# Warwick-JLR Driver Monitoring Dataset (DMD): Statistics and Early Findings

**Phillip Taylor**
The University of Warwick
Coventry, CV4 7AL
phil@dcs.warwick.ac.uk

**Nathan Griffiths**
The University of Warwick
Coventry, CV4 7AL

**Abhir Bhalerao**
The University of Warwick
Coventry, CV4 7AL

**Zhou Xu**
Jaguar Land Rover Research
Coventry, CV3 4LF

**Adam Gelencser**
Jaguar Land Rover Research
Coventry, CV3 4LF

**Thomas Popham**
Jaguar Land Rover Research
Coventry, CV3 4LF

## ABSTRACT

Driving is a safety critical task that requires a high level of attention and workload from the driver. Despite this, people often also perform secondary tasks such as eating or using a mobile phone, which increase workload levels and divert cognitive and physical attention from the primary task of driving. 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. Another approach may be to monitor workload online using readily available and robust sensors accessible via the vehicle's Controller Area Network (CAN). In this paper, we present details of the Warwick-JLR Driver Monitoring Dataset (DMD) collected for this purpose, and to announce its publication for driver monitoring research. The collection protocol is briefly introduced, followed by statistical analysis of the dataset to describe its structure. Finally, the public release of the dataset, for use in both driver monitoring and data mining research, is announced.

### Author Keywords

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

### ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

## INTRODUCTION

Driving is a safety critical task that requires a high level of attention and workload from the driver. Despite this, people often also perform secondary tasks such as eating or using a mobile phone, which increase workload levels and divert cognitive and physical attention from the primary task of driving [6, 10, 7, 17, 8]. In addition to these, the driver may 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 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, or aid the driver in other ways. Traditionally, such systems have monitored physiological signals such as heart rate or skin conductance measures [1, 16, 5]. Such approaches are not practical for everyday use, however, 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 can be unreliable in poor light conditions [9].

A alternative approach, that does not rely on intrusive or unreliable inputs, is to use telemetry data accessible via the CAN-bus [3]. When a driver is performing additional tasks unrelated to driving and is under higher workload, changes can be observed in features such as the Steering Wheel Angle (SWA) [2, 5, 14, 16]. The CAN is a central bus to which all devices in the vehicle connect and communicate by a broadcast protocol [3]. 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 offline data analysis. 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 workload of a driver [12] or the state of a vehicle, such as the current road type [11].

We have previously outlined a protocol for collecting a dataset for driver monitoring research [13]. In this paper, we present initial analysis of this dataset and detail its structure.

 AutomotiveUI '15, September 01 - 03, 2015, Nottingham, United Kingdom © 2015 ACM. ISBN 978-1-4503-3736-6/15/09$15.00 DOI: http://dx.doi.org/10.1145/2799250.2799286

Table 1: Experiment protocol and mean durations of stages.

| | Stage | Mean duration (s) | Std (s) |
|---|---|---|---|
| 1 | Habituation | 1302 | 269 |
| 2 | Baseline | 280 | 99 |
| 3 | 0-back (introduction) | 10 | 2 |
| 4 | 0-back | 82 | 9 |
| 5 | 0-back (recovery) | 256 | 59 |
| 6 | 1-back (introduction) | 10 | 2 |
| 7 | 1-back | 100 | 12 |
| 8 | 1-back (recovery) | 300 | 81 |
| 9 | 2-back (introduction) | 11 | 6 |
| 10 | 2-back | 113 | 15 |
| 11 | 2-back (recovery) | 294 | 127 |

## PROTOCOL

The experimental protocol we use is based on that performed by Reimer and Mehler *et al.* [9, 5], and is outlined in Table 1. During the habituation period, the driver is asked to drive as if they were on a highway to get used to the unfamiliar vehicle and environment. After approximately 20 minutes (or more in cases where the driver did not yet appear comfortable), a baseline period of 4 minutes began. The protocol then alternated between one of the *N*-back tasks and a recovery period of normal driving, again of around 4 minutes. The 0-, 1-, and 2-back tasks were presented to the drivers in a random order, and each participant performed each task once. For full details of the protocol, refer to [13].

All digits in the *N*-back tasks were repeated regardless of the shift (in contrast to Reimer and Mehler *et al.* [9, 5]), the 1-back task was in effect one digit longer and the 2-back task was two digits longer, than the 0-back task. This is reflected in their mean durations shown in Table 1. Other variances in durations were due to safety concerns, recording quality or human error. In some cases, for example, the physiological measures took longer to return to their baseline values and so the recovery periods were extended. Some events on the road such as low flying birds or overtaking vehicles, for example, caused reactions from the driver that were both out of the control of the experimenter and led to a pause in the protocol or extension of a stage. In other cases, recording quality led to changes in the length of the baseline periods.

