Graphical Abstract

Evaluating Train Drivers' Performance to Inform the Development of Automatic Train Operation

Baris Cogan, Esther Bosch, Birte Thomas-Friedrich, Helena Wasle, David Schackmann, Christian Klotz, Birgit Milius

Highlights

Evaluating Train Drivers' Performance to Inform the Development of Automatic Train Operation

Baris Cogan, Esther Bosch, Birte Thomas-Friedrich, Helena Wasle, David Schackmann, Christian Klotz, Birgit Milius

- Train drivers' performance can be used as a benchmark for Automated Train Operation (ATO) systems
- We tested train driver's reaction times to objects of varying size and contrast on track
- Train cab simulator experiments (n = 43) revealed significant effects of object size, contrast, and speed on reaction times
- This study presents a first step towards understanding performance shaping factors that can serve as a basis for defining the requirements for ATO systems

Evaluating Train Drivers' Performance to Inform the Development of Automatic Train Operation

Baris Cogan[a,*], Esther Bosch[b,*], Birte Thomas-Friedrich[b], Helena Wasle[b], David Schackmann[b], Christian Klotz[c], Birgit Milius[a]

[a] *Technische Universitaet Berlin, Berlin, Germany*
[b] *German Aerospace Center, Institute for Transportation Systems, Braunschweig, Germany*
[c] *German Center for Rail Traffic Research, Federal Railway Authority, Bonn, Germany*

Abstract

Understanding train driver performance can provide valuable insights for the development of automatic train operation systems. This study investigates the visual perception of train drivers under different conditions using driving simulator experiments. The 43 participating train drivers were instructed to drive the train and react to stationary objects on the tracks of varying size and contrast to the background. Two train protection systems (the German intermittent train protection system PZB and the European Train Control System with in-cab signalling ETCS) and on-sight driving were used. The results showed significant effects of size, contrast, and speed on reaction times. The effects of the train protection systems and on-sight driving were inconclusive. The approach presented in this study, along with an understanding of the relative impact of various performance shaping factors can serve as a basis for defining the requirements for ATO systems.

*Joint first authors and corrosponding authors: baris.cogan@tu-berlin.de , Esther.Bosch@dlr.de

Keywords: train drivers, visual perception, ATO

PACS: 0000, 1111

2000 MSC: 0000, 1111

1. Introduction

Digitalization and automation are driving fundamental changes in transportation systems. The introduction of Automated Train Operation (ATO) will transform railway operations across various levels of implementation. These levels of implementation, classified as Grades of Automation (GoA) in urban transit, have also gained traction in mainline railways [1]. GoA1 describes train driving without automation. From GoA2 onwards, different levels of automation are present. In GoA2, the safety responsibility remains with the train drivers as they remain in the cab and supervise the system driving the train automatically. GoA3/4 is defined as the train driving automatically with no train driver aboard. Thus, the safety responsibility shifts from train drivers to the system itself. Pilot projects for "ATO over ETCS" in GoA2 are underway for mainlines. GoA3/4 specifications are currently under development at the European level [2].

According to the common safety method of European regulations (CSM), regardless of the chosen level of automation, any new or significantly altered system must demonstrate that associated risks are justifiable [3]. One idea is to use human performance on tasks taken over by automation, such as obstacle detection, as a benchmark [4]. From this perspective, the onboard detection systems must at least reliably match the safety performance of human drivers [5]. This approach relies on two key considerations: first, the fundamental assumption that existing systems incorporating human drivers currently meet established safety criteria—a safety level which must then be achieved or surpassed by the new automated system— and second, the need to evaluate functional requirements, such as reaction times, as specification parameters.

Implementing a human-as-reference approach requires a comprehensive understanding of human performance capabilities and limitations in safety-critical tasks. Accurately describing train driver performance for safety-critical functions is challenging; it cannot be reliably derived solely from accident statistics [6], which are inherently limited and often fail to capture the frequency of crucial non-accident events like near misses. To address this challenge and obtain representative data on human performance, controlled experimental investigations utilizing driving simulators offer a promising and practical approach for obtaining large, structured datasets under controlled conditions. Drawing on this motivation, the present study set out to investigate train driver performance under different driving conditions using simulator experiment studies.

