Investigation of Train Driver Physiological Responses

Katie Crowley
Trinity College Dublin
College Green, Dublin 2
k.crowley@tcd.ie

Nora Balfe
Trinity College Dublin
College Green, Dublin 2
balfen@tcd.ie

Train Driver workload is an under-researched area. Operator workload has been extensively studied in the automotive, aeronautical and other domains using performance, subjective and physiological measures. In this exploratory study, we combine subjective self-report measures with a task-based measure of workload and physiological measures. Heart Rate and Galvanic Skin Response are collected from train drivers over the course of their journey. These signals are analysed with respect to subjective and task-based measures of workload, but no reliable correlations were found between the physiological and other workload measures. However, the results show that peaks in both the Heart Rate and GSR data are associated with particular locations or events and changes in GSR data reflect anticipatory events and are inline with subjective driver commentary. This suggests that further research on physiological measures for train drivers is warranted.

Workload measurement, rail human factors, heart rate, galvanic skin response, psychophysiology, ambulatory monitoring, wearable sensors.

1. INTRODUCTION

Train driving is a highly specialised skill requiring detailed knowledge of traction characteristics, local geography, and railway procedures. The safety critical nature of the task means that it has been the subject of human factors research as far back as 1979 [1] with a particular focus on the train driving task and high-risk events such as signals passed at danger (SPADs). More recent research, such as the studies conducted by Naweed (e.g. [2,3,4]) have developed models of train driver information processing and behaviour to better describe the task and processes involved in train driving. The focus on SPADs has also led to several studies using eye-tracking to study train driver visual behaviours (e.g. [5,6]), but other applications of physiological measures are rare.

This paper describes the collection of heart rate (HR) and Galvanic Skin Response (GSR) data from train drivers over the course of their journey. The data was collected as part of a study of train driver workload, and this paper will discuss the collection, interpretation, and analysis of the data in this context as well as more generally.

A range of methods attempt to measure workload. These methods generally focus on different facets of workload, such as self-report questionnaires (e.g. NASA-TLX: [10]) and physiological measurement. There is consensus, however, that at least three components are important for measuring workload. These components are subjective, physiological and performance measures [11]. Physiological signals are a useful metric for providing feedback about a driver’s state because they can be collected continuously and without interfering with the driver’s task performance. This information could then be used automatically by adaptive systems in various ways to help the driver better cope with their workload.
While research investigating the use of physiological signals as a proxy for driver workload has been undertaken for automobiles [e.g. [12, 13, 14] and aeroplanes [e.g. [15, 16, 17]] and also for rail signallers, to our knowledge, no research has examined the use of wearables sensors to estimate train driver workload.

In contrast to rail signalling, where several specific workload tools have been developed (e.g. [7]), and rail signaller workload has been extensively researched [8], train driver workload is under-researched. In an attempt to redress this balance, Balfe et al [9] developed a method to extract train driver taskload from downloads of on-train-data-recorders (OTDR) to capture train driver activity and thereby attempt to calculate driver taskload in terms of an ‘actions per minute’ metric. This paper compares the taskload recorded over short, commuter journeys with physiological data from heart rate and skin response recorded over the same period. This could provide insight into the work demand placed on drivers across the network.

2. PHYSIOLOGICAL MEASUREMENT

Heart rate and skin response have been used to monitor task demand on pilots [18,19,20,21], while more extensive sensor suites have been used in automotive research, including the use of electrocardiogram, electromyogram, skin response and respiration by Healey et al [12] for detection of driver stress.

Stress is a physiological response to the mental, emotional, or physical challenges that we encounter. Immediate threats provoke the body’s “fight or flight” response, or acute stress response. The body secretes hormones, such as adrenaline, into the bloodstream to intensify concentration. There are also many physical changes, such as increased heart rate and quickened reflexes. Under healthy conditions, the body returns to its normal state, homeostasis, after dealing with acute stressors.

The autonomic nervous system (ANS) regulates the body’s major physiological activities, including the heart’s electrical activity, gland secretion, blood pressure, and respiration. The ANS has two branches: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS mobilises the body’s resources for action under stressful conditions. In contrast to the SNS, the PNS relaxes the body and stabilizes the body into steady state. Under acute stress, the SNS increases heart rate, respiration activity, sweat gland activity, etc. After the stress has passed, the PNS releases hormones to restore homeostasis. Since the ANS controls the heart, measuring cardiac activity is an ideal, non-invasive means for evaluating the state of the ANS.

Galvanic Skin Response (GSR) (also known as skin conductance) is a measure of the electrical resistance of the skin. A transient increase in skin conductance is proportional to sweat secretion [22]. When an individual is under mental stress, sweat gland activity is activated and increases skin conductance. Since the sweat glands are also controlled by the SNS, skin conductance acts as an indicator for sympathetic activation due to the stress reaction.

