DRIVER WORKLOAD ESTIMATION BASED ON REALISTIC IN-VEHICLE SENSORS

Florian Ripper
Ardit Dvorani
Christoph Hemprich
Koshan Mahdi
Lukas Scholz
Joyson Safety Systems (JSS)
Germany

Leonard Cech
Joyson Safety Systems (JSS)
USA

Paper Number 19-0325

ABSTRACT

The introduction of automated driving motivates the need for driver state detection, prediction and monitoring. The first introduced systems for automation (SAE level 2 or 3) rely on the ability of the driver to resume the driving task, both manually (steering control) and visually. The automation system monitors the presence of the driver's hand(s) on the steering wheel and some systems monitor the driver's visual attention as well. This is necessary to ensure that the driver monitors the automated drive and responds manually and/or visually to HMI cues from the automation system.

Considering higher levels of automated driving (SAE level 3 or 4), several challenges and opportunities arise. First, as driver confidence in the performance of automation grows, they will switch attention to secondary tasks increasing their cognitive workload. The driver may be busy doing other tasks and a system which estimates cognitive workload can indicate that a longer take-over period is necessary, or, in the extreme, a human driver take-over may not be safe or even feasible. Secondly, the driver may be visually, manually or mentally overloaded (or a combination of these) during manual control and the automation system might encourage or intervene automation modes to enhance safety. These two use cases can be improved using accurate real-time prediction of the driver's mental workload using one or more in-vehicle sensors.

In this paper a robust method for estimating driver workload based on real in-vehicle sensors is presented. Sensors in the seatbelt and at the steering wheel rim derive heart rate metrics which are used to estimate cognitive workload. Furthermore, an analysis is conducted to determine if additional metrics derived from vehicle dynamics data have an impact on the calculation accuracy. Comparing individual-based driver classification approaches versus a generalized driver algorithm is also part of this investigation. A driving simulator study with n-back task induced workload is used to validate the driver cognitive workload estimation method accuracy.

INTRODUCTION

The introduction of automated driving motivates the need for driver state detection, prediction and monitoring. The first introduced systems for automation (SAE level 2 or 3) rely on the ability of the driver to resume the driving task, both manually (steering control) and visually. The automation system monitors the presence of the driver's hand(s) on the steering wheel and some systems monitor the driver's visual attention as well. This is necessary to ensure that the driver monitors the automated drive and responds manually and/or visually to HMI cues from the automation system.

As higher levels of automated driving (SAE level 3 or 4) are introduced, several challenges and opportunities arise. First, as drivers gain confidence in the performance of automation, switching to secondary tasks (for example, reading, talking, cellphone manipulation, playing games, etc.) will re-direct and/or increase cognitive workload. In these situations, a system which continuously estimates cognitive workload can indicate that a longer take-over period is necessary, or a human driver take-over may not be safe or even feasible [1]. Secondly, the driver may be visually, manually or mentally overloaded during manual control and the automation system might encourage or intervene automation modes to enhance safety. These two use cases require accurate real-time prediction of the driver's mental workload based on one or more in-vehicle sensors.

Societal and technology changes such as cellular phone proliferation, adoption of advanced driver assistance systems and the nearly continuous availability of virtual social engagement are also contributing factors which drive ongoing research in this field coupled with increased attention by safety regulatory agencies. As an example, EuroNCAP has acknowledged the importance of advanced driver monitoring measures by putting a special emphasis on this topic in their 2025 Roadmap [2].

STRESS, MENTAL WORKLOAD, & DISTRACTION

Stress, fatigue and proclivity for risk-taking behaviors are some examples of human factors that limit effective driver performance. Cognitive workload, defined synonymously as mental workload represents the measure of effort applied towards cognition (e.g. thinking, understanding, perception, reasoning, learning ...). Therefore, it can be expected that driver mental workload is affected in all these situations. Mental workload is a volatile concept which must be interpreted carefully in each situation. Simultaneous secondary tasks to the primary driving task increase mental workload and decrease the driving performance efficiency [3]. Furthermore, increased mental workload can occur in situations with monotonous traffic environment and is dependent on driver experience or presence of circumstantial contexts like fatigue or sickness [3]. Mental workload can be conceptualized in terms of information processing capability and demand. Information processing includes cognitive as well as motivational and emotional aspects. Each individual person will evaluate the demands which they have to cope with and self-regulate their cognitive effort for processing these demands [4].

Several papers have shown that predicting mental workload in an automobile environment can be achieved by deriving and combining vehicle performance measures, video based (behavioral) measures and driver physiological measures [5]. When a secondary task involves visual cues in the cockpit, for example glances to the center stack display or other locations within the vehicle, behavior detection by means of a driver monitoring camera is very efficient. Measuring the percentage of gaze time directed towards the forward road center, as described in the NHTSA distraction guidelines [6], provides a good way to track visual attention to secondary tasks. This paper is based on a study design which concentrates on mental workload apart from visual distraction and investigates physiological measures for further analysis. In this study, periods of high mental workload are induced through a graded mental secondary task (n-back) [7].

While a variety of physiological measures have been studied in the context of mental workload estimation; electrocardiogram (ECG) is one of the most widely utilized measures. Metrics derived through the ECG signal, specifically heart rate (HR) or inter-beat-interval (IBI) directly reflect activity of the autonomous nervous system (ANS). Several past studies indicate a relation between the ECG signal and mental workload in a driving environment. For example, Mehler et al (2009), Rodriguez-Ibanez et al (2012), and Fallahi et al. (2016) all indicate a decrease in mean IBI with increased cognitive workload and an increase in mean IBI with increased drowsiness [5, 7, 11, 12] for drivers in these studies.

