Train driving simulator studies: Can novice drivers deliver the goods?

David R Large[1], David Golightly[1], Emma Taylor[2]

[1]Human Factors Research Group, Faculty of Engineering, University of Nottingham

[2]Angel Trains Ltd, Derby

Abstract

Early research suggests that, in a simulated train-driving environment, unskilled, novice drivers may exhibit comparable behaviour and performance to experienced, professional train drivers after receiving only minimal, task-specific training. However, this conclusion is based on exiguous performance indicators, such as speed limit exceedances, SPAD violations etc., and considers only limited data. This paper presents further, detailed analysis of driving performance data obtained from 20 drivers (13 novices and 7 experienced train drivers), who took part in a previous simulator-based research study, utilising more sensitive and perspicuous measures, namely acceleration noise and control actuation. Results indicate that, although both cohorts exhibited similar performance using

the original metrics, and would thus support the same conclusions, the manner in which this performance was effected is fundamentally different between groups. Trained novice drivers (mainly comprising students and staff at the University of Nottingham) adopted far more erratic speed control profiles, characterised by longer control actions and frequent switching between power and brake actuation. In contrast, experienced drivers delivered smoother acceleration/braking profiles with more subtle (and shorter) control actions and less variance in speed. We conclude that although utilising trained non-drivers may offer an appealing solution in the absence of professional train drivers during simulator-based research, and their input remains of value, researchers should remain mindful when interpreting results and drawing conclusions from a contingent comprising non-drivers. The work also demonstrates the value of dependent variables such as acceleration noise, and quantitative measures of control actuation, which may offer an insightful portfolio of measures in train driving research studies.

Introduction

Due to the capacious demands imposed on train drivers (long working hours, short-turnarounds, flexible turns, geographical diversity etc.), and in spite of genuine enthusiasm and interest, it can often be difficult to recruit a sufficiently large and representative sample of experienced train drivers to take part in simulator-based research studies. An appealing solution to this problem is to recruit unskilled test participants with no prior train driving experience (of which there is a proliferation within research institutions such as universities), and provide them with a highly focussed, but relatively short, period of training and familiarisation. This can expose them to the most pertinent aspects of driving a train, commensurate with the task in hand, and can be conducted in-situ either as part of the testing regime or during an earlier, pre-arranged visit (1, 2). This approach is common in human factors studies for tasks such as Air Traffic Control (e.g. 3) or military command and control (e.g. 4), where domain experts may be equally hard to access. Early data for train driving simulation (e.g. 1, 2), appear to support this as a plausible solution, in this domain, with most subjective and objective performance metrics obtained during these studies showing no significant differences between groups.

For example, Large et al. (1) investigated the effect of using in-cab driver advisory systems (DAS) on driver workload and performance. DAS are designed to present information – as advice or control values (e.g. speeds, timetabled stops, energy consumption etc.) – typically incorporating real-time traffic information, to support drivers in the safe transport of people and freight in accordance with the timetable, and to minimise delays and energy consumption of the train. During the Large et al. (1) study, thirteen novice and seven experienced drivers were required to learn and negotiate a bespoke, geotypical route while using two DAS concept designs, which displayed either speed or timetable driving advice. Participants also completed a baseline (no DAS) condition. The three drives were undertaken during high and low demand conditions (with/without scheduled stops). Hence, participants completed six drives. All participants received instruction/familiarisation with the operation of the simulator and were provided with the opportunity to build route knowledge required for the study. Identifying and understanding the significance of route characteristics is essential to train driving and impacts on a driver's ability to control their train safely and efficiently. This typically requires comprehensive knowledge/recall of route features such as signal types and location, permitted speeds, level crossings, braking points and stopping distances, and stations (so-called 'route knowledge'). Drivers will also require knowledge of the stations at which they are due to stop according to the timetable on a particular day (i.e. 'scheduled-stops'). During the study, expert driver training was relatively brief (in the region of 10 to 15 minutes), and was primarily focussed on acquiring route knowledge. However, more substantial training and instruction, lasting approximately one hour, were provided to non-drivers. This comprised detailed instruction on train handling, signal aspects, driver vigilance systems etc. Novice drivers also completed several practice drives to ensure they reached a basic threshold level of performance. The results of the study indicated consistency between groups (experienced drivers versus trained novices) regarding both driving performance and overall workload ratings, captured using NASA-TLX (5), in most cases. There was, however, a significant interaction between demand and driver experience for the no-DAS condition, suggesting that in the absence of driving advice, the assessment of overall workload made by novice drivers during the high demand condition was significantly higher to that made by experienced drivers.

