moellendorff
Department of Psychology
Saarland University
Saarbrücken, Germany
mawimex@gmx.de

Christian Müller
Department of Computer Science
Saarland University
Saarbrücken, Germany
cmueller@cs.uni-sb.de

Anthony Jameson
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
jameson@dfki.de

ABSTRACT

The recognition of a user's internal states via physiological sensors is sometimes seen as a matter of detecting the direct physiological correlates of the internal states. This type of detection can be problematic when a user is moving around, as is often the case with today's mobile systems. We present a study which illustrates that detection of internal states is sometimes actually easier when the subject is moving: The affective state may be associated with overt behavior that results in detectable changes in the physiological variables.

1 INTRODUCTION

One way of assessing a user's affective state is via physiological sensors. With unobtrusive sensors, the system can obtain data without disturbing the user, collecting continuous data about her internal states and reactions to the environment.

In the present study, the variables of interest are time pressure and movement around the environment. As time pressure has proved to be one of the most common stressors in the work environment, one which has an impact on available processing capacity, estimation of and adaptation to such pressure could be beneficial. The variable of movement is becoming important as more and more mobile assistance systems support their users in mobile contexts. Detection of the user's movements is often straightforward, but the variable is of interest here as a moderating variable that is likely to affect the extent to which time pressure is detectable. The combination of the two variables in a 2×2 design allowed us to address this question directly within one particular context (see Figure 1, which will be explained below).

2 RELATED PREVIOUS WORK

There is a growing interest in the recognition of the internal states of users through physiological sensors. Research in the field of emotion recognition is especially relevant to the question of whether time pressure is identifiable on the basis of sensor data.

Picard, Vyzas, and Healey (2001) tried to discriminate among 8 different emotions played by an actor over several days using Fisher Projection as a basic method. They were successful in identifying especially the emotions of anger, sadness, joy, and adoration. Sensor fusion, the use of a several different sensors simultaneously, made a correct prediction possible, whereas no sensor alone made good predictions.

Lisetti, Nasoz, LeRouge, Oyzer, and Alvarez (2003) were successful in discriminating among 6 emotions induced by pictures. Anger and sadness were among the most recognizable emotions that they found using the k-nearest-neighbor algorithm. The measures heart rate, temperature, and galvanic skin response were used to discriminate among the emotions.

Greenwald, Cool, and Lang (1989) used the "international affective picture system" to provoke emotions, and they measured muscle tension from different muscles in the face. They compared the self-attributed feelings of arousal and valence (Lang, 1968, 1978) to changes in muscle tension and found an increase of the tension of the corrugator supercilii (muscle over the eyebrow) with pictures of negative valence. They noticed an increase in heart rate during the presentation of pictures with positive valence, and electrodermal activity correlated positively with arousal induced by pictures.

A study involving physiological measures was conducted by Conati (2002) in the context of frustration recognition during work with educational games. Electrodermal activity was used as an indicator of arousal, and an increase in the heart rate and the tension of the corrugator supercilii served as indicators of frustration and negative valence. Through a combination of the information from all sensors in a dynamic Bayesian network, the state of the user could be estimated.

Research on stress detection through sensors is important because time pressure has somewhat similar effects on human physiology as stress. Experiments conducted with pilots (Ylonen, Lyytinen, Leino, Leppaluoto, & Kuronen, 1997) used especially heart rate as an indicator of stress. In a complex study by Healey and Picard (2002), the stress induced by driving a car was examined. The 10 subjects drove the same 90-minute fixed route, while their electrocardiogram, electromyogram, electrodermal activity, and respiration were being registered. The route was divided into different categories of stress. Through sensor fusion, an identification accuracy of 88.6% for the stress level was achieved.

The recognition of mental load, which typically increases with time pressure, is an area of research in which Chen and Vertegaal (2004) conducted an interesting experiment about interruptibility. Two indicators were used, one to detect whether the person was sitting or moving and the other to detect whether the person was men-

*The fourth and fifth authors are listed in the final section of the paper. This research was supported by the German Science Foundation (DFG) in its Collaborative Research Center on Resource-Adaptive Cognitive Processes, SFB 378, project READY; and by the German Ministry of Education and Research (BMBF), Project SPECTER.