Another promising data source for estimating mental workload is the dynamic response of the vehicle itself. The effects of cognitive workload on a driver can be measured indirectly through data obtained from the vehicle dynamics data, for example, vehicle average speed or steering wheel rotation angle [5].

HYPOTHESIS

In this paper examination of vehicle dynamics data and physiological data is carried out in order to establish a classifier-system for driving with varying cognitive load. A related literature search indicates, that there are few studies with repeating subject measures (e.g. same subject over several days) in a driving environment with varying cognitive load. Repeated trials include effects of circumstantial contexts and may reveal differences in driving performance over time. Previous studies have indicated that the ANS reacts sensitively to different cognitive load conditions in comparison to a baseline for between subject comparisons [5]. A premise for this study was to include repeated trials to derive a classifier system for mental workload which includes physiological data depending on the (initial) state of the ANS for each individual driving subject. The following hypotheses are proposed in this paper:

- Driver ANS will react sensitively to cognitive load conditions between and within subjects (increasing HR)
- Vehicle dynamics performance data and driver physiological (e.g. ECG) data show complementary trends during cognitive load conditions
- Accuracy of a feature-based classification is higher when applied to an individual subjects' dataset (repeated measures) than when used across multiple subjects
- Physiological and driving performance data is affected by time of day (difference between morning and evening trials)

DRIVING SIMULATOR STUDY

For this study a small fixed-base driving simulator was used, consisting of a dashboard with cockpit display and touch screen center console, a steering wheel and a seat from a mid-sized vehicle. SILAB Version 5 by WIVW is used as the driving simulation software. The simulation is displayed on a 55-inch TV.

Five test subjects participated in a repeated measure driving study. The test subjects were randomly selected from JSS Berlin, Germany staff and consisted of four male and one female test subject aged 25 to 50. The test drives for each subject were spread over a period of five consecutive days with one trial in the morning and one trial in the evening; resulting in 10 trials in total. At the beginning of the experiment the subjects were asked to complete a 20-minute test drive to become familiar with the secondary cognitive tasks and characteristics of the driving simulator. None of the subjects was affected by simulator sickness (assessed with a simulator sickness questionnaire) [14]. The subjects were directed to focus on the primary driving task and to follow local traffic laws.

Each trial was designed to emulate a commute from home to work (morning) or work to home (evening). The trial consisted of a 30 min drive mainly through rural roads, separated into five 6-minute-sections each followed by a questionnaire presented on the center stack touch screen display. The first and the last 6-minute sections were baselines with no secondary cognitive tasks. The three middle sections included an audible n-back task presented with one of three difficulty levels to raise the total cognitive task demand on the driver. These audible secondary tasks do not conflict with the manual control or visual processing demands of the primary driving task [7]. The presented n-back tasks were 0-back, 1-back and 2-back in pseudo random order with the evening drive consisting of the exact same order as the morning drive. The driver section profile and secondary task structure are shown in Figure 1.

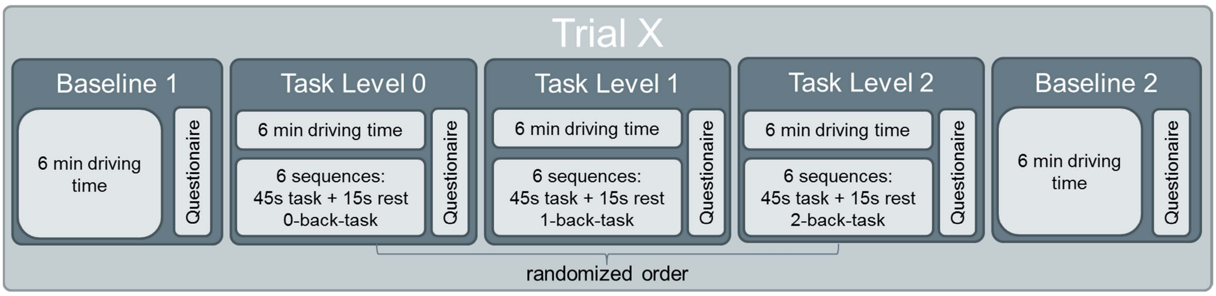

Figure 1. Test procedure per trial.

An ECG reference sensor was used to validate the heart rate measurements of the in-vehicle sensors. The reference sensor is an internal development of JSS and was previously validated against an industry/university standard reference equipment. The JSS reference is portable and uses gel electrodes placed on the bare chest of the test subject. It records a 24-bit ECG with 1 kHz sample rate and calculates the IBI (Interbeat Interval) with an accuracy of ±2mn.

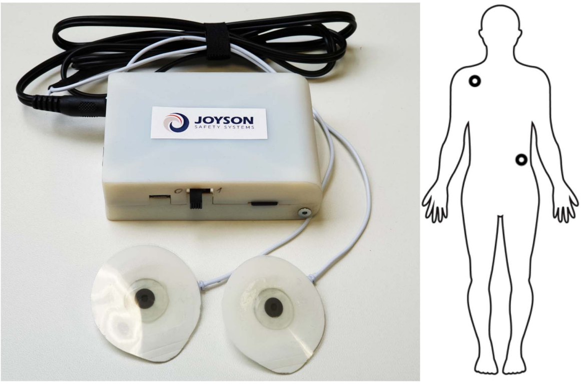

Figure 2. JSS Vital Recorder and electrode position.

VITAL SIGN SENSORS FOR IN-VEHICLE APPLICATIONS

JSS is a leading supplier of steering wheels, including several production wheels which include hands-on-wheel detection and classification sensors. JSS also developed one of the first production driver monitoring cameras which supports automated driving (SAE Level 2+) in passenger vehicles through facial head pose, eyelid closure and eye gaze tracking. As the importance of vital sign sensing for advanced driver state detection continues to evolve, JSS has extended its portfolio to deliver steering wheels capable of sensing driver physiological (vital) signals such as ECG. This steering wheel is designated the Vital Sign Steering Wheel and was used in this study.