In a similar study, Dunn and Williamson (2) recruited 28 drivers and 28 non-drivers to investigate the effect of primary task demand on driving performance. The study revealed significant, adverse, monotony-related effects on performance (induced by low primary task demand) over extended periods of driving. Participants received negligible training, other than five minutes of practice/familiarisation using the simulator, but, in terms of driving performance, were only required to achieve and maintain the designated speed limit. To overcome lack of route knowledge, changes to speed limits were previewed ten seconds in advance for all participants. Other than this, no specific training or additional measures were provided to assist the control group (non-drivers). The analysis indicated some differences in subjective ratings made by drivers and the control group – non-drivers were more worried about their performance than drivers, found the task to be lower in overall workload (determined using NASA-TLX), and rated the driving task as more boring, monotonous and fatiguing than drivers. However, when corrected for multiple comparisons using the Bonferroni method, none of these differences reached significance. There were also no differences between the driver and control groups regarding objective performance measures, such as the time and distance spent travelling at the wrong speed.

Needless to say, in these examples, novice drivers would still clearly lack the highly specialised skills, knowledge of rules and route awareness required to command a train in all situations on a conventional rail network. However, based on the reported data, one may conclude that, after a limited period of driver training, individuals with no prior train-driving experience may be sufficiently skilled at driving a train within a simulated environment for the purpose of contributing to a research study. Moreover, their participation in train driving research may be as valid as that of experienced drivers and their behaviour and opinions may be equally revealing, although this will clearly depend upon the exact nature of the research and metrics under investigation.

One consideration is that, in the aforementioned studies, some of the comparative data were collected using exiguous measures. For example, in Large et al. (1), driving performance was determined from the number of speed limit exceedances and SPAD violations, and comparisons were made between relatively small numbers of participants. In fact, speed exceedances seldom occurred during this study and there were no SPADs. Thus, most participants displayed similar, 'acceptable' performance. Clearly, determining performance based on the occurrence of exceptional events is coarse and may compromise statistical vigour. Moreover, data obtained from a small numbers of participants is likely to lack statistical power. Therefore, concluding that the absence of any significant between-groups results suggests that the behaviour of non-drivers was equally as valid as that of experienced drivers may in fact be presumptive: the absence of evidence is clearly not evidence of absence. It is therefore hypothesized that while these exiguous metrics may *suggest* comparative/adequate performance, the manner in which that performance is effected (specifically, speed control of the train) may be fundamentally different between experienced drivers and trained novices. Clearly, in a real world situation, executing smooth and timely braking is critical for effective handling of a train (6).

Measuring Driving Performance

There is an apparent absence of suitable methodology and/or measures within the train driving literature to examine driving performance and speed control at a fine-grained level, and thus we have looked to automotive research for a suitable approach. In automotive driving research, where the recruitment of experienced and available car drivers seldom presents a challenge, a number of different metrics to determine speed performance and control have been employed. A common solution or approach is to measure the statistical dispersion and variation of speed, for example using mean and standard deviation of speed, speed variance or speed reduction (7, 8). However, these measures typically consider the accumulated speed data profiles of multiple vehicles at a single measurement point and thus are more useful for predicting consolidated driver behaviour or activity at that location, such as accident risk or crash frequency, with recommendations then typically based on 85^{th} percentile characteristics or standard deviations of aggregated speed data.

In the current application, we require a metric that compares the relative performance of individual drivers. Acceleration noise is a measure of speed performance and control that is derived from the speed data of a single vehicle recorded along an entire corridor and is therefore more germane when comparing the performance of different drivers rather than capturing the collective behaviour of a platoon of vehicles. Previously, it has been difficult to capture speed data along corridor segments, primarily due to equipment limitations. However, today, GPS and on-board instrumentation allows this data to be collected with ease. In transport simulators, these data are captured as a matter of course, typically at fraction of a second sampling rates, and thus readily available for analysis. The additional benefit of measuring acceleration noise, compared to other similar speed control or performance metrics, such as acceleration signatures, which visualise driving behaviour as an array (9), is that acceleration noise provides a single metric to quantify a driver's driving behaviour, thus allowing clear comparisons to be made between drivers and different groups.

