evaluating drivers’ states in sleepiness countermeasures experiments using physiological and eye data - Hybrid logistic and linear regression model

**Summary:** Objective sleepiness evaluation is essential for the effect analysis of countermeasures for driver sleepiness, such as in-car stimulants. Furthermore, measuring drivers’ sleepiness in simulator studies also becomes important when investigating causes for task-related sleepiness, such as driving on monotonous routes which requires little driver engagement. To evaluate driver sleepiness and the effect of countermeasures we developed a model for predicting sleepiness using both simple logistic and linear regression of heart rate variability, skin conductance and pupil diameter. The algorithm was trained and tested with data from 88 participants in driving simulator studies. A prediction accuracy of 77% was achieved and the model’s sensitivity to thermal stimulation was shown.

# Introduction

In the last decades, driver fatigue theories (Lal & Craig, 2001; May & Baldwin, 2009; van Veen et al., 2014) have been developed that distinguish different types of fatigue. Most theories differentiate task-related (TR) and sleep-related (SR) fatigue. TR fatigue can be due to task overload or underload. The latter one is also referred to as cognitive sleepiness. This differentiation becomes important when considering countermeasures of driver fatigue. The literature reviews of May & Baldwin (2009) and van Veen et al. (2014) propose various countermeasures for different fatigue types. In the case of TR fatigue due to monotony, counter- measures include a variety of in-car stimulation, whereas SR fatigue is intervention-resistant (van Veen et al., 2014). To increase traffic safety and driving comfort, studies have been conducted investigating these countermeasures (Desmond & Matthews, 1997; van Veen, 2016).

When investigating driver sleepiness and the effect of different stimuli in experimental settings, a tool for measuring sleepiness are subjective questionnaires (e.g. Karolinska Sleepiness Scale, KSS). However, questionnaires can only be asked at discrete times, preventing a continuous analysis of sleepiness. Second, listening and answering questionnaires has an awakening effect on the driver which is often undesirable in driver sleepiness research.

A different tool for measuring sleepiness can be classification algorithms. Patel et al. (2011) for example describe a neural network classifying sleepiness with 90% accuracy based on drivers’ ECG (electrocardiogram) data. Friedrichs & Yang (2010) and Hu & Zheng (2009) report accuracies of 83% when differentiating three degrees of sleepiness based on eye features. These algorithms have been developed using data from sleep-deprived drivers, hence these are detecting SR fatigue. The model of Igasaki et al. (2015) uses data of non-sleep deprived drivers. Their logistic regression based on heart rate variability (HRV) measures and respiratory features yields 81% detection accuracy, however it was only generated with data from eight male drivers. None of the above studies attempted to model sensitivity to in-car sleepiness countermeasures.

We propose an algorithm that can be used as a tool for continuously evaluating driver sleepiness in simulator studies with focus on TR sleepiness caused by monotonous driving. The training data was generated by means of a secondary data analysis of three driving simulator experiments investigating in-car countermeasures on the driver’s state. Our leading research questions are:

RQ1: How accurate can the algorithm detect cognitive sleepiness in an unknown driver?

RQ2: Can the algorithm detect changes in sleepiness induced by countermeasures?

To answer these questions, this paper first describes how training data was collected from three simulator studies and the selection of features indicating sleepiness. Second, the algorithm, consisting of a cascaded logistic and linear regression model, is reported with its quality factors. Finally, our research questions are discussed and the applicational limitations are reflected on.

# Method

## Experimental Designs

Data for training and testing the classification of sleepiness was collected in a series of driving simulator experiments. The aim of these pilot experiments was the investigation of countermeasures for critical driver states. This secondary data analysis aims to model sleepiness caused by task underload, and therefore only a part of the drives completed in these studies – which caused task underload – are analyzed. Table 1 shows a description of the monotonous highway drives, information on sample sizes and a descriptions of the treatment.

The first experiment included a 24 minute long monotonous drive which is visualized in Table 1. After 14 minutes and at the end of the drive participants were asked the Stanford Sleepiness Scale (SSS) via microphone. A countermeasure was applied between minutes 20 and 23 that consisted of a combination of orange light from the car ceiling, scent, rhythmic sound and an increased fan intensity of the AC (COMB). The in-car settings were started via remote control by the investigator. The second experiment employed a one-factor within-design. In this study, each subject drove two identical highway routes for 26 minutes. In one of the drives – the order was randomized – cooling at 17°C (COOL) was applied between minute 20 and 26 by the investigator. The sleepiness of the subjects was asked after 6, 16 and 26 minutes via the KSS. The third experiment included an 18 minute long monotonous drive with no intervention (CONT). At the end of the drive sleepiness was asked with the SSS.

