Warwick-JLR Driver Monitoring Dataset (DMD): A public Dataset for Driver Monitoring Research

Phillip Taylor*
Dept. Computer Science
University of Warwick
Coventry, CV4 7AL
United Kingdom
phil@dcs.warwick.ac.uk

Nathan Griffiths
Dept. Computer Science
University of Warwick
Coventry, CV4 7AL
United Kingdom

Abhir Bhalerao
Dept. Computer Science
University of Warwick
Coventry, CV4 7AL
United Kingdom

Derrick Watson
Dept. Psychology
University of Warwick
Coventry, CV4 7AL
United Kingdom

Xu Zhou
Jaguar Land Rover
West Midlands
United Kingdom

Thomas Popham
Jaguar Land Rover
West Midlands
United Kingdom

ABSTRACT
Driving is a safety critical task that requires the full attention of the driver. Despite this, there are many distractions throughout a vehicle that can impose extra workload on the driver, diverting attention from the primary task of driving safely. If a vehicle is aware that the driver is currently under high workload, the vehicle functionality can be changed in order to minimize any further demand. Traditionally, workload measurements have been performed using intrusive means such as physiological sensors. We propose to monitor workload online using readily available and robust sensors accessible via the vehicle's Controller Area Network (CAN). The purpose of this paper is to outline a protocol to collect driver monitoring data and to announce the publication of a database for driver monitoring research. We propose five ground truths, namely, timings, Heart Rate (HR), Heart Rate Variability (HRV), Skin Conductance Level (SCL), and frequency of Electrodermal Responses (EDR). The dataset will be released for public use in both driver monitoring and data mining research.

Keywords
Driver monitoring, Data collection, EDA, ECG, CAN-bus

1. INTRODUCTION
Driving is a safety critical task that requires the full attention of the driver. Despite this, modern vehicles have many devices with functions that are not directly related to driving. These devices, such as radio, mobile phones and even internet devices, divert cognitive and physical attention from the primary task of driving safely. In addition to these distractions, the driver may also be under high workload for other reasons, such as dealing with an incident on the road or holding a conversation in the vehicle. One possible solution to this distraction problem is to limit the functionality of in-car devices if the driver appears to be overloaded. This can take the form, for example, of withholding an incoming phone call or holding back a non-urgent piece of information about traffic or the vehicle status.

It is possible to infer the level of driver workload from observations of the vehicle and the driver. Based on these inferences, the vehicle can determine whether or not to present the driver with new information that might unnecessarily add to their workload. Traditionally, such systems have monitored physiological signals such as heart rate or skin conductance [3, 13, 7]. However, such approaches are not practical for everyday use, as drivers cannot be expected to attach electrodes to themselves before driving. Other systems have used image processing for computing the driver's head position or eye parameters from driver facing cameras, but these are expensive, and unreliable in poor light conditions [9].

We therefore use non-intrusive, inexpensive and robust signals, which are already present in vehicles and are accessible by the Controller Area Network (CAN) [4]. The CAN is a central bus to which all devices in the vehicle connect and communicate by a broadcast protocol. This allows sensors and actuators to be easily added to the vehicle, enabling the reception and processing of telemetric data from all modules of the car. This bus and protocol also enables the recording of these signals, allowing us to perform offline data analysis and mining. In mining this data, we aim to build a system that can recognise when a driver is overloaded and then act accordingly. Our initial work has shown that features extracted from the CAN are able to support machine learning models for predicting the cognitive load of a driver [11] or

*Contact author for this paper.

AutomotiveUI'13 October 27-30, Eindhoven, The Netherlands.
Adjunct Proceedings.

the state of a vehicle, such as the current road type [10].

This paper proposes a procedure for acquiring a dataset for this driver monitoring problem, in the form of a supervised classification task. The ground truths are taken from both experiment timings and physiological measures, namely Electrocardiography (ECG) and Electrodermal Activity (EDA). The remainder of this paper is structured as follows. In Section 2 we outline the experimental protocol that is be used to distract the driver during data collection. Section 3 describes the CAN-bus data in more detail and states how the ground truth is be achieved. Finally, in Section 4 we give details of the format of the data and its release and briefly discuss its potential impact on driver monitoring research.

2. EXPERIMENTAL PROTOCOL

The experimental protocol we use is based on that performed by Reimer *et al.* [9] and Mehler *et al.* [7], and is outlined in Table 2. In their work, changes in physiology and driving style are observed while the driver is performing the N-back test as a secondary task to driving. The main difference in our protocol is that we perform it on a test track and the ECG electrodes are on the chest rather than the lower neck. Also, we use gel EDA electrodes with adhesive pads, as we have found these are more stable and, in our experience, produce a cleaner signal.