Acceleration Noise

Acceleration noise is defined as the standard deviation of accelerations, and was first proposed by Herman et al (10) as a means to measure traffic conditions and driving behaviour. Greenwood (11) refined the concept, adopting a standard time base, and thus acceleration noise may be formally considered as "the standard deviation of accelerations of a vehicle calculated under nominally uniform operating conditions from average acceleration readings measured over one second intervals". In practice, determining acceleration noise requires the calculation of discreet accelerations, derived from speed data sampled or captured over an entire journey. It can be calculated by first determining the change in speed during each time sample (the acceleration) and then calculating the standard deviation of the resultant accelerations for the entire journey (the acceleration noise). Given that accelerations are largely driver-dependent and not vehicle-dependent (12), acceleration noise is therefore a good predictor of driving style and behaviour.

In an automotive context, drivers who accelerate or brake harshly, or are late to respond to other road users or changes in road conditions and regulations, are likely to exhibit higher levels of noise compared to drivers who accelerate gently and prepare in advance for changes to speed limits, etc. Furthermore, experienced drivers, who are skilled at reading the road, seldom subject their vehicles to high levels of acceleration/deceleration, and are thus expected to exhibit lower levels of noise. In contrast, novice drivers may be less able to read the road, and ‘unsafe’ drivers may actively choose to engage in erratic ‘risk-taking’ behaviour, therefore resulting in higher acceleration noise (8, 9, 12). It is reasonable to expect that similar effects may be evident amongst train drivers, particularly when comparing consolidated data between different driver groups. Moreover, by recording the control actions used to achieve speed control (i.e. the frequency and mean duration of power and brake application), we would expect further differences to emerge between groups.

To investigate this, the speed-time trajectories and video data from the aforementioned study – Large et al. (1) – were re-examined to determine acceleration noise profiles and the frequency and duration of control actions (power and brake actuation). It was hypothesised that novice drivers would require longer application of control actions to maintain the speed of the train, in line with speed limits/driving advice, and this would be reflected in significantly higher acceleration noise.

Method

Overview of Original Study

The original study conducted by Large et al. (1), investigated the impact of using two designs of capacity-based DAS on train driver workload and performance. Thirteen novice and seven experienced drivers took part, with experienced drivers comprising active and recently retired passenger and freight drivers. The study took place using the University of Nottingham Human Factors train simulator, which comprises a medium fidelity, fixed-base train cab (based on a 319 commuter class train) situated in front of a large, single screen onto which the scenario is projected. A bespoke, geotypical scenario was created using Train Simulator 2013 software and took approximately fifteen minutes to complete. Participants were provided with two prototype DAS using MS PowerPoint. Driver advisory information was displayed on a screen situated in the centre console of the cab either as an explicit speed target or a timetable, highlighting the next timing point and with future points displayed in chronological order below. Before testing, participants received training. For experienced drivers, this primarily focussed on acquiring route knowledge. More extensive training lasting approximately one hour was provided to non-drivers. This also included detailed instruction on train handling, signal aspects, driver vigilance systems etc. Novice drivers were also required to reach a basic threshold level of performance – in terms of train control (speed compliance, stopping distances, passenger comfort) – before testing began. During the study, performance measures were calculated through speed data captured directly from the simulation software. This was used to create a speed-distance profile for each participant that was compared to the optimal line speed.

Current Study

The speed-time trajectories and video data were obtained from the original Large et al. study (1) and used to calculate acceleration noise for each participant during each of the six conditions – no-DAS / control group (C), speed-DAS (S) and timetable-DAS (T), in both low (L) and high (H) demand driving scenarios. For the purpose of comparative evaluation and for maximum perspicuity, acceleration noise was calculated using a time base of 0.1 seconds (the simulation software data sampling rate). Additionally, the videos captured during the three high-demand conditions (HC, HS and HT) were coded on frame-by-frame basis to record the frequency and duration of control actions employed by each participant.

Analysis

Mean values were calculated for each participant/drive, on each dependent variable. Following standard practice for statistical analysis within simulated driving studies (e.g. 1, 2) a repeated-measures analysis of variance (ANOVA) was performed. Bonferroni adjustments were used for multiple comparisons to determine if there were any significant differences in performance between groups, or between conditions. Mean values for acceleration noise were examined using a 3 × 2 × 2 (DAS type by Demand Level by Driver Experience) ANOVA. Mean values for control actuations (mean and total duration, total number and rate per minute) were analysed using a 3 × 2 (DAS type by Driver Experience) ANOVA. Demand was not included as an independent variable for control actions as the different type of drive (low demand involved no station stops; high demand involved three station stops) meant that comparisons between control actions were not meaningful.

