Statistical sensor fusion of ECG data using automotive-grade sensors

A. Koenig, T. Rehg, and R. Rasshofer

BMW Group Research and Technology, Munich, Germany

Correspondence to: A. Koenig (alexander.ak.koenig@bmw.de)

Received: 22 December 2014 – Revised: 5 March 2015 – Accepted: 16 March 2015 – Published: 3 November 2015

Abstract. Driver states such as fatigue, stress, aggression, distraction or even medical emergencies continue to be yield to severe mistakes in driving and promote accidents. A pathway towards improving driver state assessment can be found in psycho-physiological measures to directly quantify the driver's state from physiological recordings. Although heart rate is a well-established physiological variable that reflects cognitive stress, obtaining heart rate contactless and reliably is a challenging task in an automotive environment. Our aim was to investigate, how sensory fusion of two automotive grade sensors would influence the accuracy of automatic classification of cognitive stress levels. We induced cognitive stress in subjects and estimated levels from their heart rate signals, acquired from automotive ready ECG sensors. Using signal quality indices and Kalman filters, we were able to decrease Root Mean Squared Error (RMSE) of heart rate recordings by 10 beats per minute. We then trained a neural network to classify the cognitive workload state of subjects from heart rate and compared classification performance for ground truth, the individual sensors and the fused heart rate signal. We obtained an increase of 5 % higher correct classification by fusing signals as compared to individual sensors, staying only 4 % below the maximally possible classification accuracy from ground truth. These results are a first step towards real world applications of psycho-physiological measurements in vehicle settings. Future implementations of driver state modeling will be able to draw from a larger pool of data sources, such as additional physiological values or vehicle related data, which can be expected to drive classification to significantly higher values.

1 Introduction

Over the last years, the automotive industry has seen a surge in driver assistance systems (DAS) to increase security and comfort of the driver. Empowered by an increasing number of sensors such as radar, lidar, ultrasound or video based systems, the vehicle can estimate a model of the current state of the environment, its objects and properties, such as speed or orientation. Meanwhile, the DAS know very little about the state of the driver. Information on the current cognitive, emotional or medical state of the driver is, however, of high importance: according to a study by the German Automobile Club ADAC, tired drivers are responsible for one in four traffic fatalities in Germany (ADAC, 2014). In addition to fatigue, driver states such as stress, aggression, distraction or even medical emergencies continue to be yield to severe mistakes in driving and promote accidents. The vehicle should therefore know whether or not the driver is tired, aggressive, stressed or distracted to be able to adapt its DAS accordingly. Commercially available systems in vehicles currently focus on fatigue estimations and rely on measures of steering wheel interaction and break/acceleration commands of the driver to detect this single driver state. While fatigue estimation is already an important step towards increased safety, these systems must be improved in two ways: faster detection of driver states and a larger number of reliably detected driver states.

A pathway towards improving driver state assessment can be found in psycho-physiological measures to directly quantify the driver's state from physiological recordings, thereby removing the proxy of deducting the driver's state from steering wheel dynamics and pedal interactions (Koenig et al., 2014). One key measure of the psycho-physiological state of a driver is the Electrocardiogram (ECG) and variables derived from it, such as heart rate or heart rate variability. Heart rate is a particularly well established physiological variable

that reflects cognitive stress of subjects (Mehler et al., 2012; Mulderet al., 2000). One problem of previous work on automatically classifying the cognitive stress level of a driver from heart rate information was that recordings were typically done using medical grade, adhesive wet electrodes to record ECG.

For security and comfort reasons, any estimation of a driver’s state can only happen through contactless sensors that seamlessly integrate into a vehicle. Obtaining ECG data contactless and reliably to deduct heart rate is a challenging task in an automotive environment. Galvanic and capacitive sensors in the steering wheel, seat and seatbelt can record ECG. Galvanic sensors, built into a steering wheel for example, have the limitation that the driver has to have both hands on the steering wheel to obtain a differential recording between left and right hand from which the ECG can be extracted. Capacitive sensors, located in the seat or the seatbelt, have limited signal availability when electric currents generated through friction motion pollute the signal. These currents induce local electric fields stronger than the ECG signal and therefore render a useful signal recording impossible whenever the subject moves. In addition, thick clothing can prevent sufficient capacitive coupling between human and sensor. So, while galvanic and capacitive sensors are available automotive-ready, they have limited reliability due to movement artifacts.

