Investigation into Effects of System Design on Metro Drivers' Safety-Related Performance: An Eye-Tracking Study

Aleksandrs Rjabovs¹ · Roberto Palacin²

Received: 21 June 2019 / Revised: 7 October 2019 / Accepted: 24 October 2019 / Published online: 15 November 2019
© The Author(s) 2019

Abstract This paper explores performance shaping factors associated with holistic design aspects of metro systems affecting driver's behaviour based on a case study using eye-tracking techniques. Train driving is a highly visual task where the physical environment provides key visual instructions. Eye-mind theories suggest that eye movements closely follow cognitive processes when it comes to interaction with the visual world. The data for this study have been collected in a representative urban rail system serving a metropolitan area with circa 1.2 million inhabitants during 20 in-service trials. Participating drivers were asked to wear an eye-tracker while driving a train as they would normally do. The study focuses on four areas of interest at four consecutive stations. The stations have similar characteristics but differ in terms of certain design elements which are mirrors, platforms, stopping position markers and positions of a running signal. Previous studies have demonstrated the importance of these elements to metro drivers. The paper assesses the usability of the system design elements through analysis of gaze fixations. Results suggest the importance of the positions of a mirror and a signal in relation to the cab, as well as the distance between a stopping point and a signal. Such factors as openness of stations, passenger loadings and informativeness of certain elements also demonstrated influence on

drivers' performance. It was also found that drivers sometimes do not check a signal before departing a station, which is a significant concern. The paper discusses factors potentially propagating such behaviour.

Keywords Eye-tracking · System design · Human factors · Urban rail systems · Safety

1 Introduction

Train driving is a highly visual task. A train driver must scan the approaching physical environment for the presence of hazards, warnings and imposed limits. This visual interaction between the driver and the infrastructure means that the driver should not only see objects but should be able to interpret and interact with them. Understanding whether the physical environment has any factors that make information extraction more complex is an important step in reducing the risk of driver-related incidents in railway industry. This paper studies in-service gaze data of metro drivers interacting with a varied physical environment.

Castro [1] concluded that information on visual behaviour of drivers and subsequent analysis provide a powerful tool for quantifying attentional processes. Patterns of eye-movements are task-specific and “allow diagnostics of a task an observer is trying to perform”. A lot of this is based on work by Yarbus [2] who suggested that human thought process can be traced by analysing eye movements.

Eye-tracking techniques have previously been used on railways. The Rail Safety and Standards Board (RSSB) [3] studied train drivers' visual interaction with mainline signals and found higher complexity of signals at gantries and urban areas. A follow-up RSSB study focused purely on

✉ Roberto Palacin
roberto.palacin@ncl.ac.uk

¹ Network Rail Ltd, Trent House, RTC Business Park, London Road, Derby DE24 8UP, UK

² School of Engineering, Newcastle University, Newcastle upon Tyne NE17RU, UK

Communicated by Marin Marinov.

one multi-Signal Passed at Danger (multi-SPaD) [4]. The study analysed drivers' gaze behaviour qualitatively and did not find any indication of unusual visual behaviour. The RSSB [5] also considered effects of introducing a "Check Automatic Warning System (AWS)" sign on drivers' in-cab gazes and concluded that drivers exposed to such signs check an AWS device more frequently. More recent studies [6] focused on a change in allocation of visual attention with the introduction of in-cab signalling on mainline railways. In terms of personal factors, the RSSB [7] studied attention distribution of maintenance train drivers and how it is affected by factors like geographical knowledge, fatigue, and driving speed. Another study looked at fatigue and associated changes in gaze behaviour for different journey times [8].

Eye-tracking research in railway industry has primarily focused on signal sighting and the driver-machine interface with the signal passed at danger (SPaD) issue in mind. Moreover, most of the research is applicable only to mainline railways due to differences in signalling infrastructure, command and control systems and operational procedures of urban rail systems. This paper aims to address this gap by exploring metro drivers' interactions with some elements of a typical metro station design. Lack of previous research in this area means that there are no benchmark values to which the study's data could be compared. This limits the data analysis to the comparison between the locations selected by the study and the associated reduced sample size.

Nomenclature used in the paper can be seen in Table 1.

Table 1 Nomenclature

SPaD	Signal passed at danger
SASSPaD	Start against signal signal passed at danger
DOO	Driver only operation
PTI	Platform train interface
AOI	Area of interest
TFC	Total fixation count
AFD	Average fixation duration
TFD	Total fixation duration
AWS	Automatic warning system
FOV	Field of view
PSF	Performance shaping factors

2 Methodology

In order to address the aims of this research, a case study was designed to gather critical data related to performance shaping factors associated with the behaviour of drivers and the design characteristics of urban railway systems. The Tyne and Wear (T&W) Metro is a light rail system centred on Newcastle-upon-Tyne, in the north-east of England. Opened in 1980 as a 54 km route (of which 41 km had been adapted from existing heavy-rail tracks), currently the T&W Metro consists of a 78 km network that links the cities of Sunderland and Newcastle with the local airport and coastal regions. This is the second largest urban rail system in the UK and the only one powered by an overhead DC 1500 V supply network. Further details on the T&W network can be found in [9–12].