The steering wheel incorporates two electrode zones on the sides of the steering wheel rim surface and senses a 2-lead electrocardiograph (ECG) of the driver's heart activity. The inter-beat-interval (IBI) derived from the ECG has an accuracy typically in the range of ±4ms. This signal can be used to derive typical measures of heart rate variability such as the Root Mean Square of Successive Interbeat Interval Differences (RMSSD) or the number of successive IBIs that differ by more than 50ms (NN50). Both measures have been used in literature as a metric for estimating mental workload. Figure 3 shows an example of the raw ECG output of the Vital Sign Steering Wheel and the derived IBI signal. Figure 4 shows a higher resolution example of the steering wheel ECG signal for several heart beat cycles which highlight the signal deflections characteristic of the QRS complex.

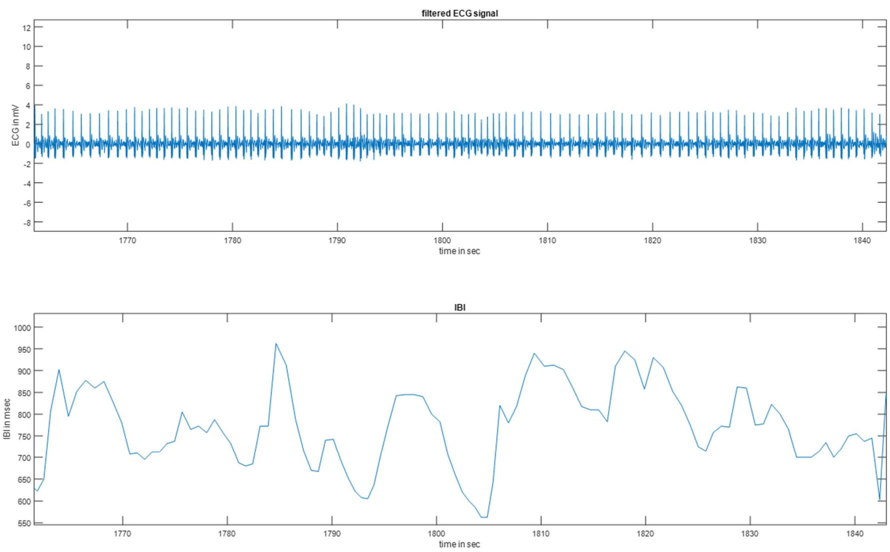

Figure 3. Vital Sign Steering Wheel ECG signal (top) and derived IBI time series (bottom).

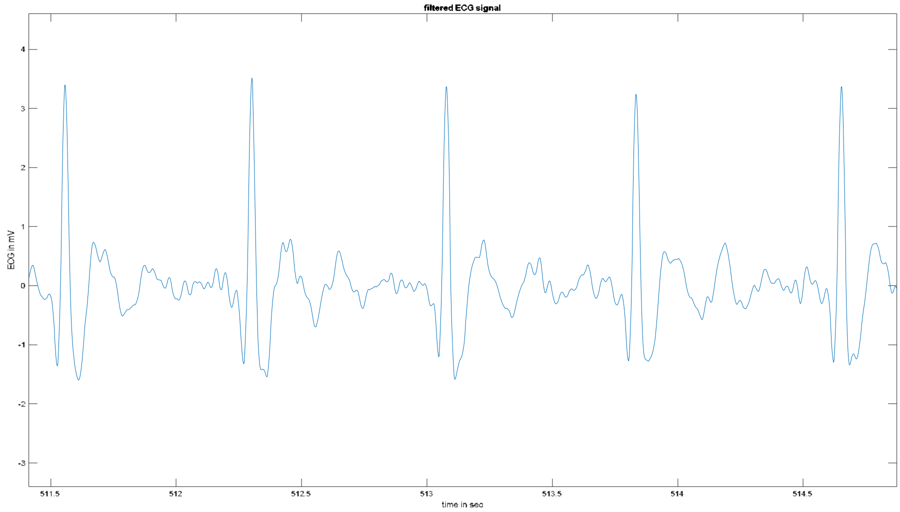

Figure 4. Zoom into the output signal of the Vital Sign Steering Wheel showing the QRS peaks.

In this study, a prototype Vital Sensing Seatbelt was included to complement the vital sensing steering wheel. This seatbelt uses a very thin sensor inlay embedded into the thorax section of the 3-point seatbelt and is capable to sense the heartbeat signal for a stationary occupant. As automated driving evolves, vital sensing in the seatbelt is essential as the hands of the driver are no longer on the steering wheel. This system can provide status assessment for drivers (or occupants) during periods of automation and can provide information to support intelligent driver/automation take-over.

Based on the sensing principle, the prototype outputs a mean heart rate (HR) estimate over a sliding 30 second window but cannot be used to derive an accurate IBI signal. As automation evolves, there will be a need to assess driver cognitive state when the hands are not placed on the steering wheel. The Vital Sensing Seatbelt can provide driver HR during these periods of fully automated control and improve feature estimates if used in combination with the Vital Sign Steering Wheel when the driver is in control.

It is important to note that in real world driving conditions, the signal of the Vital Sensing Seatbelt also includes noise induced by the road/wheel vibration and occupant movement. Accordingly, an accurate HR estimate is not always available. Real world driving studies conducted by JSS indicate, that in a typical road mix, using the current prototype, the signal availability is between 70 and 80% of driving time.

APPROACH

The study design enabled a widespread set of potential variables that could be used to derive features for a 2-level-classifier model. Considered variables were narrowed down selecting the most widely used data and their extracted features from ECG and vehicle dynamic data [5]. The objective of this paper is to maximize classifying accuracy with a small set of features to facilitate a low complex and comprehensible classifying system. The extracted features of the ECG signal were mean heart rate (HR) and heart rate variability (RMSSD). Selected features of the vehicle dynamic data were steering wheel reversal rate (SWRR) and standard deviation from center of lane (SDCL).