Sensor fusion of different sources of ECG derived signals promises higher signal availability and therefore improved driver state assessment. Our aim was to investigate, how the use of automotive grade ECG sensors would influence the accuracy of automatic classification of cognitive stress levels in drivers, and how sensory fusion algorithms could be employed to improve this accuracy. Previous research, mostly performed in the medical field, has focused on fusing existing signals of different kinds, such as blood pressure, ECG and oxygen saturation (SpO₂) to obtain more robust estimates of heart rate (Feldman et al., 1997). Only Li and colleagues looked at fusing recordings of two medical-grade ECG signal sources to obtain higher reliability of heart rate estimates (Li et al., 2008). However, so far no one has investigated how the fusion could improve heart rate recordings in an automotive setting and how classification of driver states might benefit from such sensor fusion.

In this publication, we report our experiments on inducing cognitive stress in subjects and estimating stress levels from their heart rate signals, acquired from automotive ready ECG sensors. We recorded ECG data of five subjects using automotive-grade sensors in the seat and steering wheel during a cognitively challenging task and employed statistical sensor fusion algorithms, namely Kalman filters, to increase reliability of ECG estimates. We trained a neural network to classify the cognitive workload state of subjects from heart rate and compared classification performance for ground truth, the individual sensors and the fused heart rate signal.

Table 1. The cognitive task used to induce cognitive stress levels in subjects. The stimulus (instruction by the experimenter) and the required answer of the subject are shown in the first and second line, respectively.

Stimulus by experimenter	4	8	2	6	0	8	5	2	5
2-back response	–	–	4	8	2	6	0	8	5

2 Methods

2.1 Experimental setup

We aimed at quantifying how well currently available automotive sensors are suited to automatically classify cognitive stress from physiological data. We chose to focus on heart rate as a physiological recording, as heart rate was shown to very well reflect the cognitive stress level of drivers (Mehler et al., 2012). Classifying cognitive stress levels from medical-grade heart rate recordings is well established in literature (Mehler et al., 2012; Koenig et al., 2011). Using only automotive-grade sensors, we wanted to quantify how strongly sensor fusion algorithms could influence the classification accuracy. We hypothesized that we should be able to obtain higher availability of heart rate readings by fusing several automotive-grade sensors, thereby improving classification performance of cognitive stress.

We induced cognitive stress through a well-established memory task, the two-back task (Table 1), which requires subjects to verbally repeat a sequence of numbers between 0 and 9 at a delay of two numbers (see). The experiment consisted of only two conditions of each two minutes length: a baseline condition without cognitive stress and one with cognitive stress.

Heart rate data was computed from ECG data, recorded from five healthy subjects (2f and 3m, 23 years ±2 years) at BMW Group Research and Technology, Munich, Germany, using standard beat detection algorithms. Automotive grade sensor systems were embedded into a driver seat and a steering wheel. To create realistic conditions, subjects were instructed to move in the seat and take their hands on and off the steering wheel as they felt comfortable. Ground truth ECG data was obtained from a medical grade three point leads attached through wet electrodes directly to the subjects’ chest using a Gtec g.USBamp system (Gtec, Graz, Austria) in combination with recording software written in Matlab Simulink 2010 (The Mathworks, Natick, MA, USA) at 512 samples per second. The seat sensor consists of capacitive electrodes in the sitting area and backrest that measures the? ECG as a potential difference between electrodes. The steering wheel sensor consists of a set of Plessey sensors (Plessey Semiconductors, Plymouth, UK) that galvanically record the potential difference between left and right hand. Both auto-

Figure 1. Smoothing and filtering system for the ECG data obtained from seat and steering wheel sensors.

motive sensors sampled data at 1kHz. While the ground truth data is very robust towards noise induced through movement or friction currents, the seat sensor loses signal when it loses direct contact with the driver, for example when the driver moves to adjust his/her sitting position or looks at the side mirrors. The steering wheel can only sense ECG when the driver has both hands on the wheel. Heart rate was extracted from individual ECG signals as $HR = 60 s / RR$ in beats per minute using a standard peak detection algorithm, where RR was the time in between two consecutive R waves, the most prominent peak in the ECG signal.