The in-field eye-tracking experiment was designed as a naturalistic driving study. It measured the relative cognitive impact of certain system design elements, an area which has been identified as requiring in-depth exploration [13]. The eye-tracking experiment provides a quantitative objective portrait of behaviour, whilst the interpretation of the findings is subjective in nature, using experimental observations and semi-structured interviews to generate conclusions. The overall methodological approach is summarised in Fig. 1.

2.1 Trials Design

In-service trials were conducted at four Tyne & Wear Metro stations, namely Pelaw, Heworth, Felling and Gateshead stadium. A train travelling north would arrive at these stations in the listed order. A total of 26 trials were carried out, but only 20 were deemed successful. The discarded trials lacked gaze data upon a visual inspection by the researchers. In each trial only certain timeframes were analysed: 15 s before a full stop at a station and 10 s before a departure from a station. Previous research [11, 12, 14] suggested a relationship between a type of physical environment at a station and drivers' performance, which was selected as an avenue for this research. The arrival timeframes were selected to analyse potential effects of the built environment on selection of a stopping position, or distractions causing a dip in performance. The selected departure timeframes allow an investigation into signal acknowledgement and a subsequent risk of SAS-SPaDs, as well as complexities of interaction with a platform train interface.

Four areas of interest (AOI) selected for this study are driver-only operation (DOO) equipment (a mirror), a platform, a stopping position marker and a running signal

Fig. 1 Methodological approach

at each location. The selected stations provide good variability of the AOI characteristics.

2.1.1 Stations

The stations under investigation fall into two layout types. Pelaw, Felling and Gateshead Stadium are Type 1 stations (shown in Fig. 2) whereas Heworth is Type 2 (shown in Fig. 3). The main differences between the two station types are the location of the platform, the stopping position marker and the mirror in relation to the driver's cab. It is important to note that in the Tyne & Wear Metro the driver's cab takes only half the width of a car. Furthermore, the selected stations differ in terms of passengers visibility, patronage levels and distance between the stopping point

and the running signal (Table 2). All of the Type 1 stations in the study are overground, however Gateshead Stadium is built over towards the end of the platform where the stopping position is. With a retaining wall to the left of the train at that location, it creates an effect of a built-over station when observed from the cab of a stationary train. Heworth is fully built-over, hence considered underground. The stations also differ in terms of passenger visibility on the approaches to a platform.

2.1.2 Metrics

At the time of the study, an eye-tracking body of knowledge was almost non-existent in the railway industry, hence the authors used the literature describing the experience from other industries in selection of metrics. Following this approach, three metrics were selected—total fixation duration (TFD), total fixation count (TFC) and average fixation duration (AFD). Holmqvist et al. [15] defines a fixation as “a time an eye remains still over a period of time” (p. 21). The metrics related to fixation count and duration are the most popular metrics used in modern studies [15–17], providing information on cognitive processes. In terms of equipment, Tobii® Technologies' Tobii Glasses 1 Eye-Tracker (first generation) was used in the trials. The eye-tracking set consists of eye-tracking glasses and a recording device. The set has 30 Hz sampling frequency and uses a 9-point calibration algorithm which is performed before each trial. Raw data were analysed in Tobii Studio software using the Tobii I-VT fixation filter.

Number of fixations, or total fixation count (TFC), is the sum of fixations recorded on the area of interest (AOI) in a set period. High number of fixations can suggest semantic importance or informativeness of the AOI [2, 16, 18, 19]. High TFC can also be a sign of difficulty in interpreting information [15, 20], complexity [21] and poor search

Fig. 2 Typical Type 1 design station

Fig. 3 Typical Type 2 design station **Table 2** Summary of station characteristics

Station name	Station Type	Passengers visibility on the approach	Approximate distance between a stopping point and a signal (m)	Patronage levels
Pelaw	Type 1	Good	25	Medium–high
Heworth	Type 2	Poor	35	High
Felling	Type 1	Average	25	Low–medium
Gateshead Stadium	Type 2	Average	5	Low–medium

efficiency [22]. For the purpose of this study, the TFC results are provided as an average for all the drivers included in the sample.

Longer AFD is a sign of deeper processing [15], the criticality of the elements [23, 24], a sign of issues with information extraction [15]. At the same time, shorter fixations also indicate issues with information extraction due to higher mental workload and stress [25–27] and usability problems. Holmqvist et al. [15] make a distinction between higher workload associated with higher AFD (which a human can complete without problems) and with lower AFD (causing performance issues due to additional stress).