## DATASET ANALYSIS

The DMD is analysed first by investigating the subjective ratings and data streams with respect to the secondary tasks. Finally, a ground truth for classification is produced.

### Task performance and subjective ratings

The error rates for the digit recall tasks are shown in Figure 1. The number of incorrect responses for the 0-back test were very low on average, and there were no errors for the majority of participants. In the 1-back test the number of errors were higher, and for the 2-back task there were even more incorrect responses on average. In some 2-back test blocks the participant stopped responding to numbers, and the remainder of block was counted as incorrect responses. In some other cases that were also counted as errors, the participant responded in the 2-back test was if it were the 1-back test.

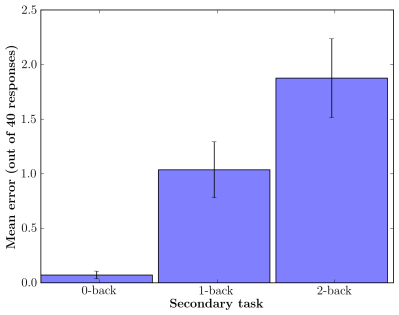



Figure 1: Mean error rates (out of 40 recalled digits) of participants during each of the secondary tasks. Error bars represent the standard error.

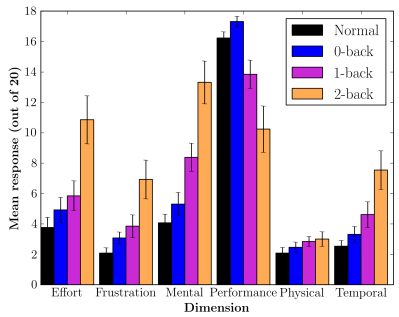



Figure 2: Mean responses to NASA TLX questions. Error bars represent the standard error.

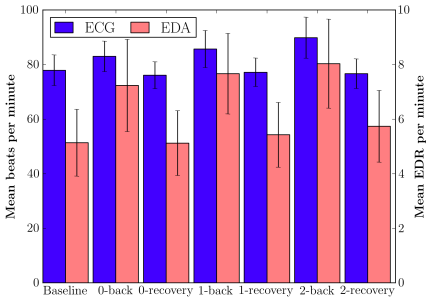
When the protocol was complete, the participants were asked to fill in four NASA-Task Load Index (TLX) questions – one for normal driving and for each of the *N*-back tasks. The TLX asks participants to rate their experiences out of 20 in 6 dimensions, namely: mental demand, physical demand, temporal demand, performance, effort, and frustration. Figure 2 presents the mean responses of the participants by the TLX Responses in general indicate that driving with the secondary tasks were harder, and that the difficulty increased with the delay in the digit recall tasks. The mental demand and effort dimensions, as expected, reported the largest increase in responses. The estimated performances decreased with the 1- and 2-back tasks, reported performance increased on average for the 0-back test over normal driving.

### Analysis of data streams

There were two data streams inspected, namely the physiological and vehicle telemetry data streams. Results of statistical analyses of both are shown in Table 2, comparing normal and distracted conditions in two ways to detail properties of the dataset. First, the mean of measurements over all subjects during normal (baseline or recovery) periods and distracted periods (during a secondary task) and were compared using a two *t*-test. Second, Analysis of Variance (ANOVA) is used to determine if there was a significant difference in means during any of the three secondary task periods and normal driving. In follow-up to this, a four way pairwise *t*-test was

Table 2: The $p$-values from two way $t$-test and ANOVA for the physiological and selected signals from the vehicle telemetry data streams. In the heading **N** represents periods of normal driving, **0**, **1**, and **2** represents periods of the 0-, 1- and 2-back tests respectfully, and **D** is periods where any of the $N$-back tasks were being performed.