2.2 Data filtering and fusion

Steering wheel and seat sensors experienced data loss due to loss of contact to the human. To filter the individual HR signals, we followed the approach of Li2008 and used Kalman filters in combination with a Signal Quality Index (SQI), which adapted the Kalman Gain at runtime depending on the current signal quality. We then used the Kalman filtered signals of both, steering wheel and seat, and selected the signal that had the higher Kalman Filter confidence (see Fig. 1). The Kalman Filter used $A = 1$ as its state space matrix with no input B . C was set to 1, D to 0. As heart rate fluctuates by several beats per minute over time (referred to as Heart Rate Variability), the only driver to our system was the process noise, which we estimated beforehand to be 8.6 from the variance of a standard heart rate recording.

The SQI was computed as a combination of three different measures: kurtosis of ECG, running variance of heart rate and absolute values of heart rate. Kurtosis (kSQI) of the ECG signal was calculated using a moving window with a window size of 0.5 s which was found to be optimal for both seat and wheel signals. kSQI was computed as

$$kSQI = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 \quad (1)$$

where x is the heart rate data, $\bar{x}$ is the mean of x and s is the standard deviation of x . If kurtosis was found to be above a threshold of 80, the signal was quantified as too noisy. Variance of heart rate was used as a second indicator of signal quality (varSQI) and computed from a 4 s window, symmetrically around the current data point. As the window was placed symmetrically around the current data point, we introduced a delay of 2 s during computation of SQI.

Usually heart rate variance, the fluctuation of heart rate at rest, lies around ± 2.5 bpm (Koenig et al., 2010). A variance of more than 40 was found to indicate that the original ECG signal had been too noisy. The third indicator for signal quality was the absolute value of heart rate (aSQI): heart rates above 200 bpm or below 30 bpm are physiologically not possible. Only if both, the absolute value and either variance or kurtosis SQI indicate good signal quality the measurement could be trusted.

$$kSQI = \begin{cases} 0 & \text{if } K > 80 \\ 1 & \text{if } K \leq 80 \end{cases} \quad (2)$$

$$varSQI = \begin{cases} 0 & \text{if } VAR > 40 \\ 1 & \text{if } VAR \leq 40 \end{cases} \quad (3)$$

$$aSQI = \begin{cases} 0 & \text{if } 30 < A < 200 \\ -5 & \text{otherwise} \end{cases} \quad (4)$$

the SQI was then computed as

$$SQI = kSQI + varSQI + aSQI \quad (5)$$

According to the SQI, the Kalman filters R matrix, the covariance matrix of the measurements, was adjusted, which directly influenced the Kalman gain. If the $SQI > 0$ the value of R was set to 0.5. If it was below zero, R was set to e^5 . The filter always trusts the signal with higher SQI and uses its inner state transition matrix for the estimation if both measurements are noisy. The computations of SQI delayed the signal by 2 s, which was found to be tolerable.

Figure 2. Exemplary subject, 60 s of data. The top panel shows the comparison between the ground truth heart rate data and the Kalman filter fused data, based upon the individual sensory data from seat and steering wheel. When both sensors fail to record reliable data, the Kalman filter switches to its internal model. The middle panel shows the comparison between the Kalman fused data and the individual sensory data. The bottom panel shows the comparison between ground truth and the individual sensory data.

2.3 Quantitative data analysis

To quantify improvement of our approach in terms of signal availability, we compared ground truth heart rate to the single sensor heart rate readings of steering wheel and seat and to the fused signal via computation of root mean square error (RMSE). In addition to quantifying the exact average improvement through RMSE, we wanted to assess if our approach could improve future efforts in driver state modeling. Our previous results indicated that fluctuations of ± 2.5 beats per minute are within the normal boundaries of heart rate variability (Koenig et al., 2011a) during constant conditions of cognitive stress. We therefore computed the percent availability as the time during which the sensory signal deviated by less than 2.5 beats per minutes from the ground truth, assuming that any deviation smaller than 2.5 beats per minute from ground truth would not result in miss-classification of cognitive stress levels derived from changes in heart rate.

Table 2. Average performance of individual sensors and improvements of heart rate estimations through data fusion as measured by percent deviation from ground truth and Root Mean Square Difference (RMSD) between sensory signal and ground truth.

	Steering wheel alone	Seat alone	Fused data
Percent deviation less than ± 2.5 beats per minute from ground truth	37 ± 18	51 ± 4	61 ± 9
RMSE (bpm)	20 ± 6	14 ± 1	5 ± 1

Table 3. Classification results of cognitive stress from heart rate.