Although the study also investigates total fixation durations to understand priorities set by drivers, it is the relationship between TFC and AFD metrics that should suggest a difference in usability of similar elements. According to Kapitaniak, et al. [17] “overall increase in requirements and complexity of the task tends to reduce

AFD and increase TFC” (p. 950). Other studies [28–30] have demonstrated that higher sampling rate (high TFC and low AFD) is a sign of an increase in cognitive demands, anxiety and high mental workload.

2.1.3 Participants

Four drivers participated in the trials, creating a sample of 20 in-service runs. The participants were selected in such a way as to eliminate as many uncertainties as possible, e.g. factors such as various health conditions and dysfunctions [31], and age [32] can affect fixation patterns. All of these factors can affect the quality of the data collected [14] and were also taken into account during the selection process. In addition, the participants were drivers responsible for driver training and were selected for their experience. This was done to avoid differences in gaze behaviour based on experience and route knowledge. As a result, all

participants had similar experience levels, ethnical background, gender and good eyesight.

The trials started at three different terminus stations, where a 9-point equipment calibration was carried out. The variety of starting points means that drivers arrived at the stations under investigation at different times after the start of each trial. The effect of different starting stations was explored by excluding trials from each terminus and running the analysis again. It was concluded that starting points did not influence the results, hence all of the 20 trials were retained.

The trials were performed in a variety of weather conditions covering three categories: rainy, cloudy and sunny, as well as at different times of day. Twenty out of 26 trials were considered successful upon preliminary inspection of the recordings. Factors that influenced exclusion of trials were direct sunshine and drivers adjusting position of the eye-tracker (without re-calibration) during the experiment.

2.1.4 Analysis Approach

The eye-tracking data for each AOI, e.g. a mirror, was compared for all of the stations. Descriptive statistics were obtained in terms of histograms and explored for differences. Trends for the metrics were studied for mean values for each of the metrics.

The statistical tests used depended on the normality of the data. If the data were normally distributed, then a paired sample *T* test was performed, whereas a Wilcoxon signed ranks test was used to compare the non-normally distributed data. The statistically significant (*p* value of 0.05 or less) results were considered as the only true indication of difference. However, the results significant at 90% confidence level (*p* value of 0.1 and less) were also highlighted, assuming such results might become truly statistically significant in a larger sample. This assumption is based on previous studies by Aitchinson and Davies [33], Moore et al. [34], Tripathi and Borrión [35].

The data were also checked for relationships between eye gaze performance and personal factors (comparing

drivers), and weather (comparing weather during a trial) using Kruskal–Wallis *H* test.

3 Results and Discussion

3.1 Arrivals

Table 3 summarises descriptive statistics for each of the AOIs in terms of mean values for arrivals. Mirrors received the largest amount of visual attention at all of the stations but Heworth, where the position of the mirror is such that it can be obscured by passengers until later in the arrival sequence. The statistics suggests that drivers' visual attention is highly focused on selecting a correct stopping point that would allow safer dispatch later. This is based on mirrors and stopping position indicators getting more than 50% of cumulative TFD for all four AOIs at each location. It is important to note that stations that are more “built-up” (Heworth and Gateshead stadium) have drivers spending less time looking somewhere else but the four AOIs. It is possible to see that the majority of this increase in TFD at Heworth is driven by more and longer gazes at the platform. The station has the highest patronage figures among the locations in the study, so one can suggest that passenger numbers affect drivers' visual interaction with the physical environment.

For “open” layout stations (Pelaw and Felling), a stopping position marker was the second most looked at AOI. According to GI/RT7033 [36], the recommended dimensions for platform stop markers are only 300 × 250 mm, with visibility requirement of only 2 s. The results show the metro drivers fixating on all the markers for more than 2 s, and those being visible significantly earlier. This shows an important difference between conventional mainline railways and DOO systems, where it is necessary to provide drivers with increased flexibility of stopping positions. Moreover, when the DOO equipment is used as a stopping position benchmark, 2-s visibility is too small, as it does not factor in issues of divided attention between mirrors and stopping position indicators.

Table 3 Descriptive statistics for station arrivals (data related to mean values of 20 samples)

Stations	Mirror			Platform			Signal			Stopping position		
	TFD (s)	TFC ^a	AFD (s)	TFD (s)	TFC ^a	AFD (s)	TFD (s)	TFC ^a	AFD (s)	TFD (s)	TFC ^a	AFD (s)
Pelaw	5.19	5.8	0.97	2.31	3.4	0.58	0.37	0.75	0.19	2.43	2.7	0.89
Heworth	3.99	4.35	1.08	4.88	7.3	0.73	0.33	0.9	0.20	2.63	2.5	0.90
Felling	5.07	5.45	0.98	1.71	2.3	0.66	0.42	0.8	0.21	2.47	2.65	0.92
Gateshead Stadium	5.37	5.4	1.17	3.28	5.15	0.62	0.2	0.55	0.13	2.09	2.85	0.58