| Signal | Feature | $p$-value | N vs. D | N vs. 0 | N vs. 1 | N vs. 2 | 0 vs. 1 | 0 vs. 2 | 1 vs. 2 |
|---|---|---|---|---|---|---|---|---|---|
| ECG-PEAKS | raw | **0.031** | **0.006** | 1.000 | 0.422 | **0.050** | 1.000 | 1.000 | 1.000 |
| EDA-PEAKS | raw | **0.034** | **0.004** | 0.605 | 0.265 | 0.122 | 1.000 | 1.000 | 1.000 |
| Adaptive Cruise Control Cancel (by brake) | STD | 0.232 | 0.056 | 1.000 | 0.419 | 1.000 | 1.000 | 1.000 | 1.000 |
| Brake on | STD | 0.239 | 0.057 | 1.000 | 0.436 | 1.000 | 1.000 | 1.000 | 1.000 |
| Engine Speed | raw | 0.237 | 0.063 | 0.414 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Engine Torque | raw | 0.053 | **0.016** | 0.067 | 1.000 | 0.672 | 1.000 | 1.000 | 1.000 |
| Engine Coolant Temperature | STD | 0.190 | **0.036** | 0.362 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Gear Selected (automatically) | raw | 0.085 | **0.012** | 0.207 | 1.000 | 0.556 | 1.000 | 1.000 | 1.000 |
| Steering Wheel Movement Speed | STD | **0.003** | 0.066 | 1.000 | 0.087 | 0.055 | **0.030** | **0.020** | 1.000 |
| Steering Wheel Angle | STD | **0.024** | 0.471 | 0.555 | 0.423 | 0.968 | **0.039** | 0.095 | 1.000 |
| Suspension Height (front-right) | STD | 0.091 | 0.213 | 0.089 | 1.000 | 1.000 | 0.527 | 0.228 | 1.000 |
| Throttle Position | raw | **0.044** | **0.010** | 0.068 | 1.000 | 0.473 | 1.000 | 1.000 | 1.000 |
| Yaw Rate | STD | **0.022** | 0.532 | 0.422 | 0.679 | 0.715 | **0.048** | 0.051 | 1.000 |
| ... | | | | | | | | | |




(a) Heart rate

Figure 3: Mean heart rate and EDR over all subjects for the different periods of the trial. Each recovery period is presented separately and error bars represent the standard error.

performed and normalized by the Bonferroni correction. All results in this table produced $p$-values of less than 0.1 in at least one of the $t$-test and the ANOVA and any $p$-value smaller than 0.05 is highlighted in bold. The authors accept that conclusions made from this analysis are limited because it is a multiple comparisons procedure, but a two-way ANOVA including all signals is impractical due to their number.

The physiological data consisted of the ECG and EDA signals, from which the heart rate and electrodermal response (EDR) frequencies were extracted respectively, both of which are measured in beats or responses per minute. The two way $t$-test showed a significant difference in all physiological measurements with $p < 0.01$, and the ANOVA produced a significant difference between at least one of the baseline or task periods ($p < 0.05$); shown in the top section of Table 2. In the pairwise $t$-tests, however, only the difference in mean heart rate of the 2-back task and normal driving periods was found to have any significant difference ($p < 0.05$). Figure 3 shows the mean heart rate (left) and EDR frequencies (right) computed over the full baseline, task, and recovery periods. Both physiological measures increased during the $N$-back tasks, and increased more with higher difficulty tests.

In the lower section of Table 2 the results of the $t$-test and ANOVA are shown for representative signals from the vehicle telemetry data that had a $p$-value less than 0.1. As well as the mean of the raw signal values, the standard deviation (STD) was computed for each signal over a one second sliding window. This produces a feature of the signals where sample values are equal the STD of the twenty samples before and after the respective sample in the signal. Signals that were expected to have a close relationship to the driver workload were those related directly to the driving controls, such as the pedals and steering wheel. The analysis shows that the throttle position and STD of the steering wheel angle speed both have a close relationship to the driving period ($p < 0.05$ in both the two way $t$-test and ANOVA). The STD of the SWA however, was not as closely related to the driving period, which was unexpected. In fact, in the data the STD of the SWA decreased from the baseline during the 0-back task and increased during the 1- and 2-back tasks.

Signals with indirect relationships to the vehicle controls were expected to have weak relationships to the driving conditions. These had larger $p$-values in general than measures of the vehicle controls, such as with the STDs of both the suspension measurements and yaw rate. The raw values of the engine speeds and target gear of the automatic gear box, however, had relationships more similar to those of the vehicle controls. Other signals that have no obvious link to the driver were of course expected to have large $p$-values, and for the majority this was the case. A small number, including ACCCancelRequest, engine coolant temperature, and others redacted from Table 2, had small $p$-values for the two way $t$-test and can only be explained by chance.