Critical to successful task performance in train driving is the effective utilization of train drivers' senses. Train drivers utilize their senses to gather information, perceive the external environment, and monitor the correct functioning of the train. When analysing the tasks of train drivers, it becomes evident that one of their primary responsibilities is to perceive and process information from various visual stimuli [7]. Accordingly, the requirements for perceptual performance in obtaining a train driver's license in Germany (TfV) primarily emphasise visual abilities [8]. Thus, the current study focuses on the visual perception of train drivers.

Extensive research has examined the visual perception of car drivers (see, e.g., [9], [10]). In contrast, fewer studies have explored the visual behaviour of train drivers. One study analyzed gaze patterns of train drivers between signals, tracks ahead, in-cab, and environment [11]. Another study focused on the visual performance of urban train drivers in Australia [12]. Even fewer studies addressed the visual performance of train drivers under different conditions. These include investigating the effect of train speed and background image complexity on driving performance [13] and studying the effect of the visual field of view on signal detection [14].

Given the critical role of visual perception in ensuring safe operations and its implications for the development of ATO systems, there is a need for more targeted research to enhance understanding of train driver performance across diverse operational modes, including different train driving models such as ERTMS/ETCS and on-sight driving. Therefore, this study aimed to investigate train drivers' visual perception performance across different scenarios through simulator experiments. Visual perception performance was operationalized as the reaction time to visual stimuli, aiming to identify critical factors influencing the perception performance of train drivers.

2. Background

Reaction time to visual stimuli is a common measure of visual perception performance, influenced by various factors [15]. It is widely acknowledged that the physical properties of stimuli significantly affect reaction time. Stimulus intensity, such as the differences in brightness or colour between the object and its background, influences reaction time [15]. Research shows that reaction time decreases with larger stimulus size [16] as well as with increasing luminance of the stimulus [17]. However, the relationship between luminance and reaction time is complex, depending on the range or level of intensities [18] and other factors, such as task difficulty [16].

Environmental characteristics also influence visual perception. Poor lighting and visibility impair performance significantly [19]. Low environmental complexity can cause tunnel vision, restricting the useful field of view (UFOV) to objects directly in the line of sight [20, 21]. Viewer state also influences visual performance: severe fatigue and alcohol can induce tunnel vision, while high cognitive workload reduces the UFOV due to limited processing capacities. Conversely, low workload conditions diminish the UFOV due to decreased attention levels[20, 22, 23, 21, 24].

In driving tasks, the driver's speed influences visual perception. People in motion look about three seconds ahead, shifting the fixation point forward, resulting in a deterioration of peripheral perception at close range [20]. In a simulated car driving task, higher driving speeds led to faster reactions to road markings [25] but narrowed the UFOV [26]. For train drivers, higher speeds were associated with more vertical and fewer horizontal gaze fixations, whereas lower speeds involved more horizontal gaze movements with a lateral sweeping motion [27], influencing the detection of visual stimuli. A study using hazard perception test to simulate foreign objects appearing on railway tracks found that drivers' response times decreased at higher speeds due to increased vigilance and visual tunnelling [28]. A study using a VR train driving simulator found that at higher speeds, drivers' reaction times to a visual task (i.e., detecting non-target visual stimuli) decreased, but the accuracy of their detections was lower. The study indicates that higher vehicle speeds significantly elevate psychological pressure, as reflected in an increased heart rate and changes in heart rate variability [29]. In a simulator study, approach speed significantly influenced drivers' ability to detect and recognize signs/signals [30]. A negative non-linear relationship was identified between time to arrival after detection/recognition and train speed. An online study using reaction time tasks to cubes on the track found faster reaction times for faster speeds and higher object contrast and size [31]. Additionally, the allocation of attention between the driver's cab and the outside area influences the detection probability of trackside hazards. The division of visual attention in ETCS with cab signalling significantly reduces the time for observing the track compared to the operation with lineside signalling [32, 33, 34, 35].

3. Materials and Method

3.1. Study design

This study aimed to determine how fast train drivers perceive visual information while driving under various conditions. Driver perceptual performance was defined as the reaction time to perceived visual stimuli. Participants were tasked with driving a train and responding to stationary stimuli placed on or near the tracks at irregular intervals by pressing the train horn. The study employed a partially crossed within-subject design, where participants responded to stimuli varying in contrast (high vs. low) and size (large vs. small) while operating under three different train protection systems (ETCS, PZB, and on-sight driving) at various speed levels.