^aMeasured as average number of occurrences across the whole study

The participants fixated the least on running signals, which was unexpected as the drivers clearly indicated checking the signals on arrival in the questionnaires [14]. This discrepancy indicates that the drivers might have been trying to answer the questionnaire in a way they thought to be correct. Morrel-Samuels [37] claims that such behaviour is not unique and favourable answer skewness can be a sign of anonymity concerns. However, in a more demanding environment of an operational railway, checking something which is not required can be skipped to address more critical tasks. It was expected that drivers would fixate more on the Gateshead Stadium signal due to its proximity to the driver's cab. However, the further a signal was from the stopping position, the more fixation it received. It is possible that the closest signals were checked by side gazes that were not being registered by the eye-tracker.

The two “open” stations (Pelaw and Felling) show comparatively poorer usability of mirrors in terms of higher TFC and lower AFD. It is possible that environmental factors exerted much higher influence on the usability of the mirrors at such locations. This is discussed in Sect. 3.3. Another performance shaping factor (PSF) at Pelaw and Felling was a slightly greater distance from the stopping point to the mirror compared to the other locations. With all mirrors being the same size, it can be assumed that a shorter distance positively affects usability.

As drivers constantly switch their attention between the stopping position markers and the mirrors on arrival, the usability of one should affect another. It was found that where informativeness of one of the two elements was relatively poor, it positively affected usability of the other element. For example, the mirror at Heworth can be obstructed by passengers early in the arrival sequence. This in turn leads to much easier interaction with the stopping position indicator at that station. A similar relationship can be observed for the mirror at Gateshead Stadium, where the stopping position indicator has very low informativeness due to a sub-optimum position (as described by the participating drivers who all stop their trains a few meters past the indicator to be able to better view the mirror). When both elements are informative it can lead to a divided attention scenario. Such a scenario is known to significantly increase stress and workload [38, 39], especially when considering more than two concurrent monitoring tasks whilst stopping the train at a station. Even though Basacik and Gibson [40] claim that station stop associated tasks can cause underload, neither of their “underload” scenarios is applicable to DOO systems.

In terms of platforms, only Pelaw demonstrated signs of poor usability. These are based on only a marginal difference in AFD with Felling, the other “open station” or, indeed, with Gateshead Stadium. The literature does not specify whether it is necessary to compare TFC/AFD ratios

in situations when one of the metrics produces the same values. However, both metrics can report the same in certain scenarios. For example, high TFC can be a sign of difficulty in information extraction [15, 20] but is more controversial when not supported by a low AFD value. One factor which might have affected the results at Pelaw is the location of the waiting area and the subsequent higher concentration of passengers towards the back of the train. Concentration of passengers in only one part of a platform is a known risk which can be caused by station design [41]. At Pelaw, this leads to the drivers needing to scan large volumes of passengers at higher speeds (early in the arrival sequence). Fitzpatrick et al. [42] showed that multi-tasking at higher speeds causes a decline in car drivers' performance. However, this is applicable to the situation at Pelaw only if observing each individual passenger is considered as a separate task.

The relationship between the signals at Heworth and Felling needs to be considered with caution due to the factors described above. If the signal at Felling is in fact easier to interact with than the Heworth signal, the distance between stopping position and signal becomes an important PSF. However, the adverse effects are not directly proportional to the distance, and become pronounced only after a certain distance. Based on other locations, it is possible to note that distances in excess of 25 m (Felling) are where this process starts.

3.2 Departures

Table 4 contains descriptive statistics for each of the AOIs in terms of mean values for departures. Similarly to the results in Sect. 3.1, more confined stations like Gateshead Stadium and Heworth have drivers focusing on AOIs for much longer compared to “open” stations. Longer fixations on the mirrors at Gateshead Stadium and Heworth are potentially linked to a closer proximity of such assets to the stopping positions. Drivers pay significantly more attention to mirrors than signals. Signals could be considered a static element that, in most cases, should not change during a station stop.

A field of view (FOV) angle, the angle between mirror, cab and signal (example in Fig. 4), is strongly related to total visual attention allocated to the two AOIs. Stations with a wider angle, Pelaw and Gateshead Stadium, demonstrate a drop in TFD for signals. This suggests that drivers are willing to sacrifice signal checks, to some extent, in layouts that require longer saccadic movements. In comparison to the indicator-mirror relationship on arrivals, the two AOIs on departures are both safety-critical and, in theory, none can be skipped. Hence the issue of divided attention is further prompted by certain layouts and PSFs. This can manifest itself in more stressful interaction

Table 4 Descriptive statistics for station departures (data related to mean values of 20 samples)

Stations	Mirror			Signal		
	TFD (s)	TFC ^a	AFD (s)	TFD (s)	TFC ^a	AFD (s)
Pelaw	5.37	6.85	0.77	0.93	1.95	0.33
Heworth	6.26	7.4	1.32	1.58	3.35	0.36
Felling	4.85	5.65	0.82	1.49	2.85	0.39
Gateshead Stadium	5.66	6.2	0.99	1.15	2.35	0.46

^aMeasured as average number of occurrences across the whole study **Fig. 4** Field of view (FOV) angle

with one of the AOIs, as demonstrated by the mirror at Pelaw.