Results

Acceleration Noise

Mean values and standard deviations of acceleration are shown in Table 1. The repeated measures ANOVA revealed significant between-subject effects for driving experience ($F(1,18)=21.12$, $p<.001$), with novice drivers displaying significantly higher acceleration noise, overall, than expert drivers (0.139m/s^2 and 0.092m/s^2, respectively). There was also a significant within-subjects effect of scenario demand ($F(1,18)=20.47$, $p<.001$), with higher acceleration noise evident, on average, during the higher demand driving conditions for both expert and novice driver groups (0.124m/s^2 compared with 0.107 m/s^2 for low demand), although there was less variability in experts' data and their absolute values were significantly lower than novices'. There were no significant differences for acceleration noise between DAS types.

Table 1: Acceleration noise - mean (standard deviation)

Group	LC	LS	LT	HC	HS	HT
Novices	0.133 (0.051)	0.116 (0.015)	0.134 (0.053)	0.151 (0.028)	0.150 (0.028)	0.148 (0.029)
Experts	0.090 (0.011)	0.086 (0.010)	0.081 (0.009)	0.094 (0.014)	0.101 (0.015)	0.101 (0.009)
Total	0.118 (0.046)	0.106 (0.020)	0.116 (0.049)	0.131 (0.036)	0.133 (0.034)	0.132 (0.033)

Control Actions

To compare driving behaviour, the video data were coded to capture all interactions with power and brake actuators. At the time of the original study, the train simulator had analogue controllers (i.e. no notching), so interactions were coded to record the start and end of ‘power-increase’, ‘power-decrease’, ‘brake-increase’ and ‘brake-decrease’ actions, thereby revealing mean and total duration, total number and rate-per-minute of these actions. Control actions were naturally punctuated by periods of ‘no-control’ (e.g. when coasting). A repeated-measures ANOVA was conducted to investigate between-subjects (driver experience) and within-subjects (DAS-types) effects for these measures.

Mean duration

The ANOVA revealed a significant difference between experts and novices for the mean duration of control actions ($F(1,12)=7.23$, $p=.0.02$), with novices making longer interactions, on average, than experts (see Figure 1). There were no significant differences for the mean duration of control actions between DAS types.

Total duration

For the total duration of control actions, the difference between expert and novice drivers was significant at $p<.10$ ($F(1,12)=4.41$, $p=.058$), with the observed data indicating that the total amount of time that novices applied power or braking force was longer than for experts (see Figure 2). There was also a significant difference between DAS types ($F(2,24)=5.87$, $p=.008$). Pairwise comparisons revealed that the total amount of time spent actively controlling the train was significantly higher when drivers were provided with speed advice, compared to timetable advice ($p=.031$).

Total number

There was no significant difference revealed between experts and novices for the total number of control actions, although there was a significant within-subjects difference evident between DAS types (F(2,24)=6.13, p=.007). Pairwise comparisons indicated that drivers made more control adjustments when provided with speed advice, compared to no driving advice (p=.006), and when following timetable advice (p=.040) (see Figure 3).

Rate per minute

There was no significant difference between experts and novices for the number of control actions made per minute. A significant difference existed between DAS types (F(2,24)=3.59, p=.043), although pairwise comparisons failed to reveal where these differences existed (see Figure 4).