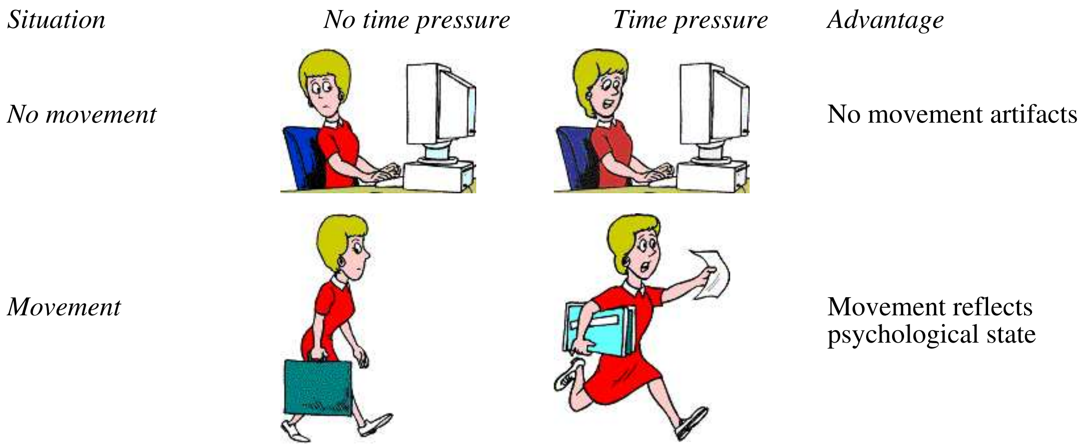

Figure 1. *Overview of the experimental design.*

tally busy or not. For the detection of movement, an electroencephalogram (EEG) sensor was used, which allows a good detection of motor-related activity. For the measurement of mental load, the sensor data from the electrocardiogram was transformed via a power spectrum analysis. The authors found a tendency toward increased heart rate variability in the lower-frequency ranges (< 0.1 Hz) of the power spectrum.

3 CENTRAL ISSUE OF THIS STUDY

The prior research in this field has promising results for the detection of emotion, stress, and mental load through sensors, especially through measures like the heart rate variability of the lower frequency range, electrodermal activity, and the tension of the corrugator muscle. But many of the experiments were conducted under artificial conditions. Emotions were induced by extreme pictures or played by an actor. The subjects were almost always stationary. In our experiment, we wanted to use a task with a strong relationship to everyday work, and we wanted to see how well a moving user could be dealt with. The project is based on the idea of a mobile assistance system that is able to detect the user's state even in a mobile context. But movement often produces movement artifacts in sensor data that make the data hard to to interpret. We therefore looked for measures that not only are unobtrusive, so that they don't bother the moving user, but also provide robust signals for interpretation. The overall question is whether the recognition of time pressure during movement is easier (or more difficult) than in a condition in which the subject is sitting still. Although a single experiment can investigate this question only in one particular context, we hope that the methods and results will be found stimulating for further work on this question.

4 PHYSIOLOGICAL SIGNALS

The following sensors and quantitative indices were used in our study:[1]

[1] Further details can be found in the master's thesis of von Wilamowitz-Moellendorff (2005).

Electrodermal Activity

We measured the electrical resistance of the skin of the nonactive hand of the subjects during the entire experiment. The skin conductance level was monitored in reference to the baseline recording of the first 15 minutes before the main part of the experiment started. All data points were z-transformed so as to make the subjects comparable, because not only the level of each subject was different but also the variability in the signals. This signal is relatively free of movement artifacts.

Electromyogram

We recorded the electrical potential of the muscle cells in two different regions:
First, it was recorded on the arm of the subject, with the goal of registering the tension of the nonactive arm to detect movement. The data had to be standardized, because of large differences between the subjects.
Second, two electrodes were attached on the forehead, over the left eyebrow. The tension of the corrugator supercilii is known to increase with stress and frustration. The problem with the recorded data was that not only the tension of the eyebrow muscle (through frowning) had an impact on the data but also blinks and eye movements. Since the pattern of frowning was virtually undetectable, this variable ultimately had to be excluded from the analyses.