Importance of the FOV angle is further corroborated at Pelaw where the mirror once again shows signs of poor usability through the TFC/AFD analysis. Although it is possible that environmental factors could influence this, Felling, another “open” station in the study, demonstrates better results. This could be due to the mirror at Felling being located further down the platform compared to Pelaw, thus creating a narrower FOV.

The results suggest usability issues with the signal at Heworth. It is located significantly further away from the stopping position compared to other locations. This suggests negative correlation between usability and distance to a signal.

One very important finding in the trials was numerous failures to check a signal during a departure. SASSPaDs are rather common in the UK railway industry, with RSSB [6] reporting 39 occurrences with passenger trains between 2010 and 2013. Moreover, SASSPaD risk is higher in DOO systems according to Basacik and Gibson [40]. As presented by the RSSB in the Precursor Incident Model [8],

many of driver-related accidents stem from smaller incidents acting as precursors. In this particular situation, failures to check a signal aspect before departure can lead to SASSPaD under certain circumstances.

In total 7 occasions of no fixations at a signal were registered during the 20 trials, with all the participants demonstrating at least one violation. The occurrences were not distributed evenly: Pelaw—three, Felling and Heworth—two, Gateshead Stadium—zero. All of the stations with such violations are at locations where signals are located at least 25 metres away from a starting position. It takes 7–10 s to cover 25–35 m, respectively, from a complete stop. Hence the drivers, especially experienced drivers, might believe that they have time to check a signal and react even if they do it only after departure. However, Multer et al. [43] indicate that very responsive controls of electric trains mean more in-cab gazes for speed information monitoring and subsequent higher risk of SPaDs. It is important to note that there is no direct correlation between distance to a signal and the number of violations. It is possible that a combination of other PSFs, such as FOV angle (widest at Pelaw out of the 3 locations with violations) and high passenger loadings, also contribute to the results.

Tripathi and Borrión [35] discovered that emphasising on-time performance affects metro drivers’ safety and security related performance. It also means that other competing goals can have similar effects on metro drivers’ performance. The RSSB [5] claim that drivers are highly motivated to check signals on departure and non-compliance is highly unlikely. On the other hand, another RSSB report highlights a risk of drivers departing a station without fully checking that all doors are closed due to time pressures [40]. In DOO systems, where most risks are related to the PTI, the drivers might prioritise DOO equipment and skip signals due to other pressures. This demonstrates why the same findings cannot always be applied to mainline railways and metro systems, as a combination of design features and risk profiles can significantly affect drivers’ priorities.

It is known for people in general that visual attention can precede fixations by about 0.25 s [15, 44]. Hence if

Table 5 Summary of importance of personal and environmental factors by metric, location and AOI ($p < 0.05$ only)

	Personal factors		Environmental factors	
	Arrival (%)	Departure (%)	Arrival (%)	Departure (%)
All metrics	54	29	2	25
TFD	69	38	0	25
TFC	56	13	4	25
AFD	44	38	0	25
Mirror	50	58	0	17
Platform	67		0	
Signal	25	0	0	33
Stopping position	75		8	
Pelaw	75	0	0	17
Heworth	58	33	0	17
Felling	42	50	8	33
Gateshead Stadium	42	33	0	33

participants wanted to fixate on a signal right before departure, the actual fixation could have happened after a train started moving (not included in the analysis). Considering that there is a slight lag between applying power and start of motion, it is possible to claim that the lag between visual attention and fixations is negligible for this experiment. Most importantly, whatever the reason for equipment not registering fixation on a signal, even if one of these cases is a genuine violation or a lapse, this lag creates a significant risk factor for metro operations.

3.3 Personal and Environmental Factors

Results of the Kruskal–Wallis H test for the dependence of the collected data on environmental and personal factors are summarised in Table 5. Personal factors are found to have more effect on gaze behaviour than environmental factors, especially at arrivals. Dependence on environmental factors is more pronounced on departures compared to arrivals. The results corroborate previous research emphasising the high importance of personal factors [45, 46]. AFD and TFC metrics demonstrate significantly lower percentages of statistically significant relationships with the personal factors than the TFD values. This is a positive result for the methodology taking into account concerns about assessment methodology of stress-inducing elements. However, it is important to note that different drivers did different numbers of trial runs, which could have affected some of the results. Moreover, passenger levels encountered could be very different, as this not only depends on the time of a trial but also the presence of peak services (relieving passenger congestion) ahead and disruptions in the system.