Two simulators were used to enhance the validity of the results. Consistent findings across both simulators could indicate a higher reliability of the findings. The driving simulators of the Department of Rail Operations and Infrastructure at the Technical University of Berlin (TUB) and the Institute of Transportation Systems at the German Aerospace Center (DLR) were used. From here on, these two phases of the experimental study will be referred to as DLR-study and TUB-study. Different routes were simulated using different software (VIRES and Zusi) in driving simulators. Track geometry and driving surroundings provide essential visual cues that inform a driver's visual behaviour [36]. For example, the optic flow of the visual scene significantly influences driver gaze behaviour [13], and external elements in driving surroundings can lead to visual distraction [37]. Additionally, differences in the physical setup of simulators, such as the location and the responsiveness of controls, could influence the motor reaction time. Therefore, another research question was formulated to examine to what extent using two distinct train driving simulators leads to statistically significant variances in train drivers' reaction times.

Stimuli were cubes of different sizes and colours, appearing at a distance of 800 m ahead of the train. Participants could view and respond to the stimuli from the moment they appeared until the train passed their position. Stimuli appeared either in the middle of the track (DLR-study) or next to the track right or left side counterbalanced) within a maximum distance of 3 m from the track centre (TUB-study). This difference was due to the technical capabilities of the simulator software, however, in both studies, stimuli appeared in the central field of view of drivers from the point of observation.

Participants were instructed to react to every recognized obstacle as quickly as possible without needing to evaluate their hazard potential. This approach enabled us to measure sensory perception time without considering further cognitive processing time. A within-subject design was employed, where each participant was exposed to all experiment conditions. This approach allowed for a comparison of their performance across different scenarios while controlling for individual differences.

3.2. Simulation environment

The TUB study was conducted in a driving simulator at TU Berlin; see Figure 1a. The driver's cab of the simulator meets the requirements of the European Driver's Desk. All relevant technical and operational information was displayed on several touch screens. The driving simulator was operated with the Zusi 3 Professional software [38], which provides various route modules, realistic operating rules, and accurate driving physics. The simulated view was presented on a modern 32-inch UHD monitor (3,840 x 2,160 resolution, 60 Hz refresh rate, 2500:1 contrast ratio, 1500R curvature). Since the simulator was not located within a dedicated train cab mock-up, curtains were installed on the windows to minimize reflections and glare.

The DLR study was conducted in the RailSET® (Railway Simulation Environment for Train drivers and operators, Figure 1b), a train driver's cab simulator at the DLR Institute of Transportation Systems (for the simulator specifications, see [39]). The simulator was operated using an original control panel of a traction unit. The simulation environment is based on the VIRES software (VIRES Simulationstechnologie GmbH, Bad Aibling, Germany). A video projector shows the simulated view to the front, while the view from

(a) TU Berlin train driving simulator.

(b) DLR train driving simulator.

Figure 1: Driving simulators used in the study.

the side windows is displayed on screens. An audio system in the cabin provides ambient sounds modelled on the interior of a real train driver's cab.

3.3. Independent variables

The simulator experiments were designed to reflect the impact of selected influencing factors and to capture the range of human performance utilizing the sense of sight. Several influencing factors were chosen as independent variables, varying within the scope of the simulator study. This variation allowed for exploring the relationships between these factors and reaction time, providing insights into how different conditions affect train drivers' perceptual performance.

3.3.1. Operational parameters

The operational parameters in the study included the train protection system (PZB, ETCS in-cab signalling and on-sight driving) and train speed (40 km/h, 100 km/h and 160 km/h) at the time of object appearance. Both ETCS and PZB routes included a speed level of 100 km/h, while 40 km/h was implemented for PZB and on-sight driving (OS). Only the ETCS route allowed driving at 160 km/h. On-sight driving refers to scenarios where train drivers cannot rely on a clear track indicated by signals; instead, they must visually identify hazards and, stop if necessary. Drivers choose their speed based on visibility and track conditions, not exceeding 40 km/h, while focusing primarily on observing the tracks. However, when a train safety system is used, additional attention must be directed towards the displays in the driver's cab, thus dividing their attention. Although the routes used in the TUB and DLR studies differed, efforts were made to place stimuli at comparable locations, such as on straight, level track sections.