	Ground truth	Steering wheel alone	Seat alone	Fused data
Percent correctly classified	77 ± 12	68 ± 4	61 ± 10	73 ± 11

2.4 Classifying cognitive stress

Heart rate was used as a measure for cognitive stress. We trained a neural network (Neural Network Toolbox, Matlab, The Mathworks) with heart rate estimates as input and performed leave one out classification for all five subjects, i.e. we trained the classifier on all but the i th subject data and classified the data of subject i with it. The neural network used a hidden layer with 20 neurons and a 70, 15 and 15 % split of the training data for training, validation and test respectively. These parameters were identified empirically. Four different sets of input data were provided to train the net:

- ground truth to evaluate the maximally possible correct classification with perfect, unpolluted data
- only heart rate data extracted from the seat sensor
- only heart rate data extracted from the steering wheel sensor
- fused heart rate data obtained from both, seat and steering wheel sensor, as described above

Leave one out classification then allowed computing the percent correctly classified.

3 Results

The fusion algorithm allowed dynamic switching between sensor sources. As seen in Fig. 2 the Kalman Filter used its internal model of $A = 1$ when both sensory signals were unavailable. As soon as one of the sensors was available again, the algorithms switched immediately back to trust its sensor readings.

Figure 3. Aggregate data of all five subjects. The top panel shows mean and standard error of percent of all data that was found to deviate by less than ± 2.5 beats per minute from ground truth for the steering wheel sensor alone, the seat sensor alone and the fused data. The bottom panel shows mean and standard error of the root mean square error (RMSE) between the ground truth and the respective sensors.

Through our sensor fusion concept, we were able to improve availability of heart rate information during our tests from 37 % for the steering wheel sensor and 51 % for the seat sensor by 10 to 61 % total signal availability (as compared to 100 % availability of ground truth data). The availability of the steering wheel sensor between subjects varied greatly by ± 18 %, depending on whether or not the subjects were keeping their hands on the wheel (Table 2). Signal fusion decreased RMSD by 15 and 10 bpm respectively from steering wheel and seat alone.

The classifier used heart rate data as input to classify whether or not the subject was experiencing cognitive stress or not. It has to be noted that 50 % correct classification would denote chance, as we only differentiated between two classes. The best possible classification, given ground truth data, was found to be 77 %. Using only heart rate from seat data, classification dropped to 68 %, a decrease by 9 % (Table 3). Steering wheel data alone only provided correct classification in 60 % of cases. By fusing the signals, we could reach a total of 73 %, only 4 % lower than with ground truth.

Figure 4. Results of neural network based classification of cognitive stress levels of subjects. Ground truth allowed 77 % correct classification, which was almost reached by fused data. ECG steering wheel and ECG seat sensor proved to be worse an average of 7 and 10 %, respectively.

4 Discussion

We investigated the effects of heart rate fusion algorithms on classification of driver states using automotive-grade sensors. The results presented in this paper are a first step towards real world applications of psycho-physiological measurements in vehicle settings. An ever increasing number of driver assistance systems need to understand the current cognitive, emotional or vigilance state of the driver to adapt their support functionality, for example, by adapting warning thresholds.

Previous experiments on automatic classification of subjects cognitive stress levels featured a larger array of sensory input, such as additional physiological signals and task performance data (Koenig et al., 2011a; Mehler et al., 2012; Mulder et al., 2000). Additional physiological values often recorded include Galvanic Skin Response (GSR), a measure for stress and arousal (Boucsein, 2005; Dawson et al., 2007), a measure for cognitive workload or breathing frequency, a measure for stress and mental effort (Carroll et al., 1986; Suess, 1980). In previous experiments, when using only physiological recordings of health subjects, we were only able to classify cognitive workload of subjects at 38 % correct classification for a 4 class problem, where chance level was 25 % (Koenig et al., 2011b). With the additional input of task performance indicators, the correct classification of this 4-class problem rose to 84 %. In this light, our classification results of 73 % correct classification of cognitive stress levels looks very promising, given that ground truth

only provided 77 % correct classification. In comparison, the best single sensor was at 68 %, which corresponds to a 5 % increase in correct classification through the use of filtering software.

Future implementations of driver state modeling will be able to draw from a larger pool of data sources, such as additional physiological values or vehicle related data. Already now, fatigue is quantified from steering wheel and pedal interactions alone. While vehicle based driver state classification has the disadvantage of requiring long calibration times during which the vehicle learns its internal model, this data is constantly available during driving. Combining such available data with data from future sensory systems such ECG sensors, eye trackers and driver cameras can be expected to drive ground truth classification to significantly higher values.