All drives were highway drives with very little traffic. Based on the observations of Schmidt et al. (2016) who investigated the possibility to induce cognitive sleepiness by means of traffic scenarios, we conclude that 17 minutes of monotonous driving in a simulator is sufficient to evoke high SSS ratings.The test vehicle was a street-legal car, placed in a static driving simulator with a curved screen providing a 220° view (Figure 1). Two monitors facing the side mirrors placed behind the car and a rear mirror display enabled a rear view. For each experiment 50 subjects were invited. All subjects were asked to maintain their regular sleep schedule. Due to technical problems the usable datasets add up to 122. ECG and skin conductance level (SCL) were measured with medical sensors (g.tech, Austria) with a sampling frequency of 512 Hz. Gaze coordinates and pupil diameter of each eye were recorded at 60 Hz using a remote eye tracker (Tobii, Sweden).

**Table 1. Overview of simulator drives used for algorithm development**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Sample** |  |  | **Sleepiness intervention** |
| **Study I****December 2015** | n=36 (♂ 25, ♀ 11)age 31.3±9.7, (min. 18, max. 57) |  |  | COMB: combination of light, sound, scent and climate in minutes 20-23 |
|  |
| **Study II****February 2016** | n=44 (♂ 31, ♀ 13)age 33.0±11.3, (min. 21, max. 59) |  |  | COOL: climate change in minutes 20-26 |
|  |
| **Study III****August 2016** | n=42 (♂ 33, ♀ 9)age 30.7±8.6, (min. 22, max. 52) |  |  | None, CONT: control condition |
|  |



**Figure 1. Simulator setup**

## Signal Processing

*Predictor variables.* The data was processed and analyzed using Matlab 2013b and the algorithm was developed with weka 3.6.13 (Hall et al., 2009). The different processing steps are visualized in Figure 2. From the ECG recordings the time domain HRV measures SDNN (standard deviation of normal to normal intervals) and RMSSD (root mean square of successive differences) for the period of 3 minutes were extracted. Furthermore, a spectral analyses of a moving 3-minute sequence of interbeat intervals was performed and the frequency domain HRV measures LF (low frequency component, 0.04-0.15 Hz), HF (high frequency component, 0.15-0.4 Hz) and total power were obtained. Pupil diameters of the left and right eye were averaged for each subject. The SCL as its raw value was used instead of its phasic and tonic component because of better classification accuracy.

To increase the classification accuracies, the features were further transformed. Instead of using absolute values, we found that the relative changes of features compared to the first driving minute (third minute respectively for HRV measures) yields better results (equations 1, 2 and 3). Furthermore, the exponents of SDNN, HFrel, total powerrel, SCLrel and diameterrel were adjusted.

|  |  |
| --- | --- |
| $$diameter\_{rel}\left(t\right)=\frac{diameter\left(t\right)-diameter\left(1\right)}{diameter\left(1\right)}, t…time [min]$$ | (1) |
| $$SCL\_{rel}\left(t\right)=\frac{SCL\left(t\right)-SCL\left(1\right)}{SCL\left(1\right)}, HRV\_{rel}(t)=\frac{HRV\left(t\right)-HRV(3)}{HRV(3)}$$ | (2), (3) |

*Labels.* The responses of the subjects in the KSS and SSS were used for labeling. The corresponding predictor variables are the values of the adjusted features of the minute before the subjective rating (e.g. the KSS rating after 16 minutes was matched with the features from the 16th minute). The classes “awake” and “sleepy” were formed by means of the subjective sleepiness rankings in the following way: data observations with KSS-values of 1, 2, 3 and 4 as well as SSS-values of 1, 2 and 3 formed class 1 – “awake”. Data observations with KSS-values of 8 and 9 as well as SSS-values of 6 and 7 formed class 2 – “sleepy”. This way, a total data set of 171 observations from 88 different drivers was generated. The dataset includes 85 observations of awake drivers and 86 observations of tired drivers.