The participants had different strategies in looking on the most dynamic AOIs—platforms on arrival and mirrors

on departure. Varied passenger levels might have contributed to this. This is supported by the fact that during arrivals the variability is mostly observed at the two most crowded stations—Pelaw and Heworth.

The environmental factors' influence was rather low, implying that sun is not the contributory factor at Pelaw for the mirror during departures. However, the drivers' in their questionnaire assessment of contributory factors stated that the direct sunlight is the issue, not sun in general [14]. As the trials were conducted at different times and in different months, it is impossible to say in retrospect whether some of the trials had that issue or not. On the other hand, a measure for wind was not recorded. The issue of “shaky” mirrors was raised by drivers during the semi-structured interviews [14]. As it can be very windy in Tyne & Wear during any type of weather, the wind should be considered a major contributory factor to poor usability of the mirrors at the open stations.

3.4 Statistical Significance of the Discussed Relationships

Sections 3.1 and 3.2 contain discussion based on the analysis of descriptive statistics. When the data are analysed using paired samples T test or Wilcoxon signed ranks test, a number of differences were found to be not statistically significant, for example, the majority differences in metrics for signals or stopping position indicators at arrivals. The results were double checked using sign test which corroborated the previous results. Table 6 below illustrates this analysis for the arrivals case.

It is possible that the sample size is simply not big enough. This is supported by further statistically significant relationships found when expanding the p value. If differences with $p < 0.1$ are taken into account, 25% of arrival

Table 6 Difference in gaze behaviour at different locations (arrivals)

	Metric								
	TFD			TFC			AFD		
	Pelaw	Heworth	Felling	Pelaw	Heworth	Felling	Pelaw	Heworth	Felling
<i>Mirror</i>									
Heworth	– 1.979**			– 2.235**			– 1.008		
Felling	– 0.093	– 1.717*		– 0.699	– 1.463		– 0.243	– 0.224	
Gateshead Stadium	– 0.448	– 1.867*	– 0.635	– 0.549	– 1.225	– 0.048	– 1.941*	– 0.597	– 1.792*
<i>Platform</i>									
Heworth	– 2.856**			– 2.639**			– 1.419		
Felling	– 1.494	– 3.173**		– 1.820*	– 3.685**		– 0.560	– 0.560	
Gateshead Stadium	– 2.501**	– 1.568	– 2.333**	– 2.803**	– 1.431	– 2.991**	– 0.597	– 1.456	– 0.560
<i>Signal</i>									
Heworth	– 0.157			– 0.206			– 0.235		
Felling	– 0.078	– 0.157		– 0.420	– 0.233		– 0.471	– 0.408	
Gateshead Stadium	– 1.479	– 0.756	– 0.863	– 0.863	– 0.511	– 1.026	– 1.071	– 0.979	– 1.02
<i>Stopping Position</i>									
Heworth	– 0.871			– 0.477			– 0.0457		
Felling	– 0.149	– 0.037		– 0.263	– 0.177		– 0.336	– 0.149	
Gateshead Stadium	– 0.610	– 1.112	– 1.408	– 0.257	– 0.810	– 0.317	– 1.917*	– 2.343**	– 1.730*

* $p < 0.1$, ** $p < 0.05$

relationships would be found statistically significant. It cannot be expected that all of the AOIs show differences in all metrics. Furthermore, it cannot be expected that all metrics will be significantly different.

The selected eye-tracking device potentially created a lot of deviation in the results, making it harder to find statistical significance. This could have been caused by the necessity to explore a dynamic physical environment using static AOIs, low sampling rate and lower sampling measures in some trials. The exploratory nature of this experiment showed the areas that need to be investigated, even if those were not highlighted as statistically significant at all locations. However, lack of simultaneous statistically significant differences in both AFD and TFC variables means that all assumptions about increased workload/stress are purely based on descriptive statistics.

4 Conclusions

This paper has shown that the eye-tracking research can be conducted safely in metro systems. It allows building a basis of the exploratory research in front-line staff interaction with the physical environment. Moreover, the more sophisticated eye-tracking tools available now and

advancements in experiment methodologies should allow establishing baseline eye fixation values in the future. Such baseline values would facilitate assessing the infrastructure and human operators in a non-intrusive way.

Metro drivers clearly prioritise mirrors over any other elements of a station design, suggesting a higher semantic value of PTI risks to them compared to SPaD risks. Data suggest a division of visual attention between mirrors and stopping position indicators on arrivals. When one of the elements in this relationship exhibits poor informativeness, the issue of divided attention is not as acute as the driver's focus on the second available AOI. Gaze duration on platforms is dependent on passenger levels and how "open" a station is.