There are many challenges relating to ambulatory measurement of physiological data that are beyond the scope of this paper but should be noted. While developing physiological monitoring algorithms in real-life ambulatory situations, it is crucial to take physical activity (e.g. walking) and posture (e.g. sitting or standing) into account. Cardiovascular variability is highly affected by changes in body posture (people may have higher heart rate when standing than when sitting) and physical activity [23, 24]. In Van Steenis et al.’s sample, subjects’ mean HR increased significantly from a supine to sitting posture (from 66 to 77 bpm), from a sitting to standing posture (86 bpm), and from a standing posture to dynamic body movements (92 bpm). As activity is influenced by ANS activation, heart rate may also increase when people are mentally stressed. A major obstacle for ambulatory monitoring is that physiological dysregulation or emotion effects can be confounded by physical activity and hence, using heart rate alone as an indicator to detect mental stress may lead to misclassification. It is also important to note when conducting ambulatory physiological measurement that signal artefacts caused by motion, electrode placement, or respiratory movement can affect the accuracy of measured recordings.

3. DRIVER WORKLOAD

Solovey et al [13] note in relation to automotive driving that it is a “dynamic, complex activity involving visual, cognitive and manual tasks: the driver has to form strategic goals, monitor the roadway environment and the vehicle systems,
process information and make tactical action plans as well as execute control level activities” [13, 25]. The driving task imposes varying levels of workload on all driver types, including train drivers. Understanding the workload induced during driving is important for preventing accidents and hazards, and in a relatively controlled system such as a railway workload measurement could also be used to determine changes to the network that could improve human driving performance. It has been shown that operators perform better at intermediate levels of workload compared to extreme levels (i.e. too low or too high workload) [26]. The Yerkes Dodson law of arousal [27] (inverted U) that suggests that during periods of underload, added workload may improve performance, while during heightened demand, higher workload may reduce performance. With the added factor of advances in technology, additional demands in many working environments from the largely physical to more supervisory oversight of automation, have in many cases increased cognitive demand. This development can be observed in the field of driving as well [28, 29].

Although mental workload is not directly observable [13, 30], researchers deploy several measurement approaches to assessing mental workload. Empirical measures are generally divided into three categories: subjective measures, performance measures and psychophysiological measures. Each method has advantages as well as disadvantages [13, 31]. Subjective measures, while cost-effective and suitable in prototype testing, can be, however, subject to biases, can suffer from time delay, and can be intrusive which ultimately makes them unsuitable for continuous workload assessment. Conversely, performance and psychophysiological measures can be measured non-intrusively and continuously throughout the tasks [3, 31]. Combining these objective and subjective approaches in a balanced manner yields minimally-intrusive contextualised ambulatory data.

4 Method

In this study the authors combine performance, physiological and subjective measures to investigate the estimation of train driver workload as measured in an ambulatory task setting. The on-train-data-recorder (OTDR) captures all actions taken by the driver on the train, and hence can be used to calculate an ‘actions per minute’ metric to estimate driver taskload. This could provide insight into the work demand placed on drivers across the network. Subjective workload ratings were measured using a nine point Integrated Workload Scale (IWS). Physiological measurements of heart rate and Galvanic Skin Response were also recorded using two wrist based data collection devices.

Data were collected from eight return journeys (16 journeys in total) over two routes. Five different drivers participated in the study. A researcher travelled with the driver on each journey and noted any unusual events or deviations from standard practices. The researcher also collected subjective and physiological data over each journey. The journeys were all scheduled passenger services, and the research did not require any changes to the timetabled journey.

Route 1 was a suburban commuter route between an urban city centre and outlying commuter town. The route covers almost 28km, is timetabled to take 48 minutes, and serves 13 stations, including the start and terminating stations. Route 2 was a shorter commuter route. The journeys were chosen to provide sufficient timetabled turnaround time to set up the study equipment for both the outbound and inbound journeys.

Balfe et al [3] describes the data signals analysed, data preparation and a method to extract train driver taskload from downloads of on-train-data-recorders (OTDR). The paper describes the type of data held in OTDR recordings and how they can be transformed into driver actions throughout a journey. A detailed description of this method is beyond the scope of this paper but more detail can be found in Balfe et al [3]. The signals analysed from the OTDR provided data on train driver movement of controls in the train cab. All signals were logged as bitcodes (0/1); in addition, analogue signals of time, speed, and GPS coordinates were downloaded and exported for each journey. This allowed the OTDR data to be mapped to the subjective and physiological data collected simultaneously. The data were then pre-processed as outlined in [3]. This dataset provided the basis for the analysis of train driver actions or taskload.