**Figure 2. Signal processing**

## Algorithm

The detection of sleepiness can be handled in two different ways: classification or regression. This is possible because the classes 1 and 2 do not only describe nominal classes (awake and sleepy) but can also serve as numeric values for the degree of sleepiness, allowing for regression approaches. Regression approaches often fail to model the individual differences in prediction problems because the prediction is often approximating the mean of all labels. Therefore we chose a classification approach over regression in the first step to distinguish the separate classes. To improve the sensitivity of the prediction results to external stimuli, a linear regression model with the inputs diameterreland the class values was developed in the second step.

# results

After comparing several classification algorithms we found that the logistic regression classifier performed the best. The logistic regression model was developed using a 10-fold-cross-validation on the 171 observations. The logistic regression function is given by equations (4) and (5) with a coefficient vector (a, b, c, d, e, f, g, h, i) of (8.4∙10-5, 0.81, -0.84, 0.16, -6.5∙10-7, 4.1∙10-3, 4.6, -6.6, 0.45). The classification accuracy is 77.19%, with a ROC area of 0.781 for both classes. Figure 3 shows that a total of 92 observations were classified as awake and 79 as sleepy. The confusion matrix is shown in Table 2.

|  |  |
| --- | --- |
| $$probabilty of class 1 =\frac{1}{1+exp⁡(-x)}$$ | (4) |
| $$x=a∙SDNN^{2}+b∙RMSSD\_{rel}+c∙HF\_{rel}+d∙HF\_{rel}^{2}+e∙LF$$$$+f∙total Power\_{rel}^{-1}+g∙GSR\_{rel}^{3}+h∙diameter\_{rel}^{2}+i$$ | (5) |
| $$probability of class 2=1-probability of class 1$$ | (6) |

Even though the classification of cognitive sleepiness is fair, it is not sufficiently accurate to model the activation of the driver through the countermeasures because the class is either “1-awake” or “2-sleepy”. Hence, any intermediate state, such as slight reductions in sleepiness due to in-car stimulation are not represented by the model.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2. Confusion matrix of simple logistic regression**

|  |  |  |
| --- | --- | --- |
| Classified as → | 1 | 2 |
| 1 | 69 | 16 |
| 2 | 23 | 63 |

 | **Table 3. Confusion matrix of linear regression**

|  |  |  |
| --- | --- | --- |
| Classified as → | 1 | 2 |
| 1 | 70 | 15 |
| 2 | 24 | 62 |

 |
| **Figure 3. Histogram of predicted sleepiness with simple logistic regression** | **Figure 4. Histogram of predicted sleepiness with linear regression**  |

Therefore, we improved our model by cascading a linear regression model after the logistic classifier which reproduces such intermediate states. The classification result serves as input for the linear regression model, along with the relative change in pupil diameter. The pupil diameter was chosen as an input because it delivers best results as it is a very sensitive measure of sympathetic activation and hence replicating slight changes of driver activation. The linear regression model for sleepiness was generated using a 10-fold-cross-validation. The regression function is given by equation (7) with a coefficient vector (j, k, l) of (0.45, -1.5, 0.8). The correlation coefficient of the regression model is r=0.53, p<.001. The classification accuracy of the rounded numeric sleepiness level is still at 77.19%. Table 3 shows the new confusion matrix and Figure 4 the new distribution of predicted sleepiness levels.

|  |  |
| --- | --- |
| $$sleepiness level=j∙class+k∙diameter\_{rel}+l$$ | (7) |

When comparing the distributions of Figure 3 and Figure 4, the effect of cascading the linear regression is visualized. As it is a characteristic of linear regression, it approximates the mean of all observations which constitutes that the two bars of Figure 3 are moved closer towards the mean of 1.5 in Figure 4. This hardly alters the classification accuracy (see Table 2 and Table 3), but allows for the detection of slight changes in sleepiness through external stimuli. If the linear regression would have been performed without the logistic classifier beforehand, the distribution of predictions would peak at 1.5 and therefore confuse sleepy and awake drivers.