The recorded signals from the different sensors (reference ECG sensor, vital sensing steering wheel, vital sensing seatbelt) were synchronized using timestamps recorded with each data set. All signals were resampled to 40 Hz to allow synchronous processing of the recorded data. Features were calculated within a time window of 60 seconds which was repeatedly shifted by 5 seconds. The data was subsequently normalized using the minimum value of the first baseline section and the maximum value of the 2-back-task period to eliminate the day-to-day baseline fluctuation and improve the classification performance.

As a first attempt at classifying the workload level, a decision tree was created both for each individual subject and all subjects together using different combinations of the derived features. After analysis, only three of four features turned out to be meaningful and were taken into further consideration. The fourth feature, steering wheel reversal rate, did not show a measurable effect between the different task periods and the baseline.

RESULTS OF THE SIMULATOR STUDY

The verbal answers to the n-back-tasks of each trial were recorded with an automated voice recognition (Microsoft Speech API) and compared to the list of correct answers. Seven trials were left out of the analysis, because the subjects did not manage to complete individual n-back tasks. Overall 143 out of 150 observations were analyzed (3 per trial for each subject).

Mean performance in the easiest task condition (0-back) was nearly perfect and below 1% error rate in both daytime conditions (Figure 5). As the level of cognitive challenge rises the error rate rises also. The mean 1-back error rates in the morning were 3.3%, and, in the evening 2.5%. The mean 2-back error rates in the morning were 10.2%, and, in the evening 11.2%. The observed error rates show that mental workload increased across the task periods independently from daytime. The mean error rates were on a comparable level both in the morning and evening. Wilcoxon tests were carried out in between each task period and show a significant difference in mean values.

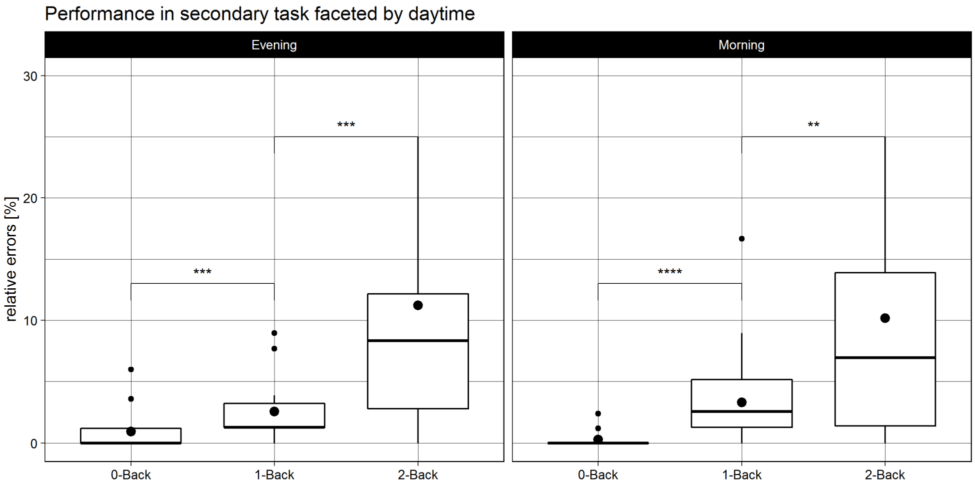

Figure 5. Mean error rates in different n-back-task difficulties sorted by daytime (**: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001).

The subjective rating of mental workload was assessed with NASA-TLX questionnaire, 248 of 250 observations were analyzed (5 per trial for each subject). Two samples had to be omitted due to loss of data. Wilcoxon tests were carried out in between each task period. Both diagrams in Figure 6 show a significant difference in means of each task period. As an exception the 0-back task condition in the evening does not differ significantly from the baseline. Overall the diagrams show that the subjects experience differences in workload between the task periods in both morning and evening.

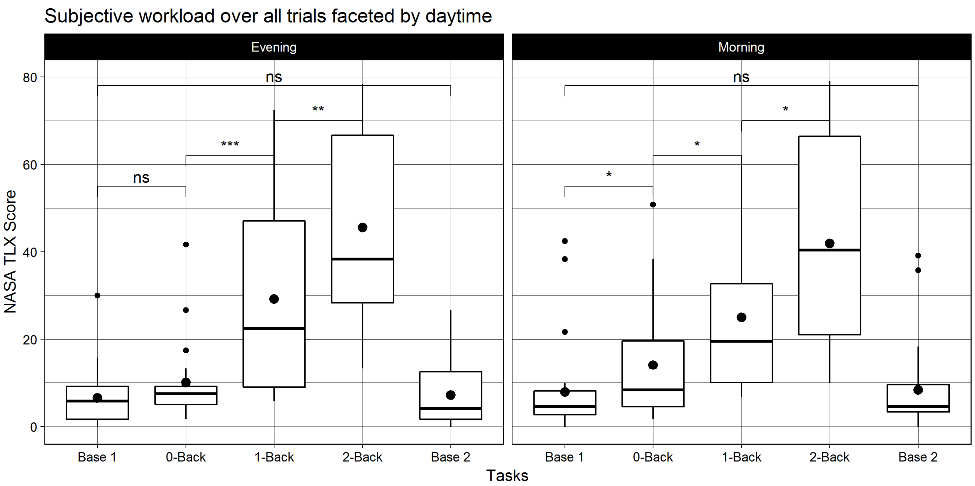

Figure 6. Boxplot of subjective multidimensional workload rating (ns: >0.05, *: p <=0.05, **: p <= 0.01, ***: p <= 0.001).