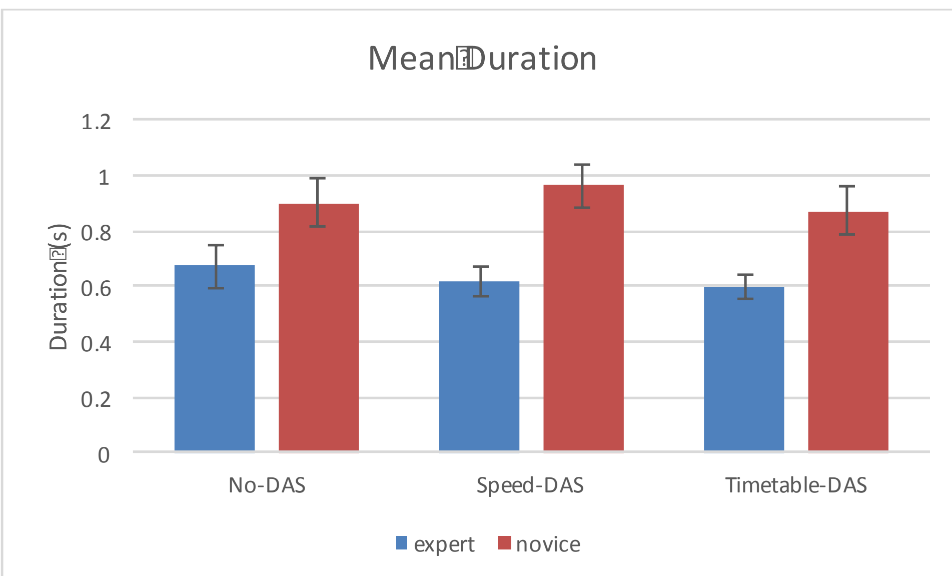

Figure 1: Mean duration of control actions, with standard error bars

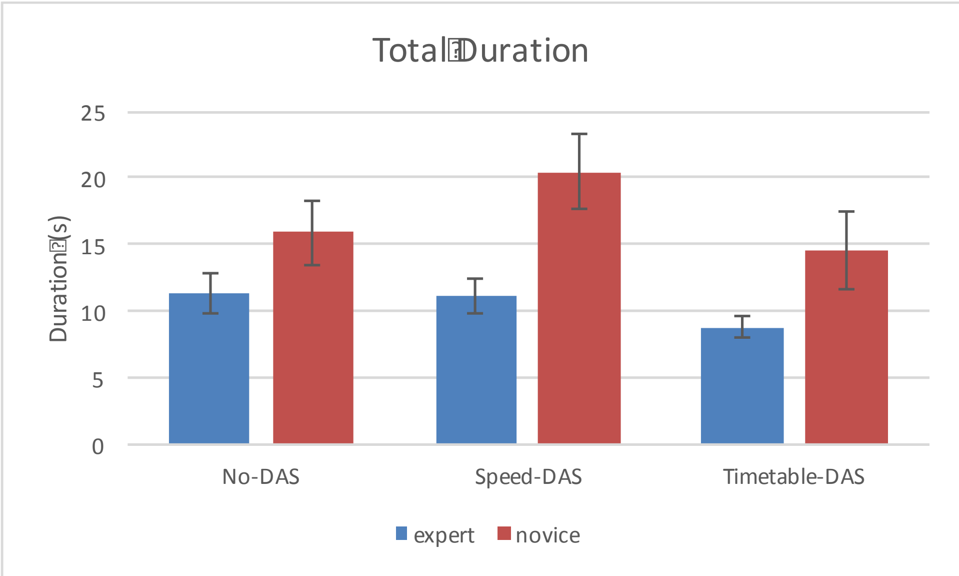

Figure 2: Total duration of control actions, with standard error bars

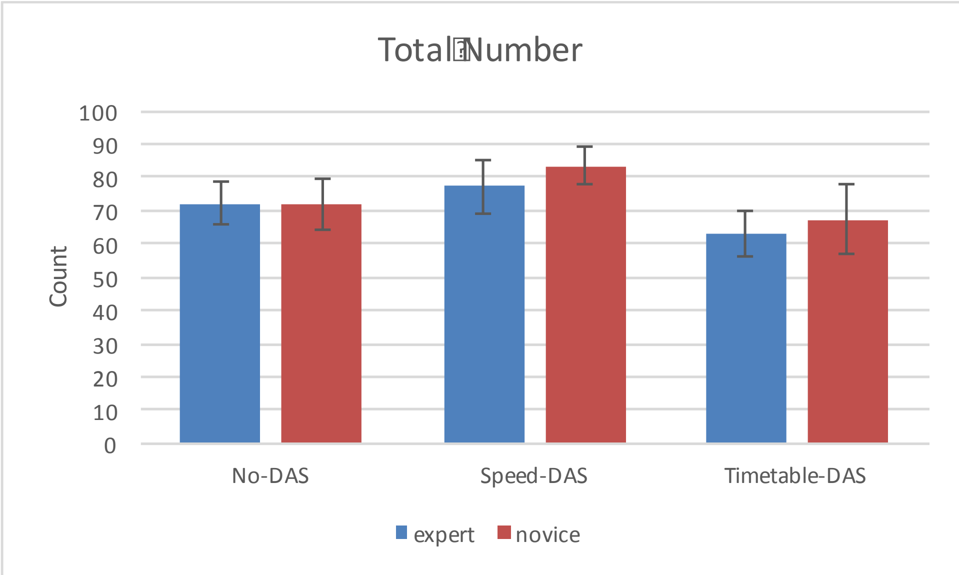

Figure 3: Total number of control actions, with standard error bars

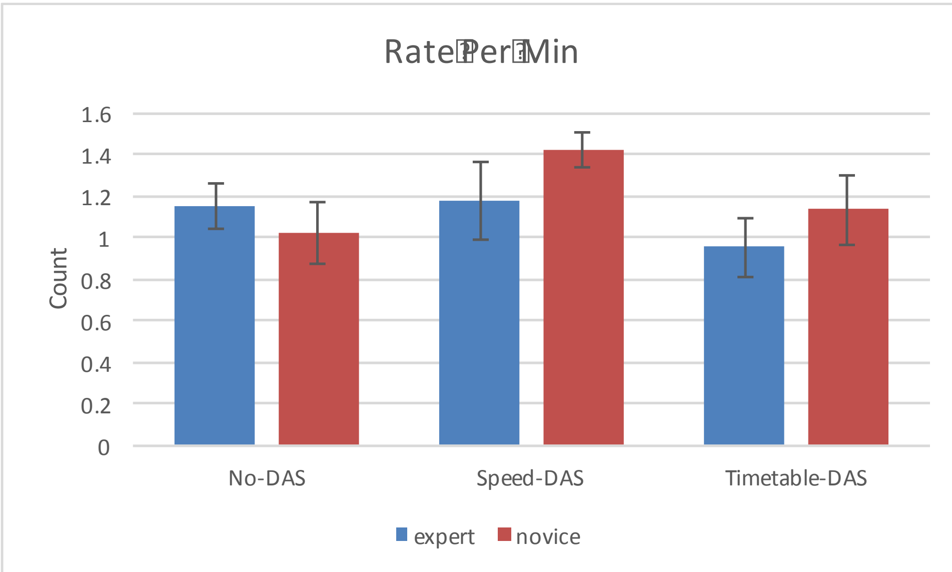

Figure 4: Rate per minute of control actions, with standard error bars

Discussion

Despite previous research (1, 2), suggesting that unskilled drivers exhibit comparable behaviour and performance to experienced, professional train drivers, it is perhaps unsurprising that the further analysis presented here reveals that the way in which this performance is achieved is fundamentally different between groups. Results show that novice drivers deliver far more erratic speed control profiles, revealed by higher values for acceleration noise.

Acceleration noise can be considered as the disturbance of the vehicle's speed from a uniform trajectory or speed. Thus, higher noise is associated with greater speed variations. The analysis revealed that novice drivers displayed significantly higher acceleration noise, overall, than expert drivers in all conditions, suggesting that the manner in which they controlled the train, though sufficient to achieve adequate performance (as determined by speed limit exceedances and SPAD violations in Large et al., (1)), was significantly different to the control imposed by experienced drivers. The data suggest that novice drivers applied cruder control actions (brake and power actuation), resulting in more erratic speed control.

Although novice drivers applied power and/or braking force for longer periods, on average, than experts, the total number of actions was similar between groups, suggesting that novices were slower to select the appropriate magnitude of power or braking force. There was also some evidence that novices made gross adjustments, often locating controllers at full amplitude (i.e. moving from ‘minimum’ to ‘maximum’ rather than selecting an appropriate intermediate value), resulting in frequent episodes of overcorrection and switching between power and brake actuators – this was especially evident at the approach to stations, where some novice drivers slowed too abruptly, coming up short, and were consequently required to re-apply power to reach the platform. This also corresponds with the higher acceleration noise demonstrated by novices.