In terms of workload and stress inducing PSFs, several design features are suggested by the data. Firstly, a wide angle between mirror, driver and signal (FOV angle) creates situations in which the drivers need to produce longer saccadic movements to shift their attention from one AOI to another. Secondly, the distance between signal and stopping position affects drivers' interactions with the signals. Thirdly, the passenger levels and distribution on a platform can create additional stress for drivers during both arrivals and departures. However, the analysis shows that different drivers approach complications caused by

crowding differently. Despite almost negligible influence of weather (sunny or rainy) on drivers' performance, other environmental factors, e.g. wind or direct sunlight, can still affect drivers' visual interactions with the physical environment. This is demonstrated by differences between "open" and built-over stations as well as drivers' perceptions.

The biggest concerns are related to drivers potentially not checking a signal before powering up the train. This is a serious precursor in SaSSPaDs and should be addressed. Even though the data could have been affected by quality issues, the absolute number is too high to be disregarded. In fact, even one occasion in 80 station departures is alarming, as all of the participants are very experienced drivers. With many tasks competing for metro drivers' visual attention in DOO systems, the drivers prioritise monitoring of the PTI over mitigation of SPaD risk. This is significantly different from conventional mainline railways, where a train guard or platform staff ensure safe dispatch of a train.

The combination of a distance to a signal, FOV angle, and passenger loadings seems to affect the propagation of such violations. Metro drivers have up to 10 s between powering up and passing a signal, which can create a perception that there is always time to mitigate a risk of SPaD. Moreover, there is a possibility that such perception is developed by more experienced staff with greater understanding of such technicalities of the system. Finally, while the relatively small size of the sample might have an effect in the specific values analysed in this paper (e.g. Table 6), the overall conclusions extracted are, nevertheless, considered valid.

Acknowledgements The authors would like to thank the Tyne & Wear Metro for the opportunity to carry out this research and the access to the relevant data.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Castro C (2009) Human factors of visual and cognitive performance in driving. CRC Press, Boca Raton
2. Yarbus AL (1967) Eye movements and vision. Plenum Press, New York
3. Rail Safety and Standards Board (RSSB) (2005) T152—Train driver visual strategies. RSSB, London
4. Rail Safety and Standards Board (RSSB) (2006) Development of SPAD risk management tools—summary report. RSSB, London
5. Rail Safety and Standards Board (RSSB) (2007) Understanding multi-SPAD signals and the effects of line identifiers on driver behaviour: an investigation into multi-SPAD signal D301 (T632). RSSB, London
6. Rail Safety and Standards Board (RSSB) (2013) SPAD and TPWS activity report. Quarter 2—2013/14. RSSB, London
7. Rail Safety and Standards Board (RSSB) (2008) Understanding human factors: a guide of the railway industry. RSSB, London
8. Rail Safety and Standards Board (RSSB) (2011) Precursor indicator model (PIM) for train accident risk. RSSB, London
9. Howard DF (1976) Tyne and Wear metro—a modern rapid transit system. *Inst Mech Eng (Lond) Proc* 190(18):121–136
10. Prickett BR (1981) Electrification of the Tyne and Wear metro. *IEE Proc B Electric Power Appl* 128(2):81–91. <https://doi.org/10.1049/ip-b.1981.0012>
11. Rjabovs A, Palacin R (2015) Attitudes of metro drivers towards design of immediate physical environment and system layout. *Urban Rail Transit* 1(2):104–111. <https://doi.org/10.1007/s40864-015-0020-y>
12. Rjabovs A, Palacin R (2017) The influence of system design-related factors on the safety performance of metro drivers. *Proc Inst Mech Eng Part F J Rail Rapid Transit* 231(3):317–328. <https://doi.org/10.1177/0954409716630007>
13. Naweed A, Balakrishnan G (2014) Understanding the visual skills and strategies of train drivers in the urban rail environment. *Work* 47(3):339–352
14. Rjabovs A (2017) Performance shaping factors affecting driver safety-related behaviour in urban rail systems: Tyne & Wear Metro case. Newcastle University, Newcastle-upon-Tyne
15. Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J (2011) Eye tracking: a comprehensive guide to methods and measures. OUP, Oxford
16. Jacob RJ, Karn KS (2003) Eye tracking in human–computer interaction and usability research: ready to deliver the promises. In: Radach R, Hyona J, Deubel H (eds) *The mind's eye*. Elsevier, Amsterdam, pp 573–605
17. Kapitaniak B, Walczak M, Kosobudzki M, Jozwiak Z, Bortkiewicz A (2015) Application of eye-tracking in the testing of drivers: a review of research. *Int J Occup Med Environ Health* 28(6):941–954
18. Henderson JM, Weeks PA Jr, Hollingworth A (1999) The effects of semantic consistency on eye movements during complex scene viewing. *J Exp Psychol Hum Percept Perform* 25(1):210
19. Poole A, Ball LJ (2006) Eye tracking in HCI and usability research. In: Ghaoui C (ed) *Encyclopedia of human computer interaction*. IGI Global, Hershey, pp 211–219
20. Ehmke C, Wilson S (2007) Identifying web usability problems from eye-tracking data. In: *Proceedings of the 21st British HCI Group annual conference on people and computers: HCI... but not as we know it—volume 1*. British Computer Society, London, pp 119–128
21. Dzung R-J, Lin C-T, Fang Y-C (2016) Using eye-tracker to compare search patterns between experienced and novice workers for site hazard identification. *Saf Sci* 82:56–67
22. Goldberg JH, Kotval XP (1999) Computer interface evaluation using eye movements: methods and constructs. *Int J Ind Ergon* 24(6):631–645
23. Harris RL, Sr, Christliff DM (1980) What do pilots see in displays? In: *Proceedings of the human factors society annual meeting*, vol 24, No. 1. Sage Publications, Los Angeles, pp 22–26
24. Shinar D, McDowell ED, Rockwell TH (1977) Eye movements in curve negotiation. *Hum Factors* 19(1):63–71
25. Miura T (1990) Active function of eye movement and useful field of view in a realistic setting. In: Groner R, d'Ydewalle G, Parham R (eds) *Studies in visual information processing*, vol 1. From eye to mind: Information acquisition in perception, search, and reading. North-Holland, Oxford, pp 119–127