Electrocardiogram

The electrocardiogram (ECG) signal was measured from the manumbrium sterni and the lower left costal arch. The most basic index derived from the ECG signal was the heart rate, defined as the number of R-spikes per minute. The heart rate is sensitive to movement and relaxation. It was calculated with an algorithm that detects the QRS complex.
Heart rate variability concerns the regularity of the interval between successive heartbeats. It was measured through the mean square of successive differences, which is defined as the accumulated dissimilarity of the durations in the sequence of interbeat intervals. This measure is known typically to decrease during attention-demanding tasks (Schandry, 1988).
A related measure is based on the power spectrum of the electrocardiogram. A tendency toward increased heart rate variability in the lower frequency ranges (< 0.1 Hz) was expected to accompany

stress and time pressure (Chen & Vertegaal, 2004).

Recording Device

The recording device used to detect the incoming sensor data was the Varioport from Becker-Meditec. The recorder is about as tall as a packet of cigarettes, and it is able to register the incoming signals from the four sensors. The data was was registered with a rate of 256 Hz for about one hour. The Varioport is so small and light that it could be fixed to the subjects' clothes and was carried during the experiment without problems.

5 METHOD

Design. The main goal was to determine under which circumstances a better prediction of time pressure is possible: in a stationary context or when a person is moving. To make a direct comparison, all four combinations of movement (vs. no movement) and time pressure (vs. no time pressure) were realized for every subject. The task was divided into four parts. Two parts were designed so that they had to be dealt with on the computer, and two parts were designed so that the subjects had to move around an office in order perform the task. The other distinction was that half of the tasks were without any time limit and the other half had a time limit of 5 minutes (money was subtracted from the payment when the time was exceeded, at the rate of 2 euro cents a second).
Figure 1 gives an overview of the 4 conditions. The conditions were presented to the subjects in 8 different sequences so that order effects could be avoided.

Material. The task that was used for the experiment is a so-called *inbox task*, a test task often used in German assessment centers. This instrument is close to normal everyday work life in that it simulates the act of organizing incoming letters (or emails) and making quick decisions about what to do about them. In all, 16 different letters were presented to each subject; 1 letter was presented at the beginning to allow the subject to get familiar with the task.

Subjects. In this study, 18 subjects took part, 8 of them female and 10 male students. Each subject invested between 1.5 and 2 hours for the whole experiment. The subjects were paid for their participation. To make the time limit conditions more effective , money was distracted from the reward (the money loss being displayed by a watch on the monitor counting backward in money after the time limit had elapsed).

Procedure. Each subject first performed a small part of the inbox task so as to get accustomed to the task and to produce a baseline value. They then performed each of the four parts of the task as was sketched above. Finally, they filled in a questionnaire that asked about their subjective responses to various aspects of the experiment, and they were debriefed.

Hypotheses. We expected the electrodermal level to increase during the conditions in which time pressure was induced. In previous research, Wang, Prendinger, and Igarashi (2004) had found a correlation between stress and the electrodermal response of the skin. Also, an increase in the electrodermal level was expected during the movement conditions, because of the bodily reaction to movement. The EMG sensor attached to the forearm was expected to register a greater standard deviation due to the activity of the arm during the movement conditions.
It was hypothesized that heart rate is associated with movement and should increase during the two movement conditions.
Another measure from the electorcardiogram was calculated through a power spectral analysis. It was expected that under time pressure there would be a tendency for increased heart rate variability in the low frequency ranges of the ECG data (Chen & Vertegaal, 2004).
The muscle tension of the corrugator supercilii was expected to increase while the person was under time pressure and was feeling stressed. Conati (2002) had found an relationship between frustration and tension of the muscle over the eyebrow.

Storage and processing of sensor data. Incoming signals from four sensors with a storage rate of 256 Hz had to be managed. The enormous amount of data from each one of the 18 subjects required the use of a special database. This database (developed by the second author) was adapted to fit the data from this experiment. About 3,686,400 data points per subject were saved in the LDAAT (Large Data Amount Analysing Toolkit). LDAAT is based on PHP (a modern script language for the creation of websites) and MySQL (a database implementation); it makes it possible to execute analyses easily through a web interface. The database could perform analyses fast, because it made use of a cluster of computers. The data from the Varioport memory card had to be read out and be transformed into ASCII files. Calculations were made via SQL commands. Formal conversions and transformations were also carried out with the database; for example, the z-transformation (every data point minus the overall average for the person in question divided by the standard deviation) of the raw data. A special command executed the segmentation of the raw data into intervals of the desired length (for example 1 minute), for operations like the calculation of the heart rate per minute.