It would not have to be necessary to employ Kalman Filters for this problem, as simple if-then rules could have performed in a similar way. However, one major advantage of our approach is the possibility to extend the system by additional sources of ECG or heart rate recordings. Possible are, for example, heart rate estimates extracted from camera images (Wu et al., 2012), which could be recorded from a driver camera system or additional capacitive sensors in the seat belt. These signals could then be fed as additional input sensory data into the Kalman filter.

Future fusion algorithms will need to not only fuse heart rate signals computed from noisy ECG sources, but perform fusion on the ECG level directly. Probability based approaches, such as particle filters, will be used to estimate RR intervals, which would result in a binary representation of the CG signal. From an estimated RR signal, additional parameters such as Heart Rate Variability which was shown to correlate with cognitive load and emotional stress levels (Malik and Camm, 1990).

Apart from only quantifying drivers' stress levels or tiredness and reacting to it, future developments call for manipulation of the driver state. Active modification, as for example through light, sound, scent or climate control, will allow the vehicle to de-stress the driver or wake him or her up. The basis for these applications will be a reliable driver state monitoring system.

Acknowledgements. The authors thank their colleague Justus Jordan for his support.

Edited by: M. Chandra

Reviewed by: S. Baumgartner and one anonymous referee

References

- ADAC e.V.: Müdigkeit im Straßenverkehr – unterschätzt, verkannt, tödlich, Artikelnummer 2831141, available at: www.adac.de/infotechat/ratgeber-verkehr, 2014.
- Boucsein, W.: Electrodermal measurement, in: Handbook of human factors and ergonomics methods, edited by: Stanton, N., Hedge, A., Brookhuis, K., Salas, E., and Hendrick, H., London, CRC Press, 2005, 18-11–18-18, 2005.
- Carroll, D., Turner, J. R., and Prasad, R.: The effects of level of difficulty of mental arithmetic challenge on heart rate and oxygen-consumption, *Internat. J. Psychophysiol.*, 4, 167–173, 1986.
- Dawson, M. E., Schell, A. M., and Filion, D. L.: Handbook of Psychophysiology, 3rd Edn., 2007.
- Feldman, J. M., Ebrahim, M. H., and Bar-Kana, I.: Robust Sensor Fusion Improves Heart Rate Estimation: Clinical Evaluation, *J. Clin. Monit.*, 13, 379–384, 1997.
- Koenig, A., Omlin, X., Zimmerli, L., Sapa, M., Krewer, C., Bolliger, M., Mueller, F., and Riener, R.: Psychological state estimation from physiological recordings during robot-assisted gait rehabilitation, *J. Rehabil. Res. Dev.*, 48, 367–385, 2011a.
- Koenig, A., Novak, D., Pulfer, M., Omlin, X., Perreault, E., Zimmerli, L., Miheljić, M., and Riener, R.: Real-time control of cognitive load in neurological patients during robot-assisted gait training, *Trans. Neural. System. Rehabil. Engin.*, 19, 453–464, 2011b.
- Koenig, A., Decke, R., and Raschofer, R. H.: Fahrerzustandsmodellierung, *Elektronik automotive*, 33–37, 2014.
- Li, Q., Mark, R. G., and Clifford, G. D.: Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter, *Physiol. Meas.*, 29, 15–32, 2008.
- Malik, M. and Camm, A. J.: Heart rate variability, *Clin. Cardiol.*, 13, 570–576, 1990.
- Mehler, B., Reimer, B., and Coughlin, J.: Demand From a Working Memory Task, An On-Road Study Across Three Age Groups Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand from a Working Memory Task, An On-road Study Across Three Age Groups, *Human Fact.*, 54, 396–412, 2012.
- Mulder, G., Mulder, L., Veldman, J., and van Roo, A.: A psychophysiological approach to working conditions, in: Engineering psychophysiology: Issues and applications, edited by: Backs, R. W. and Boucsein, W. M., Lawrence Erlbaum Associates, 139–159, 2000.
- Suess, W. M., Alexander, A. B., Smith, D. D., Sweeney, H. W., and Marion, R. J.: The effects of psychological stress on respiration – a preliminary-study of anxiety and hyperventilation *Psychophysiology*, 17, 535–540, 1980.
- Wu, H., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. T.: Eulerian Video Magnification for Revealing Subtle Changes in the World, *ACM Transact. Graph.*, 31, 2012.