In contrast, experienced drives made more subtle adjustments and refinements to brake and power actuation, thereby delivering smoother acceleration/braking profiles, and were better able to judge the power and braking force required to achieve target speeds and on the approach to stations. This is consistent with the acceleration noise profile data – lower noise would be expected with more refined speed control – and is also in line with our hypothesis and expectations of real world behaviour. Indeed, in a real world situation, drivers are unlikely to apply power or brake 'full-on', other than under exceptional conditions. This naturally benefits passenger comfort, fuel efficiency etc. (factors that are clearly absent in a simulated driving environment), but also provides additional control capacity for drivers to respond to an emergency situation, should the need arise.

Although periods of control were shorter in duration, on average, for experts, they nevertheless made a similar number of control actions as novices. Consequently, the total duration of control actions made by experts was lower than for novices.

The video analysis also revealed further insights regarding the control actions required to follow the different formats of driving advice. There was no difference in the mean duration of control actions between DAS types, suggesting that drivers applied control in the same manner irrespective of the format of advice presented. However, following speed advice demanded a larger number of control actions, compared to no advice or timetable advice, thereby resulting in both expert and novice drivers applying control for longer periods overall to achieve target speeds. Presenting timing points to drivers, allows them to formulate their own driving strategy to meet these goals, and is therefore likely to be more aligned with drivers' normal expectations. In contrast, speed advice is task-oriented, requiring drivers to achieve specific speed targets throughout the journey, and therefore detaches the control of the train from the overall goals of the journey.