4.1 Driver taskload computation

The dataset was used to calculate driver taskload by identifying the times of driver actions. The actions identifiable from the dataset are described in full in Balfe et al [3], but include application of
train power and brakes, door operation, and operation of the safety systems on board. The dataset provides information on all routine actions performed by the driver to control the train and driver a measure of taskload can be calculated from these data by summing the number of actions within a set time period (e.g. the number of actions per minute). However, there are a number of driver tasks that are not logged in the data, particularly communications and passenger interactions (e.g. operating the passenger information system, responding to passenger queries, etc.) as well as perception and interpretation of the visual scene. The OTDR also gives limited insight into the cognitive processing associated with the actions but despite these limitations it may still be useful for monitoring and comparing different journeys and different journey phases.

4.2 Physiological Data Recording

In addition to driver taskload estimation, the following physiological signals were recorded during each journey:

4.2.1. Heart Rate Data

Collected at a frequency of 1Hz using a wrist based optical sensor – Polar 360 wristband. Heart rate is known to increase with mental workload, but is also sensitive to physical workload.

4.2.2. Galvanic Skin Response

Collected at a frequency of 4Hz using the Empatica E4 wrist based sensor. GSR measures the electrical conductivity of the skin, and hence provides a measure of stress or workload based on perspiration. Fingertip sensors are the most accurate measure but these are relatively obtrusive and are affected by any pressure placed on the sensor so were inappropriate for this application.

4.3 Subjective Data

Drivers were also asked to rate their subjective workload at several points through each journey. The nine point Integrated Workload Scale (IWS) was used for this. This scale has been validated for railway signallers, and was used with train drivers as a simple and validated scale for real-time subjective workload measurement. Drivers were requested to use the nine point numerical scale and ignore the text labels, as these were more applicable to the signalling environment.

5 RESULTS

Due to the individual nature of physiological response and subjective data, journeys were analysed separately (1-16). Taskload, physiological, subjective and observational data were examined for potential correlations. No correlation was found between Heart Rate (HR) and GSR on an individual per journey basis, as identified in some previous studies [32, 33]. The taskload measure did not correlate reliably or strongly with either HR or GSR, or with the subjective ratings. However, the journeys analysed were all similar commuter journeys with regular starts and stops from stations. Comparison with longer journeys, featuring higher speeds and fewer stops, may reveal some differences between journey types.

A complete review of the results is beyond the scope of this paper, however, some journeys and events were of particular interest and provide insight for future research into the application of physiological measures in this environment.

5.1. Heart Rate Data

Positive correlations were consistently found between the heart rate and initiating and changing power, meaning that as power was applied and changed, heart rate tended to go up. Similarly, negative correlations were often found between heart rate and initiating and changing braking, meaning that heart rate tended to go down during these activities. Heart Rate also went up during doors closures.

Figure 1 shows the heart rate data for Journey 6 overlaid with actions per minute. The peaks in the HR graph align with station stops when the driver stood up to close the doors. Three clear peaks can be identified in the signal, corresponding to the only three stops during this journey (Figure 1b). Weak positive correlations between the heart rate and initiating power (0.48) and closing doors (0.37) support this. Two weak negative correlations were found with initiating braking (-0.34) and changing braking (-0.38) (Figure 1a). Increases in HR data aligning with stops/driver standing occur in several journeys, demonstrating an increase in physical workload and in line with previous research [23, 24]. Journey 12 demonstrates a further clear example of this with two peaks in HR corresponding to two instances where the driver alighted the cab to assist a wheelchair-user passenger (Figure 2).
It was also possible to align some of the highest peaks in heart rate with higher risk/workload areas of the network. For example, peaks were more frequently present at a particular unmanned station with a history of anti-social behaviour. Drivers may have been reacting to the possibility of encountering a difficult situation at this station. Similarly, peaks were also present more frequently at city centre stations where there is a higher density of passengers and other train traffic.

5.2. Galvanic Skin Response Data
The GSR data collected in this study showed more relevance to the subjectively experienced workload than the HR data. Peaks in the GSR data were often associated with particular locations or events, and driver comments frequently indicated that higher subjective workload scores were less to do with control actions on the train and more to do with anticipating possible events. There were several instances of the GSR signal peaking as the train approached a speed restriction - drivers knew about these before they start their journey indicating a possible anticipatory effect which would be inline with GSR response.

Figure 3 shows a clear peak in GSR in the middle of Journey 3. The peak occurred during a period of braking, and commentary from the driver during the journey indicated that there was a risk of low rail adhesion (LRA, i.e. train slipping) on this journey due to the current weather conditions. The peak occurred around the time the train was approaching a station with a history of LRA, and so rather than being related to any particular aspect of the taskload, may indicate the driver concern about the driving conditions at that time. This is inline with the anticipatory nature of GSR activation [34].