Our implementation of the protocol runs as follows. First, when the participant arrives, electrodes are attached for both the ECG and EDA measurements. After this, the participant is taken to the vehicle and seated in the driving position. Once the seat, steering wheel, and mirrors are adjusted as appropriate, data recording is commenced. The protocol then continues with checking that the sensors are providing a clean and reliable signal, followed by practice runs of the N-back tests (stages 1 and 2).

The N-back test requires the participant to repeat digits provided to them in a list with a delay. Here it is operated with three forms of increasing difficulty, with delays of 0, 1 and 2 and referred to as the 0-, 1- and 2-back tests respectively. These three difficulty levels have been shown to have an increasing impact on the participant's physiology and driving style [7, 9]. In the 0-back test, the participant is required to repeat digits back as they are said. The 1-back test requires the participant to repeat the digits with a delay of 1, and the 2-back test with a delay of 2. Each task is presented in 4 blocks of 10 digits, with a time separation between each digit of around 2.5 seconds. An example block of 10 digits is shown in Table 1, with expected responses for the 0-, 1- and 2-back tests. In order to continue with the experiment, the participant must show a minimum proficiency of 8 out of 10 correct responses for two consecutive blocks of each task.

In order to have a controlled environment and minimize unexpected events, the protocol must be performed on a simulated highway test track. This track is quiet in comparison to real world roads, has 4 lanes, and is used solely by automotive engineers who may be using the track at the same time as the experiment. The participants are instructed to drive in the second lane at usual highway speeds of around 70mph, changing lanes to overtake when necessary. Because this is likely to be an unfamiliar vehicle and a new environment for the participants, a habituation period is used (stage 3). Before the commencement of the habituation period, the vehicle is driven onto the track by the participant.

Stimulus	1	5	9	3	0	2	3	3	2	9	&	&
0-back	1	5	9	3	0	2	3	3	2	9		
1-back	-	1	5	9	3	0	2	3	3	2	9	
2-back	-	-	1	5	9	3	0	2	3	3	2	9

Table 1: Example of the N-back test with a block of 10 numbers. In place of "&" the word "and" is said by the experimenter, requiring the participant to provide a response. Where there is a "-" no response is required by the participant.

	Stage	Time (minutes)
1.	Sensor verification	2:00
2.	Task practice	5:00
3.	Habituation period	25:00
4.	Drive (reference)	3:00
5.	N-back test A	2:30
6.	Drive (recovery)	3:00
7.	N-back test B	2:30
8.	Drive (recovery)	3:00
9.	N-back test C	2:30
10.	Drive (recovery)	3:00
	Total	51:30

Table 2: The protocol for the experiment, employing three N-back tests of different difficulties, presented in a random orders.

Once the driver is comfortable on the track, a reference period under normal driving is used (stage 4), with all sensors being recorded. At stage 5, after this reference period, the protocol alternates between N-back tests and recovery periods of normal driving (stages 5–10). Each participant undergoes each of the 0-, 1- and 2-back tests in a random order. Each of the N-back tests consists of 4 blocks of 10 digits, with a block separation of 5s. At the beginning of the first of the 4 blocks, a brief explanation and reminder of the test being performed is provided. This explanation takes 30s, while the four blocks take the remaining 2 minutes posted in Table 2. The recovery periods are each of normal driving, with no secondary task. Once each task has been performed and the final recovery period has taken place, the vehicle is then taken off the track and data recording is ended.

3. DATA COLLECTION

There are over 1000 signals that can be recorded from the vehicle's CAN-bus. Those signals which are expected to have relevance to driver workload include, steering wheel angle, pedal positions and vehicle speed. Many others are likely of no relevance to driver monitoring and should be removed before attempting to predict driver workload. However, to ensure that all the relevant signals are present in the dataset, we recorded the full set of signals at a sample rate of 20hz during the experiment. Each of these signals was written to a hard disk by a data logging system located under the passenger seat.