6 RESULTS

Two different analyses were performed to study the data. First a multivariate analysis of variance (MANOVA) was conducted to test whether the two factors "movement" and "time pressure" had an effect on the dependent variables. Second, a logistic regression was conducted to test the possibility of identifying the subject's condition on the basis of the recorded data.

6.1 Results of the Multivariate Analysis of Variance

A multivariate analysis of variance (MANOVA) is used when there is more than just one dependent variable. This method helps to determine whether changes in the independent variables have effects on the dependent variables and if there are interactions between the independent variables. The (two-level) factors are time pressure and movement. The dependent variables are electrodermal response, electromyogram, heart rate, power of the spectral density analysis, and heart rate variability.
The main effect for the factor "movement" became significant with $F(1,15) = 10.00, p < 0.05$. Neither the main effect for time pressure nor the interaction became significant. But there are differences between the four conditions for every dependent variable, a fact that encouraged made us take a closer look at the univariate results. Figure 2 shows the results for each dependent variable.
The results of the electrodermal activity show a tendency toward an interaction between movement and time pressure. The combination of time pressure and movement has the highest values, and

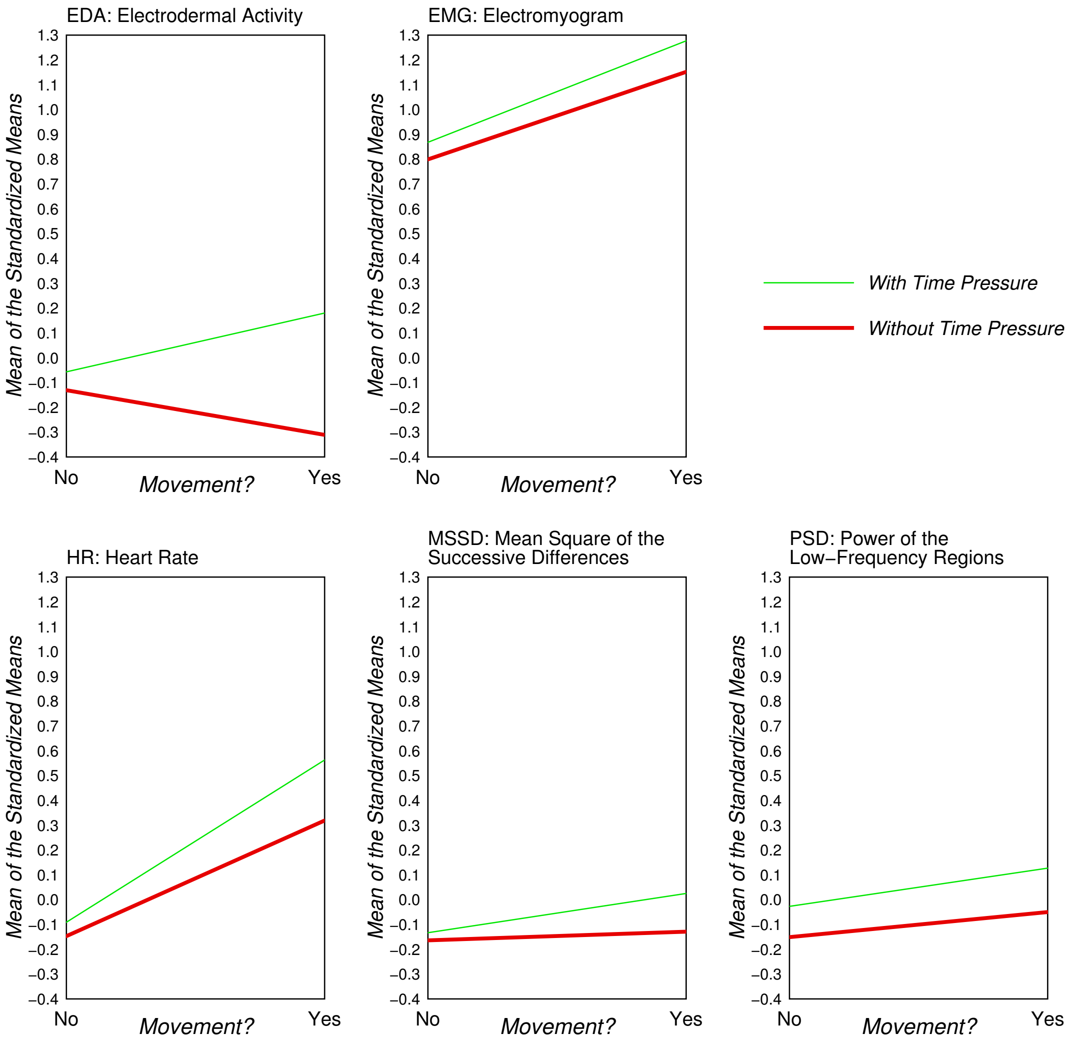

Figure 2. *Mean standardized values of the various indicator variables under each of the four experimental conditions.*

the combination of no time pressure and movement the lowest average. This pattern is not consistent with our original hypotheses, but it is consistent with the ratings of the subjects in the questionnaire, in which they reported that the situation with movement and no time pressure induced the least cognitive load.