It is also noteworthy that there was more acceleration noise associated with the higher demand condition for all participants, indicating, perhaps unsurprisingly, that additional control actions were required to achieve the scheduled stops associated with the high demand condition, and thus speed fluctuations were greater than during the low demand condition. Data from the original study analysis shows that the higher demand condition was also associated with higher subjective ratings of overall workload (specifically in the absence of DAS). However, there were no significant differences in acceleration noise related to the different DAS types, even though these invited different assessment of subjective workload during the original study. For example, the timetable DAS attracted significantly higher workload ratings than the speed DAS, and in turn, the no-DAS condition (1). This suggests that, although participants perceived different levels of workload associated with each DAS type, they were nevertheless able to control the speed of the train in a similar manner.

It is difficult to draw firm conclusions on the cognitive differences between experts and novices, but the patterns of data found in this study point to an interpretation. Roth and Multer (6) propose a number of elements that together comprise a performance model for train driving. These include aspects such as route knowledge, which would be one area where novices and experts would differ, or another in terms of skills of train handling (e.g. understanding the speed control requirements for the situation but not having the skills in train handling to implement them). However, the speed DAS condition does not require any specific route knowledge, as drivers are always able to drive to the speed target allocated to them without reference to where they might lie along the route. Nonetheless, novice drivers showed more acceleration noise, longer control actions, and a higher rate of control actions. This would suggest that their model of train handling was fundamentally less developed. Of course, route knowledge may also be less developed, and it is a limitation that at this stage, specific components of knowledge cannot be assessed with greater specificity.

Implication for measures of driving behaviour

This work also demonstrates the value of adopting new performance measures as part of train driving simulator methodology. During the original Large et al. study (1), driving performance was determined by exiguous measures (speed limit exceedances and SPADs). It is recognised that other measures could have been employed within the simulated environment (e.g. performance against timetable), and other more comprehensive and rigorous measures clearly exist in the real world (railway capacity utilisation, fuel efficiency, passenger comfort etc.). These measures are notoriously difficult to apply in a simulated environment, and thus we have enhanced the statistical rigour of the original analysis by adopting acceleration noise and control actuation as additional, easy to apply, performance metrics. To the authors' knowledge, there are no examples of acceleration noise being applied in a train-driving context. In contrast, within automotive research, acceleration noise has been applied as a surrogate measure of the effectiveness in the operation of traffic control systems (13), as a predictor of accident risk (9, 14) and to estimate the effects of traffic congestion on fuel consumption and vehicle emissions (15, 16). These factors are equally relevant to train driving, where drivers' speed control and acceleration behaviour are key indicators of safety (prediction of accident/SPAD risk), driving efficiency (fuel economy/energy consumption), passenger comfort (speed uniformity/smoothness of ride) and network capacity utilisation (effectiveness of traffic control systems). The higher levels of acceleration noise associated with the novice drivers (and more demanding driving conditions) revealed during this analysis would naturally elevate these risks and costs in a real world situation, should these drivers be allowed to command a train on a conventional network. However, this would clearly be an ill-advised strategy as novice drivers obviously still lack the highly specialised skills, knowledge of rules and experience required to deal with all manner of events, and we do not recommend it. Nevertheless, it is evident that evaluating train-driving performance using measures such as acceleration noise and control actions, appears to be a sensitive method to distinguish drivers, driving conditions, driving advice etc.

Furthermore, the fact that acceleration noise was elevated during the higher demand condition for *all* drivers, and that the total number and rate-per-minute of control actions were similar between experts and novices, indicates some parity between groups within the simulated driving environment. In essence, the approach one might take when using novices in a simulator study is not to rely on their performance levels as being representative of experts, but to consider their *performance change* in response to factors such as different sources of demand or different forms of information (as in both (1) and (2)). The use of more sensitive measures, such as acceleration noise, within such studies can make any experimental conclusions more robust.

Limitations of the current study include the still relatively low numbers of participants included in the analysis. While these numbers are appropriate for the statistical tests applied in the paper, it would be important to validate these results with a larger sample, and potentially with a different task. The second limitation, more of the method than the study *per se*, is the time intensive manner of the analysis. Low cost simulators, such as the one used in this study, capture only limited control data (in our case, speed). Detailed data regarding control movements can only be extracted through frame-by-frame video analysis, which proved to require 5-10 hours of analysis for every 1 hour of video. This serves as another limiting factor on how many participants can realistically be used in the study. Finally, as noted earlier, while we have offered an early interpretation of how skills or knowledge differ between novices and experts, the method could be developed to be more diagnostic. For example, video analysis could be used to identify whether novices apply control actions later than experts, which would suggest a more reactive style of driving with sub-optimal route knowledge.