To identify the cognitive workload using sensors available in vehicles, two features derived from vital data (mean HR, RMSSD) and two features derived from vehicle data (SWRR, SDCL) were analyzed:

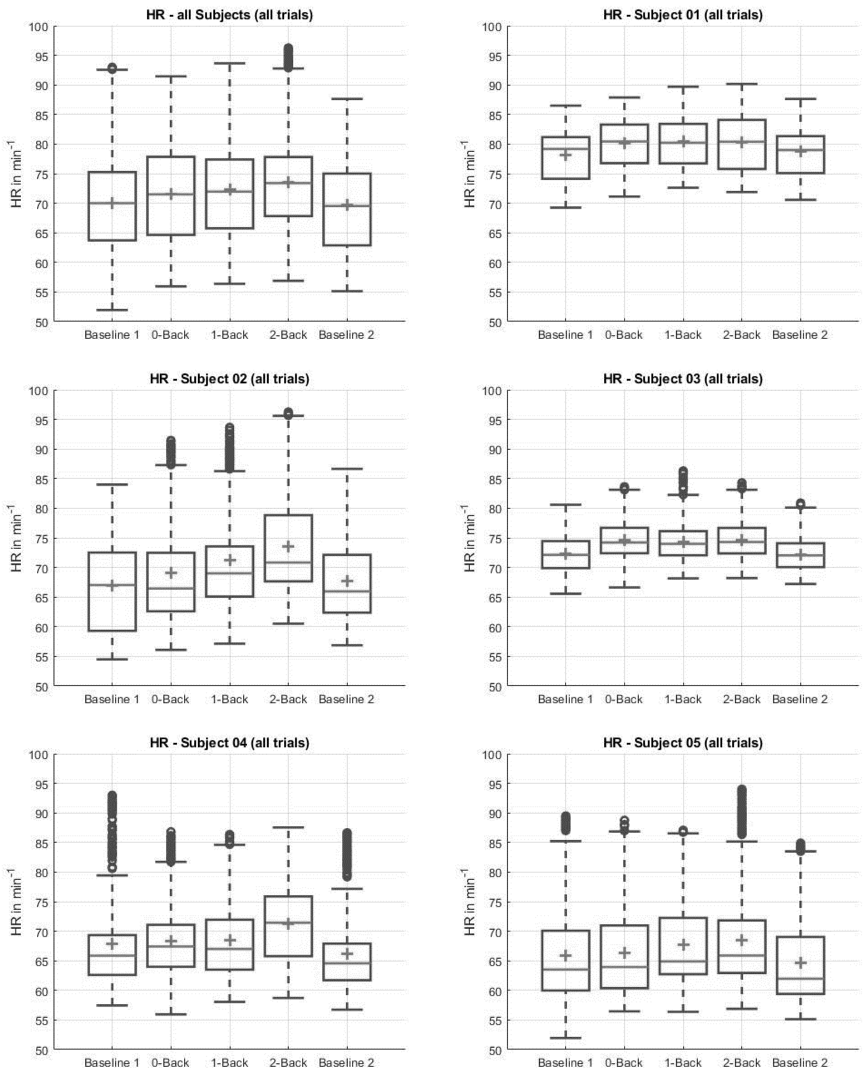

Figure 7. Distribution of HR.

Figure 7 supports the hypothesis that cognitive workload activates the ANS: The mean heart rate of all subjects rises with the increasing difficulty of the secondary task. Over all subjects, this feature increases by 5% from the first baseline to the 2-Back task. However, finding a classifier that identifies workload based on this data is difficult since the distributions are widely spread and overlap each other to a high degree. Two reasons for this observation were identified. First, the typical HR range varies from subject to subject, which distorts the boxplot over all subjects resulting in a wider spread of distributions. Second, a fluctuation of HR baseline from trial to trial was observed with a comparable magnitude to task difficulty variation.

To minimize these two influences for classification, the features were normalized using the minimum value of the first baseline section and the maximum value of the 2-back period of each trial. This normalization reduced the spread of the distributions (see Figure 8) and improved the results of the classification.