26. Unema P, Rotting M (1990) Differences in eye movements and mental workload between experienced and inexperienced motor-vehicle drivers. In: Brogan D (ed) *Visual search*. Taylor & Francis, London, pp 193–202
27. Van Orden KF, Limbert W, Makeig S, Jung T-P (2001) Eye activity correlates of workload during a visuospatial memory task. *Hum Factors* 43(1):111–121
28. Beck AT, Emery G, Greenberg RL (2005) *Anxiety disorders and phobias: a cognitive perspective*. Basic Books, New York
29. Crundall DE, Underwood G, Chapman PR (1998) How much do novice drivers see? The effects of demand on visual search strategies in novice and experienced drivers. In: Underwood G (ed) *Eye guidance in reading and scene perception*. Elsevier, Amsterdam, pp 395–417
30. Schulz CM et al (2010) Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment. *Br J Anaesth* 106(1):44–50
31. Coeckelbergh TR, Cornelissen FW, Brouwer WH, Kooijman AC (2002) The effect of visual field defects on eye movements and practical fitness to drive. *Vis Res* 42(5):669–677
32. Ho G, Scialfa CT, Caird JK, Graw T (2001) Visual search for traffic signs: the effects of clutter, luminance, and aging. *Hum Factors* 43(2):194–207
33. Aitchinson J, Davies K (2009) *Drivers' response to junction signalling: a simulator study (T669 Report)*. London
34. Moore DS, McCabe GP, Craig BA (2009) *Introduction to the practice of statistics*, 6th edn. W.H. Freeman, New York
35. Tripathi K, Borrión H (2016) Safe, secure or punctual? A simulator study of train driver response to reports of explosives on a metro train. *Security Journal* 29(1):87–105
36. GI/RT7033 *Lineside Operational Safety Signs*, Railway Group Standards, London, 2009
37. Morrel-Samuels P (2002) Getting the truth into workplace surveys. *Harv Bus Rev* 80(2):111–118
38. Davies DR, Matthews G, Stammers RB, Westerman SJ (2013) *Human performance: cognition, stress and individual differences*. Psychology Press, London
39. Prottengeier J et al (2016) The effect of a standardised source of divided attention in airway management: a randomised, cross-over, interventional manikin study. *Eur J Anaesthesiol (EJA)* 33(3):195–203
40. Basacik D, Gibson H (2015) Where is the platform? Wrong side door release at train stations. Presented at the Contemporary Ergonomics and Human Factors 2015
41. Cynk S et al (2015) Qualitative study of platform-train interface incidents. In: Transport Research Laboratory (TRL) (ed) *PTI strategy technical report*. London
42. Fitzpatrick K, Chrysler S, Park ES, Nelson A, Robertson J, Iravavarapu V (2010) *Driver workload at higher speeds*. Texas Transportation Institute, College Station
43. Multer J, Safar H, Roth E (2015) Understanding how signals passed at danger occur through the lens of sociotechnical systems. In: 5th International rail human factors meeting (London, UK)
44. Deubel H (2008) The time course of presaccadic attention shifts. *Psychol Res* 72(6):630
45. Gibson H (2016) Industry human factors SPAD review-project summary report. Rail Safety and Standards Board (RSSB). <https://www.sparkrail.org/Lists/Records/DispForm.aspx>. Accessed 1 June 2019
46. Li G, Hamilton I (2003) *Railway signal sighting: scientific evidence for sighting principles*. Rail Safety and Standards Board, London