Conclusion

The original analysis presented in Large et al. (1) indicated adequate and comparable performance between expert drivers and trained novices. The further analysis presented within this paper suggests that the manner in which that performance was effected (i.e. the control of the train) was fundamentally different between these groups, with novice drivers adopting more erratic speed control profiles, characterised by longer control actions and frequent switching between power and brake actuation. Although the paper therefore exposes some of the limitations of the original analysis, we maintain that the results and conclusions from (1) remain applicable, in the context that they are discussed (i.e. to determine relative performance and behaviour). More generally, these results indicate that simulator-based research utilising expert and/or novice drivers remains of value but may be more perspicuous for relative or formative investigations rather than absolute or summative assessment. Nevertheless, caution should be applied when attempting to draw absolute conclusions or recommendations regarding train handling strategies and techniques, such as speed control or driving performance, when using a cohort comprising primarily or exclusively of non-drivers.

Additionally, this work further demonstrates the utility and value of using acceleration noise and video coding to provide a more sensitive assessment of train driving performance, not only for the task of discriminating between experts and novices, but as a fundamental dependant variable in train simulation. It is hoped that these quantitative measures will prove to be of value to train driving behavioural analysis in the future. One future step, addressing the issue of the time-consuming nature of the data analysis, is to develop software that can both capture and process more sensitive measures in simulation studies.

Acknowledgements

The research was funded by and undertaken as part of EU FP7 ONTIME: Optimal Networks for Train Integration Management across Europe project, Agreement no. 285243.

References

(1) Large, D. R., Golightly, D., Taylor, E.L. 2014. *The effect of driver advisory systems on train driver workload and performance*. Proceedings of Ergonomics and Human Factors Conference (EHF2014)

(2) Dunn, N., & Williamson, A. 2012. Driving monotonous routes in a train simulator: the effect of task demand on driving performance and subjective experience. Ergonomics, 55(9), 997-1008.

(3) Durso, F. T., Bleckley, M. K., & Dattel, A. R. (2006). Does situation awareness add to the validity of cognitive tests?. Human Factors: The Journal of the Human Factors and Ergonomics Society, 48(4), 721-733.

(4) Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. International Journal of Industrial Ergonomics, 39(3), 490-500.

(5) Hart, S. G., & Staveland, L. E. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52, 139-183.

(6) Roth, E., & Multer, J. 2009. Technology implications of a cognitive task analysis for locomotive engineers (No. DOT-VNTSC-FRA-08-06).

(7) Aljanahi, A. A. M., Rhodes, A. H., & Metcalfe, A. V. 1999. *Speed, speed limits and road traffic accidents under free flow conditions*. Accident Analysis & Prevention, 31(1), 161-168

(8) Garber, N.J. and R. Gadiraju. 1989. *Factors affecting speed variance and its influence on accidents*. In Transportation Research Record. Vol.1213, pp. 64-71.

(9) Robertson, D. I., Winnett, M. A., & Herrod, R. T. 1992. *Acceleration signatures*. Traffic engineering and control, 33(9), 485-491.

(10) Herman, R., Montroll, E. W., Potts, R. B., & Rothery, R. W. 1959. *Traffic dynamics: analysis of stability in car following*. Operations research, 7(1), 86-106.

(11) Greenwood, I. D. 2003. A new approach to estimate congestion impacts for highway evaluation—Effects on fuel consumption and vehicle emissions. Ph.D. thesis, University of Auckland, Auckland, New Zealand.

(12) Koppa, R. J., & Hayes, G. G. 1976. *Driver inputs during emergency or extreme vehicle maneuvers*. Human Factors: The Journal of the Human Factors and Ergonomics Society, 18(4), 361-370.

(13) Chung, C. C., & Gartner, N. H. 1973. Acceleration Noise as a Measure of Effectiveness in the Operation of Traffic Control System.

(14) Underwood, G. 2013. On-road behaviour of younger and older novices during the first six months of driving. Accident Analysis & Prevention, 58, 235-243.

(15) Bester, C. J. 1981. *"Fuel consumption of highway traffic."* Ph.D. thesis, Univ. of Pretoria, Pretoria, South Africa.

(16) Greenwood, I. D., Dunn, R. C., & Raine, R. R. 2007. Estimating the effects of traffic congestion on fuel consumption and vehicle emissions based on acceleration noise. Journal of Transportation Engineering, 133(2), 96-104.