The standard deviation used as indicator with the electromyogram signal shows the expected tendency, an increase during movement ($F(1, 15) = 33.3, p < 0.01$). It is possible that with a higher number of subjects this effect would become significant in the MANOVA.

The effect for heart rate went in the expected direction, too, showing an increase during movement. Even a small increase with time pressure is observable in the data.

The heart rate variability (MSSD) did not show the expected decrease with time pressure. The slight increase with time pressure might be ascribed to the movement artifacts, which make more difficult the correct identification of the QRS complex.

The results of the Power Spectral Density Analysis show no significant tendency of the power in the low-frequency range to increase. But in the graph, a small tendency in the direction of the hypothesis is visible, which might become significant with a larger group of subjects.

6.2 Detecting the User's State

The results just discussed give us some idea of the diagnostic value of the indices used. But the real question of interest is whether it is possible to detect a user's time pressure (and movement) on the basis of the sensor data.

There are a variety of possible approaches to this user modeling problem. For example, various machine learning techniques could be applied to the learning of classifiers, and the data could be preprocessed in various ways. A model could be learned for each user individually or for users in general. We are currently exploring various possibilities (e.g., pattern classification based on Bayesian decision theory, cf. Duda, Hart, & Stork, 2000); but for now we present a relatively simple analysis based on the technique of logistic regression.

In general, a logistic regression produces a model for predicting the value of a dichotomous variable (e.g., presence vs. absence of time pressure) on the basis of a set of continuous predictor variables. We used as predictors simply the mean values, for each subject and