To evaluate the sensitivity of the proposed algorithm to changes in cognitive sleepiness due to stimulation, we compared the predicted sleepiness of the two drives “CONT” and “COOL” of the simulator study II. The mean sleepiness of the 44 drivers is visualized in Figure 5 and it can be seen that sleepiness increased over the course of both drives. The t-test results between “CONT” and “COOL” for each driving minute yield statistically significant differences in the 21st (p<0.003), 23rd (p<0.049), 24th (p<0.034), 25th (p<0.023) and 26th (p<0.001) minute, in which cooling was applied. The graph also shows that there is a trend for decreased sleepiness after 6 and 16 minutes, when the KSS was asked which caused an activation of the drivers.



**Figure 5. Mean and standard error of predicted sleepiness with t-test result between conditions**

# Conclusion

Based on the presented results, the research questions can be answered:

A1: The cognitive sleepiness of car simulator drivers can be detected with an accuracy of 77.19% using ECG, SCL and pupil diameter as inputs. The classification accuracy achieved with the proposed algorithm is fair, taking into account that the data was collected from non-sleep-deprived drivers. Better performing algorithms with similar signal input requirements found in the literature were trained with data from sleep-deprived drivers. However, it is questionable whether these algorithms can also predict cognitive sleepiness as accurately.

A2: The algorithm detects changes in sleepiness of the sample in our second study which were induced by thermal stimulation. Furthermore, the predicted values also reflect tendencies of decreased sleepiness when the driver is answering the KSS. Though, when applying the model for the evaluation of sleepiness countermeasures, we recommend having a large sample to overcome the imprecision of the prediction.

The proposed algorithm will perform well only if the light settings are kept constant during the experiment. The reason is that changes in the illuminance cause pupillary restrictions unrelated to an increase in sleepiness and would therefore skew both classification and regression results. For this reason, the algorithm is not suited to evaluate light as an intervening stimulant for sleepy drivers. Since we trained the algorithm with the extreme KSS- and SSS-values, the reported accuracy can only be guaranteed for awake or sleepy ratings. For intermediate values (5, 6, 7 for KSS and 4, 5 for SSS) the logistic regression probabilities are close to an equal likelihood for both classes, increasing the risk of misclassifications.

Taking both the accuracy and sensitivity of the regression model into account, the algorithm is a suitable tool for continuously evaluating TR sleepiness due to monotony in driving simulator studies. The model can also serve as an objective measure for the effectiveness of countermeasures, such as in-car stimulants. In further studies, we would like to evaluate the performance of the sleepiness prediction for repeated and even cooler thermal stimuli. Moreover, the algorithm should also be tested for different causes of TR sleepiness, such as automated driving.

# References

Desmond, P. A., & Matthews, G. (1997). Implications of task-induced fatigue effects for in-vehicle countermeasures to driver fatigue. *Accident Analysis & Prevention, 29*(4), 515-523.

Friedrichs, F., & Yang, B. (2010). Camera-based drowsiness reference for driver state classification under real driving conditions. *Proceedings of the IEEE Intelligent Vehicles Symposium*, San Diego, California, 101-106.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter, 11*(1), 10-18.

Hu, S., & Zheng, G. (2009). Driver drowsiness detection with eyelid related parameters by Support Vector Machine. *Expert Systems with Applications, 36*(4), 7651-7658.

Igasaki, T., Nagasawa, K., Murayama, N., & Hu, Z. (2015). Drowsiness estimation under driving environment by heart rate variability and/or breathing rate variability with logistic regression analysis. *Proceedings of the 8th IEEE International Conference on Biomedical Engineering and Informatics*, Shenyang, China, 189-193.

Lal, S. K., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological psychology, 55*(3), 173-194.

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour, 12*(3), 218-224.

Patel, M., Lal, S. K. L., Kavanagh, D., & Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert systems with Applications, 38*(6), 7235-7242.

Schmidt, E., Decke, R., & Rasshofer, R. (2016). Correlation between subjective driver state measures and psychophysiological and vehicular data in simulated driving. *Proceedings of the IEEE Intelligent Vehicles Symposium*, Gothenburg, Sweden, 1380-1385.

Van Veen, S., Vink, P., Franz, M., & Wagner, P.-O. (2014). Enhancing the vigilance of car drivers: a review on fatigue caused by the driving task and possible countermeasures. *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics*, Kraków, Poland, 516-525.

Van Veen, S. (2016). Driver Vitalization - Investigating sensory stimulation to achieve a positive driving experience. PhD thesis, TU Delft.