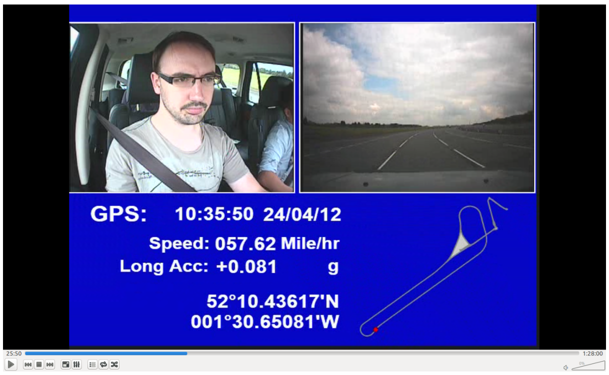

Figure 1: Screen shot of the video output recorded during the experiment, with driver and forward facing cameras and GPS details overlaid.

ECG and EDA signals were recorded via a GTEC USB biosignal amplifier (USBamp). Three point ECG gel electrodes were attached on the driver's chest, close enough together to minimize any noise generated through shoulder movement. The adhesive gel EDA electrodes were attached on the participant's non-dominant hand, on the underside of the index and middle fingertips. Surgical tape was then used to further secure them in place, minimizing any movement of the sensor contacts while driving. The wires from the ECG electrodes came out of the top of the participants shirt, while the EDA wires were positioned to the side of the non-dominant hand. Note that the vehicle used has an automatic transmission and the driver does not need to use their hands for gear selection.

The GTEC USBamp resides in the rear of the vehicle, with sensor wires positioned away from any intrusion of the driver. This connects to a laptop, where the data was recorded at 256Hz. The laptop also had input from the CAN-bus time signals for synchronization purposes, which is provided at 10Hz. In order to match these signals in time, therefore, some re-sampling is performed. Further to this, driver and forward facing cameras record video throughout the experiment, with GPS time overlaid on the image, as shown in Figure 1.

From this data, there are five ground truths that we use to produce classification problems. These are extracted from the timings of the tasks during the experiment, the EDA signal, and the ECG signal. The timings of the tasks provides a ground truth of what the participant was doing at a given point in time. The EDA signal provides two measurements, the Skin Conductance Level (SCL) and frequency of Electrodermal Responses (EDR), both of which are known to increase while a participant is under high workload [7, 5, 1]. The skin conductance level is provided by the absolute value of the EDA signal, whereas EDRs are found by spikes, as illustrated by the red dots on the EDA signal in Figure 2. Finally, two ground truths can be extracted from the time differences between R-peaks, highlighted by the red dots on the ECG signal in Figure 3. Heart Rate (HR) is calculated as the number of R-peaks per minute, whereas Heart Rate Variability (HRV) is a measure of the variation of the time

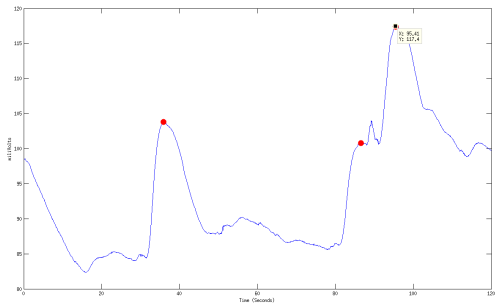

Figure 2: Two minutes of an EDA signal recorded during driving. The red dots highlight EDRs, which increase in frequency under workload. The SCL is given by the signals absolute value.

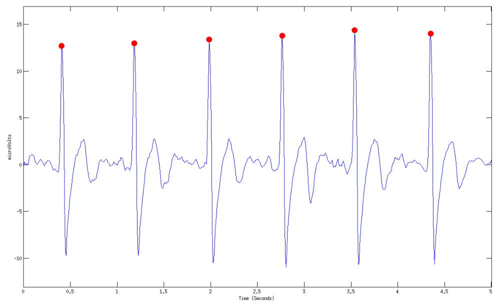

Figure 3: Five seconds of an ECG signal recorded during driving. The red dots highlight the R-peaks, which can be used to compute the HR and the HRV.

delays between R-peaks [7, 2, 5, 8]. Under higher workload demands, HR is known to increase and HRV has been shown to decrease. In computing HRV we opt to use *Standard Deviation of Successive Differences* (SDSD) of RR-intervals, as a result of findings by Mehler *et al.* [8].

From each of these ground truths, both binary and multi-class classification problems are constructed. The binary classification problems all have class labels of *Normal* driving and *Distracted* driving. If the timings ground truth is used, the label is *Normal* unless a secondary task is being performed, in which case it is *Distracted*. For all the other ground truths, a value close to the baseline is *Normal*, and a significant change from the baseline is *Distracted*.