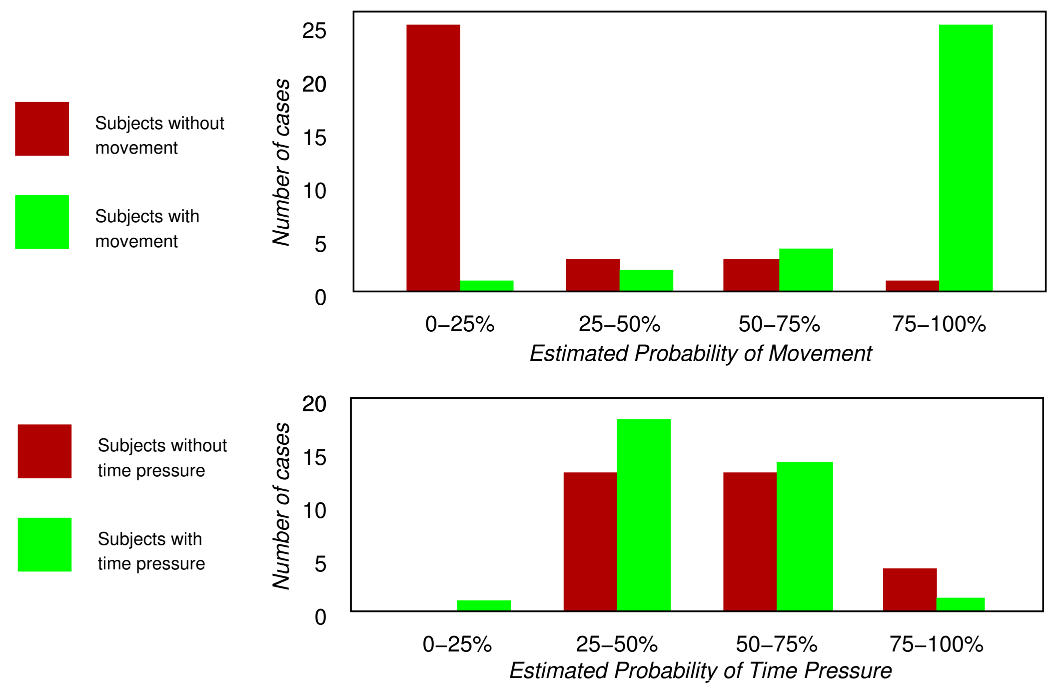

Figure 3. *Overview of the predictions made via logistic regression concerning the question of whether a subject was moving or not (top) or whether she was under time pressure or not (bottom). (For each probability interval (e.g., 0–25%), the histogram shows, for those subjects for whom the model made a prediction with a level of confidence in that interval, how many subjects were actually in each of the conditions.)*

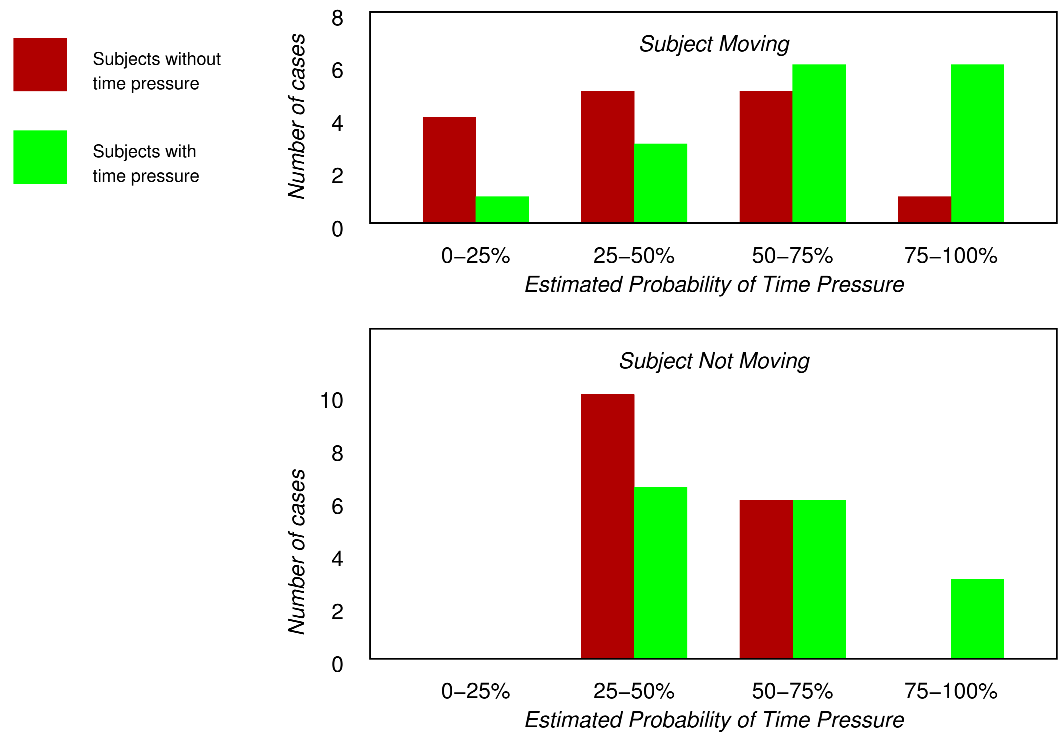

Figure 4. *Overview of time pressure predictions made via logistic regression with and without movement by the user, respectively.*

condition, of the standardized variables shown in Figure 2. The question then becomes: Given the mean values for a given subject in a given condition, how accurately can you predict whether this subject was under time pressure (or in motion)?

The results of the MANOVA already suggest that a detection of movement through sensors should be easy. The first histogram in Figure 3 confirms this expectation. Overall, a correct prediction was made in 89.1% of the cases.

The structure of the histograms in this figure can be explained easily for this simple case. With excellent predictability, we would see only one tall, lighter-colored bar in the right-hand interval and one darker bar on the left. This situation is closely approximated for the prediction of movement.

With the similar logistic regression for the prediction of time pressure, the five indices predict time pressure with only 56.3% accuracy overall, just a little better than the chance level of 50%. In the histogram, the poor predictability is reflected in the clustering of the bars in the middle two intervals.