Journey 6 is represented in Figure 1a and 1b. As previously outlined, three peaks in HR data correspond to three station stops on this journey. A clear GSR peak is also noted for Journey 6. This rise is inline with timeseries commentary relating to approaching a speed restriction zone. The GSR rise begins to decay at approximately minute 18 of this journey and the driver began to apply the brakes for the speed restriction at minute 19, marking an end of the anticipatory phase and the beginning of the action phase, which is reflected in the GSR signal.

In Journey 4, the subjective commentary was also broadly in line with the GSR signal and in Journey 10 a peak in the GSR signal aligns with the driver receiving a phone call. Peaks in GSR broadly align with approaching difficult stations and/or other anticipatory events.

6 DISCUSSION
The taskload measure did not correlate reliably or strongly with either of the physiological measures or the subjective rating. This is perhaps unsurprising as workload is a multi-dimensional concept, and the OTDR taskload measures only one aspect of workload. Peaks in both the Heart Rate and GSR data were often associated with particular locations or events. Driver comments frequently indicated that higher subjective workload scores were less to do with control actions on the train and more to do with anticipating possible events.

Positive correlations were consistently found between the heart rate and initiating and changing power, meaning that as power was applied and changed, heart rate tended to go up. Similarly, negative correlations were often found between heart rate and initiating and changing braking, meaning that heart rate tended to go down during these activities. Heart rate also went up during doors closures. Drivers must apply the brakes before stopping and apply power to move away from a station stop, so it is likely that the increase in heart rate as the train powered away from a station stop is related to the increase from standing to close the doors [23, 24]. The lower heart rate associated with braking may be the result of the driver typically sitting in this phase of driving. However, it is also possible that there may also be increased workload associated with acceleration and reduced workload associated with braking.

Further research would be needed to establish whether the increased heart rate is fully due to physical activity during station activities or whether acceleration also has an impact.

The GSR data collected in this study showed more relevance to the subjectively experienced workload than the heart rate data. The changes seen in the heart rate data were generally explained by physical activity during station stops. However, the GSR data showed more variation and peaks could often be explained by events or driver comments. Increases in GSR signal were regularly inline with anticipatory events such as approaching a speed...
Investigation of Train Driver Physiological Responses
Katie Crowley • Nora Balfe

restriction zone, challenging areas such as Low Rail Adhesion and single occurrence events such as the driver receiving a phone call. Combining GSR data with GPS coordinates might be a fruitful avenue for identifying infrastructure areas which routinely result in elevated stress and workload for drivers.

None of the workload measures (taskload, subjective, or physiological) was sufficient on its own to measure or explain driver workload, but each has its own strengths and applications. The subjective workload score gave detailed insight into the experienced workload, but could not be constantly applied so could not capture small changes in workload. The taskload measure did not correlate with any of the other workload measures and therefore may be limited with respect to directly measuring driver workload. However, it may be useful in monitoring driver underload. The routes analysed in this study were all commuter routes with relatively high taskloads. Longer train journey through rural areas may feature periods of several minutes or more with little or no actions by the driver. Human performance is known to drop during prolonged periods of low activity and this measure could be usefully applied to identify any routes with a significant risk of underload. Further research to develop a more useful taskload measure could also examine the possibility of weighting some tasks more highly. The physiological measures described in this paper are not readily interpretable without additional data, but in conjunction with the driver commentary many of the peaks observed could be aligned with locations, events or anticipation over the course of the journey. The constant nature of the physiological measurement means that it is a powerful tool for monitoring driver workload, and further research under more controlled conditions is recommended to explore the applications of physiological workload measurement in this domain in more detail. Such research may develop a model for interpreting the data, allowing a more nuanced understanding of train driver workload which can then be applied to monitor and manage the workload, and ultimately improve railway safety.

![Figure 1a: Journey 6 Detailed Actions Per Minute and HR per minute](image-url)
Investigation of Train Driver Physiological Responses
Katie Crowley ● Nora Balfe

Figure 1b: Journey 6 Actions, IWS, Normalised HR and GSR per minute. Three HR peaks corresponding to three station stops. GSR peak corresponds to approaching a speed restriction

Figure 2: Journey 6 Actions, IWS and Normalised HR per minute. Two HR peaks corresponding to driver alighting cab

Figure 3: Journey 3 Actions, IWS, Normalised HR and GSR per minute. GSR peak corresponding to a period of braking and potential Low Rail Adhesion
7. REFERENCES


32. Yerkes, R. M., Dodson, J.D. (1908) The relation of strength of stimulus to rapidity of habit-formation, *J of Comparative Neurology and Psychology 18*.