The multi-class classification problems are very similar, but the *Distracted* label takes account of different amounts of difficulty, workload or physiological response. For instance, the timings ground truth can provide three levels of difficulty of the secondary task, relating to which of the 0-, 1- and 2-back tests were being performed. From the HR, HRV and EDA signals, the amount of change can be used in providing more detail on the level of workload, such as a small change, medium change, or large change. In these cases, the labels are be *Normal*, *Low*, *Medium* and *High*, relating to the difficulty or workload level.

For this dataset we executed the protocol with 20 participants, selected from people who are regular drivers, but who have not previously driven on the test track. A Range Rover Sport was used, and was the same vehicle throughout to maintain consistency for both the CAN-bus data and each participant. The direction of the test track is reversed once per week, meaning that around half the participants travel clockwise, and around half travel anti-clockwise.

4. DATA RELEASE

The dataset is available for download via `www.dcs.warwick.ac.uk/dmd/` in a comma separated variable (csv) format, with samples in temporal order at 20Hz. Each of the 10 class labels are be provided for each of these samples. The physiological data are also be available, as this may have other uses to researchers. This physiological data has timestamps, so that it can be associated with the CAN-bus data, but the sample rate remains at 256Hz.

Because many of the signals recorded are be irrelevant to the problem, these have been removed before the release of the dataset. To avoid any human selection bias, correlation analysis with Mutual Information (MI) [12] is used; where features with a MI below a threshold have been removed. Some of those which are kept have been obfuscated so that commercially sensitive details of the CAN-bus and telemetry signals are not made publicly available.

The production and release of such a dataset may benefit both the driver monitoring and data mining communities. The data naturally has high autocorrelation, and several irrelevant and redundant signals; all of which affect the performance of a classification system [6]. As well as this, some of the signals may be correlated with time, introducing biases. Overcoming these issues is not only essential to predicting driver behaviour, but they are also difficult problems for data mining in general. We provide a central dataset against which driver workload monitoring methods and temporal data mining techniques can be evaluated and compared.

5. CONCLUSION

In this paper we have outlined a procedure for collecting a dataset for the driver monitoring problem. Five ground truths are provided, taken from experiment timings and physiological data. The experiment timings contain when a secondary task is being performed, and which task that was. The physiological data, namely ECG and EDA, provide HR, HRV, SCL and frequency of EDRs as ground truths, each providing two sets of class labels.

This dataset will be released for public use, with several vehicle telemetry signals and the 10 class labels. As well as this, the raw physiological data will be released, as this may be used for other forms of analysis.

If the outcomes of analysis of this dataset and collection procedure are positive, then we intend to use a similar set-up for collecting a second dataset, which is more representative of real world driving. For instance, it would be more realistic if EDA or ECG could be used for ground truth, independent of a secondary task such as the N-back test. In future, therefore, subjects may be made to drive for long periods of time under normal circumstances on public roads. The ECG and EDA sensors might then provide a reliable ground truth for real world workload, for use in a classification task.

6. REFERENCES

[1] W. Boucsein. *Electrodermal activity.* Springer, 2012.

[2] B. Cain. A review of the mental workload literature. Technical report, Defence Research and Development Toronto (Canada), 2007.

[3] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):596–614, 2011.

[4] M. Farsi, K. Ratcliff, and M. Barbosa. An overview of controller area network. *Computing Control Engineering Journal*, 10(3):113–120, 1999.

[5] J. Healey and R. Picard. Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 6(2):156–166, 2005.

[6] R. Kohavi and G. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

[7] B. Mehler, B. Reimer, and J. Coughlin. 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 Factors*, 54(3):396–412., 2012.

[8] B. Mehler, B. Reimer, and Y. Wang. A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, pages 590–597, 2011.

[9] B. Reimer, B. Mehler, Y. Wang, and J. F. Coughlin. A field study on the impact of variations in short-term memory demands on drivers' visual attention and driving performance across three age groups. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(3):454–468, 2012.

[10] P. Taylor, F. Adamu-Fika, S. S. Anand, A. Dunoyer, N. Griffiths, and T. Popham. Road type classification through data mining. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 233–240, 2012.

[11] P. Taylor, N. Griffiths, A. Bhalerao, A. Dunoyer, T. Popham, and Z. Xu. Feature selection in highly redundant signal data: A case study in vehicle telemetry data and driver monitoring. In *International Workshop Autonomous Intelligent Systems: Multi-Agents and Data Mining.* Springer, 2013.

[12] I. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (Morgan Kaufmann Series in Data Management Systems).* 2005.

[13] M. Wollmer, C. Blaschke, T. Schindl, B. Schuller, B. Farber, S. Mayer, and B. Trefflich. Online driver distraction detection using long short-term memory. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):273–324, 2011.