But perhaps time pressure can be more accurately predicted if we restrict ourselves to cases where it is known that the user is moving (or not moving). Figure 4 shows the results of the two relevant logistic regression analyses. As is reflected in the lower histogram, the accuracy of predictions when the subject is not moving is still very poor, only 56.3%.[2] On the other hand, in the conditions with movement, time pressure but was easier to distinguish, with an average accuracy of 71.9%.

One way of checking whether this difference might be due to chance is to count the cases in which the system made a correct prediction with high confidence. (These are the cases in which it would be reasonable for a system to act on the basis of a prediction.) There are 10 such cases (out of 36 possible cases) in the conditions with movement and only 3 such cases (out of 36) in the conditions without movement. The difference between these two frequencies is significant according to a two-tailed χ^2 test (χ^2 = $5.57, p < .05$).

7 DISCUSSION

A likely explanation of the somewhat better detection of time pressure during movement is as follows: The direct physiological effects of time pressure are in themselves rather subtle, too small to be picked up reliably by physiological sensors. When the user is moving, these subtle direct effects are even more difficult to detect; but on the other hand, there are indirect effects that result from the way in which time pressure influences the subjects' movements. The subject tends to move faster in order to finish the task quickly enough, and this difference is in turn reflected in the physiological indices.

Further work will be required to reveal the extent to which similar patterns appear in other contexts and with other psychological states. The main result of the current study is a straightforward illustration of the fact that internal states can sometimes be recognized better via physiological sensors on the basis of their expression in behavior than on the basis of their direct physiological effects. This general point can be taken into account whenever we think about the use of physiological sensors for the estimation of internal states in real-world situations.

[2]The three cases on the right in which the system was able to make a correct prediction of time pressure with high confidence suggest that some users may exhibit unusually strong symptoms of time pressure even when they are sitting still; but this possibility would have to be explored with a larger number of subjects.

8 ADDITIONAL AUTHORS

Additional authors: Boris Brandherm and Tim Schwartz, Department of Computer Science, Saarland University, brandherm@cs.uni-saarland.de / schwartz@cs.uni-saarland.de.

REFERENCES

Chen, D., & Vertegaal, R. (2004). Using mental load for managing interruptions in a physiologically attentive user interface. In *Extended Abstracts for CHI'04* (pp. 1513–1516). Vienna.

Conati, C. (2002). Probabilistic assessment of user's emotions during the interaction with educational games. *Journal of Applied Artificial Intelligence*, *16*(7-8), 555–575.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification* (2nd ed.). New York: Wiley.

Greenwald, M. K., Cool, E. W., & Lang, P. J. (1989). Affective judgement and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology*, *3*, 51–64.

Healey, J. A., & Picard, R. (2002). Smartcar: Detecting driver stress. In *Proceedings of ICPR 2000 (Vol. 4)* (pp. 218–221).

Lang, P. J. (1968). Fear reduction and fear behaviour: Problems in treating a construct. *Research in Psychotherapy*, *111*, 90–103.

Lang, P. J. (1978). Anxiety: Toward a psychophysiological definition. In H. S. Akiskal & W. L. Webb (Eds.), *Psychatric diagnosis: Exploration of biological predictors* (pp. 365–389). New York: Spectrum.

Lisetti, C., Nasoz, F., LeRouge, C., Oyzer, O., & Alvarez, K. (2003). Developing multimodal intelligent affective interfaces for tele-home health care. *International Journal of Human-Computer Studies*, *59*, 245–255.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(10), 1175–1191.

Schandry, R. (1988). *Lehrbuch Psychophysiologie.* Munich: Psychologie Verlags Union.

von Wilamowitz-Moellendorff, M. (2005). *Assessment of time pressure via physiological sensors with and without movement.* Unpublished master's thesis, Saarland University, Department of Psychology.

Wang, H., Prendinger, H., & Igarashi, T. (2004). Communicating emotions in online chat using physiological sensors and animated text. In *Extended Abstracts for CHI'04.* Vienna.

Ylonen, H., Lyytinen, H., Leino, T., Leppaluoto, J., & Kuronen, P. (1997). Heart rate responses to real and simulated BA Hawk Mk 51 flight. *Aviation, Space, and Environment Medicine*, *68*, 601–605.