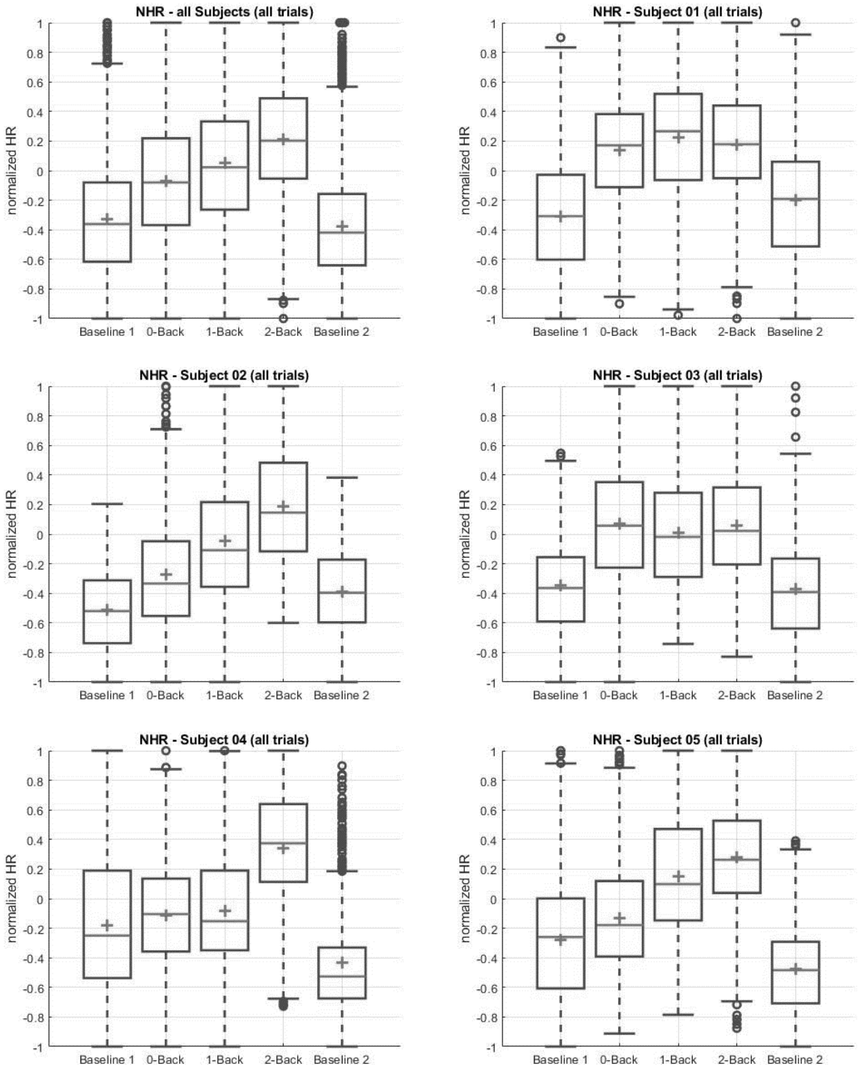

Figure 8: Distribution of normalized HR.

Figure 9 shows the distribution of mean HR over all subjects divided into morning and evening trials. The increase in mean HR is almost identical, but the baseline level of mean HR is, as hypothesized, different during the two times of day. The morning HR is 7.1 bpm higher on average.

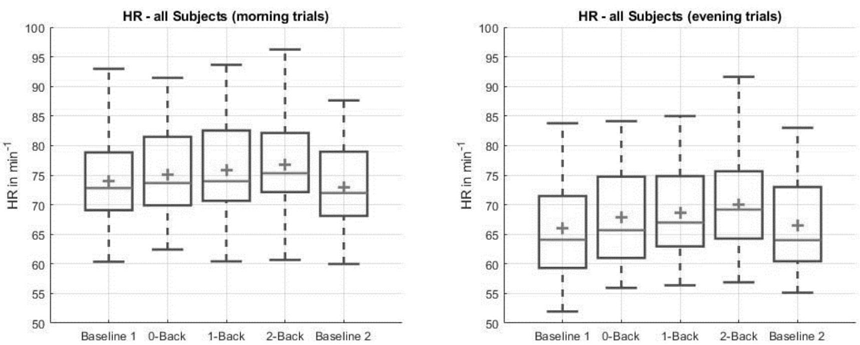

Figure 9. Distribution of HR during morning and evening trials.

The second vital data feature examined was the RMSSD. The boxplots in Figure 10 show that the RMSSD has a similar correlation to the applied task demands as the HR. The mean values of the distributions show a similar, but inverse response to task demand. The RMSSD decreases as the difficulty of the secondary task increases. However, the overlap of the distributions is even higher in relation to the difference of the mean values when compared to the HR. Both the range differences between the subjects and the fluctuations between the trials can also be observed in the RMSSD.

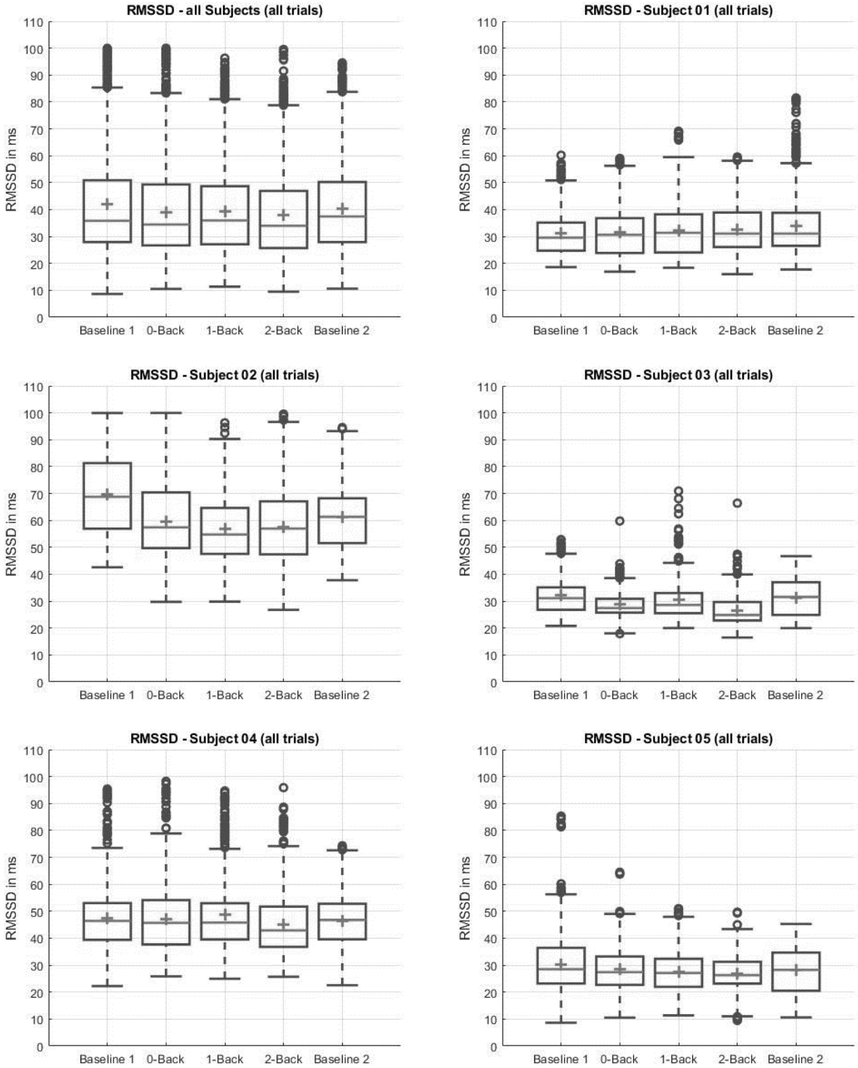

Figure 10: Distribution of RMSSD.