### Ground truth for classification

Both the timings of tasks and the physiological data streams are used to produce ground truths. The task timings can be used as one ground truth to create a binary labelling to describe whether there was a secondary task being performed or not. Here, the label *normal* relates to driving under normal conditions and *distracted* signifies that a secondary task was being performed. The *distracted* label is then also split into three to signify which of the 0-, 1- or 2-back tasks was

being performed, to produce a multi-label classification problem with four labels.

Each of the physiological data streams can be used to produce binary classification tasks, with a label of *normal* when the observations are close to those found during the baseline period, and *distracted* otherwise. Other levels can also be used to produce a multi-label classification problem. For example, increases of 5% or less can be assigned label *A*, of between 5% and 10% given label *B*, and of more than 10% label *C*.

## 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 class labels are provided for each sample. The physiological data are also available. This physiological data has timestamps, so that it can be associated with the CAN-bus data, but the sample rate remains at 256Hz.

Several features have been removed from the dataset to either protect intellectual property or because they are irrelevant to the problem. To avoid any human selection bias, correlation analysis with Mutual Information (MI) [15] is used; where features with a MI below a threshold have been removed.

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 [4]. 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.

## REFERENCES


1. Y. Dong, Z. Hu, K. Uchimura, and N. Murayama. 2011. Driver Inattention Monitoring System for Intelligent Vehicles: A Review. *IEEE Transactions on Intelligent Transportation Systems* 12, 2 (2011), 596–614.
2. T. Ersal, H. Fuller, O. Tsimhoni, J. Stein, and H. Fathy. 2010. Model-Based Analysis and Classification of Driver Distraction Under Secondary Tasks. *IEEE Transactions on Intelligent Transportation Systems* 11, 3 (2010), 692–701.
3. M. Farsi, K. Ratcliff, and M. Barbosa. 1999. An overview of controller area network. *Computing Control Engineering Journal* 10, 3 (1999), 113–120.
4. R. Kohavi and G. John. 1997. Wrappers for feature subset selection. *Artificial Intelligence* 97, 1-2 (1997), 273–324.
5. B. Mehler, B. Reimer, and J. Coughlin. 2012. 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 (2012), 396–412.
6. D. Redelmeier and R. Tibshirani. 2001. Car phones and car crashes: some popular misconceptions. *Canadian Medical Association Journal* 164, 11 (2001), 1581–1582.
7. M. Regan. 2005. Driver Distraction: Reflections on the Past, Present and Future. *Journal of the Australasian College of Road Safety* 16, 2 (2005), 22–33.
8. M. Regan, C. Hallett, and C. Gordon. 2011. Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention* 43, 5 (2011), 1771–1781.
9. B. Reimer, B. Mehler, Y. Wang, and J. F. Coughlin. 2012. A field study on the impact of variations in short-term memory demands on drivers' visual attention and driving performance across three age groups. *The Journal of the Human Factors and Ergonomics Society* 54, 3 (2012), 454–468.
10. J. Stutts, J. Wilkins, and B. Vaughn. 1999. *Why do peolpe have drowsy driving crashes: input from those who just did*. Technical Report.
11. P. Taylor, F. Adamu-Fika, S. Anand, A. Dunoyer, N. Griffiths, and T. Popham. 2012. Road type classification through data mining. In *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 233–240.
12. P. Taylor, N. Griffiths, A. Bhalerao, A. Dunoyer, T. Popham, and Z. Xu. 2013a. 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, 25–36.
13. P. Taylor, N. Griffiths, A. Bhalerao, D. Watson, Xu Zhou, and T. Popham. 2013b. Warwick-JLR Driver Monitoring Dataset (DMD): A public Dataset for Driver Monitoring Research. *Proceedings of the Cognitive Load and In-Vehicle Human-Machine Interaction Workshop* (2013).
14. K. Torkkola, N. Massey, and C. Wood. 2004. Detecting driver inattention in the absence of driver monitoring sensors. In *Machine Learning and Applications, 2004. Proceedings. 2004 International Conference on*. 220–226.
15. I. Witten and E. Frank. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
16. M. Wollmer, C. Blaschke, T. Schindl, B. Schuller, B. Farber, S. Mayer, and B. Trefflich. 2011. Online Driver Distraction Detection Using Long Short-Term Memory. *IEEE Transactions on Intelligent Transportation Systems* 12, 2 (2011), 273–324.
17. K. Young and M. Regan. 2007. Driver distraction: A review of the literature. *Distracted driving. Sydney, NSW: Australasian College of Road Safety* (2007), 379–405.