In addition to the features derived from vital data, the features from vehicle dynamics, standard deviation from center of lane (SDCL) and steering wheel reversal rate (SWRR) were analyzed to assess their value as an identifier for periods of high cognitive workload.

In this study, analysis shows the SDCL reacts to cognitive workload like the RMSSD. The mean of the distributions decreases as the task difficulty increases, but the distributions overlap each other as shown in Figure 11. In contrast to the HR and the RMSSD we observed less fluctuations between the individual trials and subjects. Accordingly, applying normalization brings less benefit to the classification.

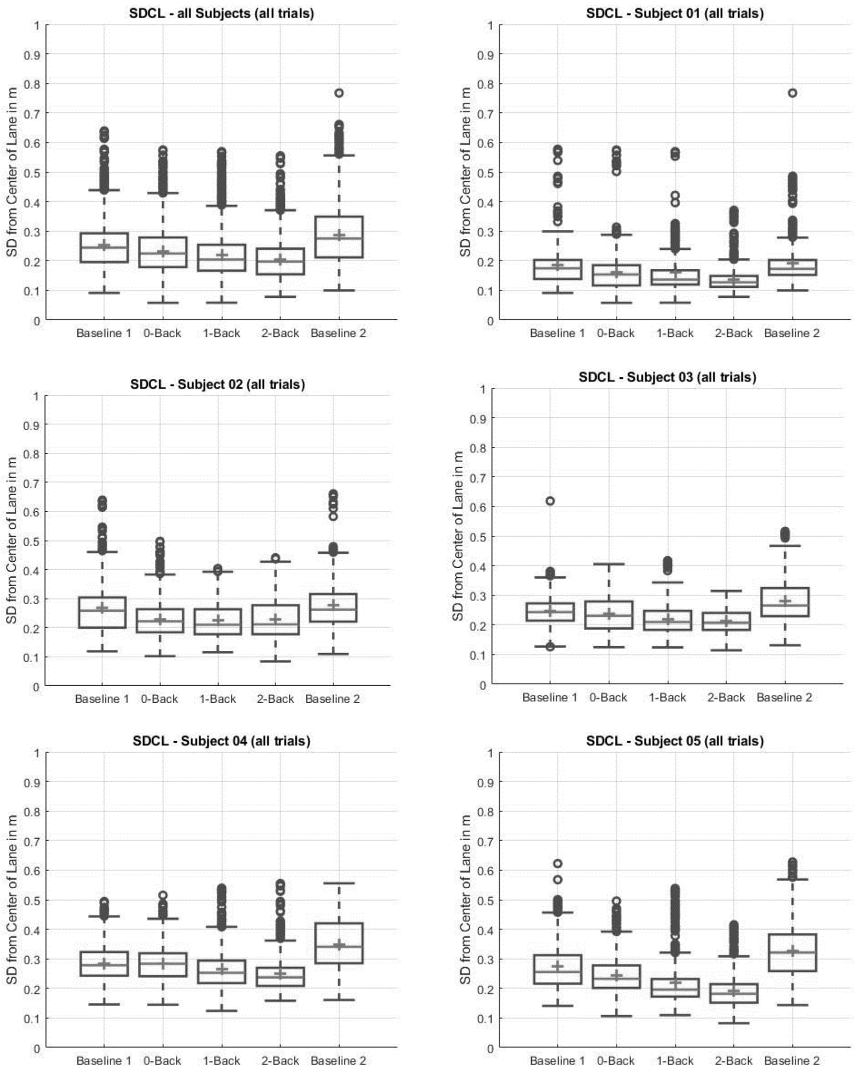

Figure 11: Distribution of SDCL.

Despite being used as a measure for stress recognition in several relevant papers [8, 9] the SWRR in this study, shows no evident connection to the difficulty of the secondary task. Figure 12 shows that the distributions of the first baseline and the 2-back task are almost identical. Even when analyzing the individual trials, no relation of the SWRR to the task difficulty was observed.

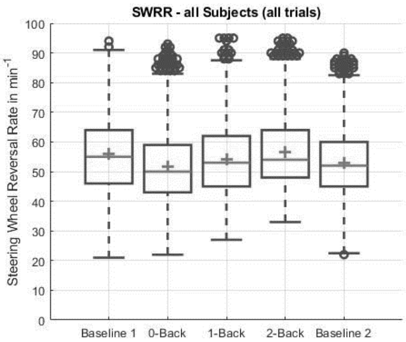

Figure 12: Distribution of SWRR.

To build a workload estimation model, a decision tree was applied to the features HR, RMSSD, and SDCL. This model could correctly separate baseline and 2-back task with an accuracy of 79.2 %. For single subjects, classification accuracies up to 88.4 % were shown.

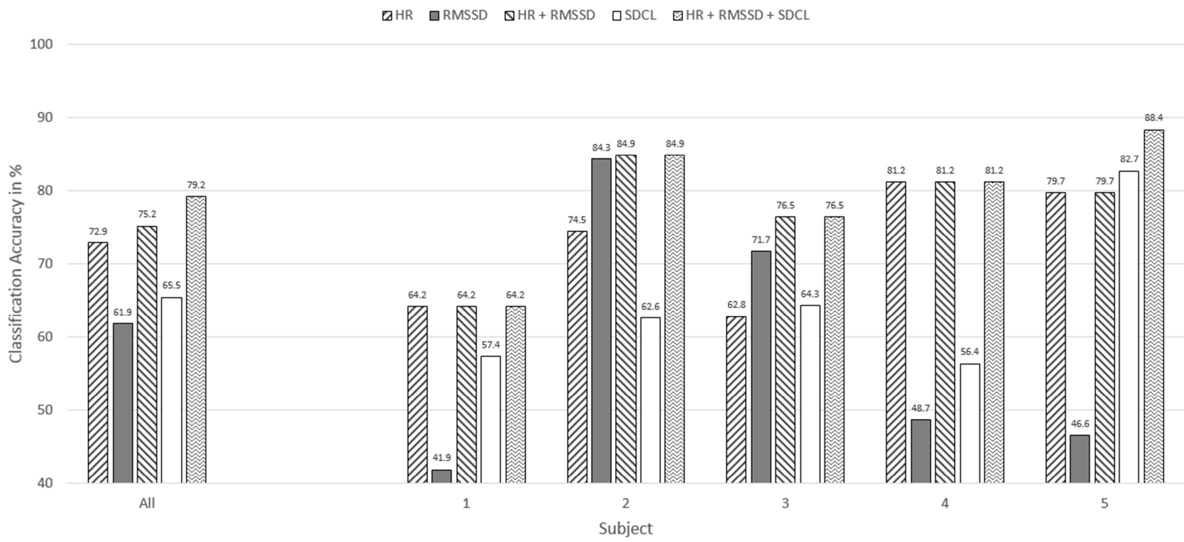

Figure 13: Classification results for different feature combinations.

Analyzing the classification results shown in Figure 13 leads to the following observations. When training a generalized classifier, each added feature improves the accuracy. In general, the HR seems to be the most reliable of the examined features. In most subjects using only the HR is good enough for estimating mental workload while the RMSSD and SDCL improve the accuracy of the prediction in some cases. A decision tree is one method to determine the impact of each feature, but more complex classification models like Support Vector Machines or Artificial Neural Networks have the potential to achieve even better results.

CONCLUSIONS

The results of this study support the proposed hypotheses and lead to several conclusions regarding the approach for implementing an estimation algorithm for mental workload:

Based on this study, the results of the classification using a simple decision tree, classification on individual subjects appears to be the superior approach. The response of the different features to mental workload differs strongly from subject to subject. With such individually optimized classification an accuracy of more than 80% was shown. In order to achieve these results, it was necessary to determine a baseline for each single drive since the features vary strongly depending on the time of the day and the driver's current condition.

For a vehicle application it will be beneficial (and may be necessary) to combine a workload detection system based on physiological data with a driver monitoring camera system. These systems can reliably detect distraction-based workload as described in the NHTSA distraction guidelines [6]. Recent research also indicates that eye gaze features, such as the range of pupil movement can be good indicators for mental workload [13].

While the results from this study were promising, the authors recommend that they be verified with a larger database containing more subjects and more conditions such as different levels of mental workload and different types of workload induction. Considering the strong effect of the time of day on the heart rate shown in this small study, it is also clear, that a deeper investigation of the effect of drowsiness, circadian rhythm, and possibly other contributing factors which could affect mental workload is necessary.

REFERENCES

[1] National Transportation Safety Board. 2017. „Collision Between a Car Operating with Automated Vehicle Control System and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016"Highway Accident Report NTSB/HAR-17/02

[2] EuroNCAP. 2017. "EuroNCAP Roadmap 2025". Revision 4. Brussels

[3] da Silva, F.P. 2014. "Mental Workload, Task Demand and Driving Performance: What Relation? "Procedia - Social and Behavioural Sciences

[4] Hacker, W., 2018. "Definitions of Mental Workload Psychology Essay.". Retrieved from https://www.ukessays.com/essays/psychology/definitions-of-mental-workload-psychology-essay.php?vref=1. Essays UK

[5] Rastgoo, Naim & Nakisa, Bahareh & Rakotonirainy, A & Chandran, V & Tjondronegoro, Dian. 2018. "A Critical Review of Proactive Detection of Driver Stress Levels Based on Multimodal Measurements." ACM Computing Surveys. 51. 88:1:35. 10.1145/3186585.

[6] Garrett, R. 2012. "Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices" (No. NHTSA-2010-0053) (p. 177). Washington, DC, USA: National Highway Traffic Safety Administration (NHTSA)

[7] Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. 2009. "Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers." Transportation Research Record, 2138(1), 6–12. https://doi.org/10.3141/2138-02

[8] Lanatà, Antonio, et al. 2015. "How the autonomic nervous system and driving style change with incremental stressing conditions during simulated driving." IEEE Transactions on Intelligent Transportation Systems 16.3 (2015): 1505-1517.

[9] Bořil, Hynek, Pinar Boyraz, and John HL Hansen. 2012. "Towards multimodal driver's stress detection." Digital signal processing for in-vehicle systems and safety. Springer, New York, NY, 2012. 3-19.

[10] De Waard, D. 2002. Mental workload. In R. Fuller and J. A. Santos (Eds.) Human Factors for Highway Engineers. (pp. 161-175). Netherlands: Pergamon Press

[11] Rodrigues, João & Kaiseler, Mariana & Aguiar, Ana & Cunha, João Paulo & Barros, Joao. 2015. "A Mobile Sensing Approach to Stress Detection and Memory Activation for Public Bus Drivers." IEEE Transactions on Intelligent Transportation Systems.

[12] Fallahi, Majid et al. 2016. "Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study." Applied ergonomics 52 (2016): 95-103.

[13] Fridman, L., Reimer, B., Mehler, B., Freeman W. 2018. "Cognitive Load Estimation in the Wild." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Paper No. 652. http://dx.doi.org/10.1145/3173574.3174226

[14] Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, *3*(3), 203-220.