3.3.2. Physical properties of the stimuli

The study included two key variables related to the visual perception of stimuli: the size of the stimuli and their contrast to the background. In both simulators, the contrast was manipulated by varying the colour of cubes, with conditions of high and low contrast. For the high-contrast condition, a bright orange (HEX Code #f18e2a), similar to the colour of the safety vest was used. The colour contrast between this high-contrast colour and an average background colour were then calculated to determine the low-contrast condition. An orange-brown hue (HEX code #9d6830) was selected, offering approximately half the colour contrast of the high-contrast colour. The colour contrast was calculated using the delta-E formula in RGB colour space proposed by [40]. The colour difference (ΔE) is defined as the Euclidean distance between points in the RGB colour space. The background colour was calculated based on the immediate surrounding colour of each stimulus. For example, the average colour difference between the stimulus and its immediate background in the TUB simulator was 222 for the high contrast and 112 for the low contrast condition, which corresponds to an average contrast ratio of 50.4%.

The stimuli varied in size based on cube edge lengths in both simulators: relative sizes of 90 cm (small) and 180 cm (large) within each simulator, approximating the heights of an adult and a child. To account for differences in display size and viewing distance between the simulators, the visual angle for each stimulus at the time of the stimulus appearance was calculated for comparability (Table 1). The angular size, measured in arcminutes ($'$), represents the visual angle subtended by an object at the eye, considering both the physical size of the object on the monitor and the viewing distance. To clarify these differences without relying solely on physical measurements, we coded the stimulus sizes as S1, S2, S3, and S4, with S1 representing the smallest visual angle and S4 the largest. This ensured consistent comparison of the size variable across simulator setups.

Table 1: Relative stimulus sizes and corresponding visual angle in arcminute

Stimuli	**Selected Size (cm)**	**Apparent Size (arcminute)**
TUB_Small (S1)	90	3.78
TUB_Large (S3)	180	7.56
DLR_Small (S2)	90	6.72
DLR_Large (S4)	180	13.43

In total, the combination of the independent variables resulted in 35 experimental conditions, each represented by one stimulus. Differences in the number of stimuli between scenarios resulted from constraints imposed by the simulator setup, such as the limited availability of suitable route sections for stimulus placement. Table 2 shows an overview of the experimental conditions.

Table 2: Experimental conditions. Angular sizes are given in arcminutes (′). Rows for high-contrast conditions are in bold font.

	40 km/h		**100 km/h**		**160 km/h**	
	Angular size	**Contrast**	**Angular size**	**Contrast**	**Angular size**	**Contrast**
OS	**13.43**	**High**				
	7.56	**High**				
	3.78	**High**				
	13.43	Low				
	7.56	Low				
	3.78	Low				
PZB	**13.43**	**High**	**13.43**	**High**		
	7.56	**High**	**7.56**	**High**		
	6.72	**High**	**6.72**	**High**		
	3.78	**High**	**3.78**	**High**		
	13.43	Low	13.43	Low		
	7.56	Low	7.56	Low		
	6.72	Low	6.72	Low		
	3.78	Low	3.78	Low		
ETCS			**13.43**	**High**	**13.43**	**High**
			7.56	**High**	**7.56**	**High**
			3.78	**High**	**3.78**	**High**
			13.43	Low	13.43	Low
			7.56	Low		
			6.72	Low	6.72	Low
			3.78	Low	3.78	Low

3.3.3. Other independent variables

In addition to the independent variables previously presented, further data were collected that may have a possible influence on the participants' visual perception performance, namely gender, age, work experience, and prior experience with PZB and ETCS. However, these variables were not manipulated in the experiment. Participants' alertness was assessed before the first trial and after each experimental block using the Karolinska Sleepiness Scale [41].

3.4. Dependent Variables

In this study, the dependent variable was the reaction time to a stimulus. Reaction time was measured as the duration from the appearance of the stimulus at a visible distance of approximately 800 meters until the recognition of the stimulus. Participants were instructed to activate the train horn in response to seeing the stimuli. This action aligns with a behaviour commonly practised in reality after recognizing an object on the track. Thus, the recognition of the stimulus was measured by the activation of the train horn.

3.5. Procedure

First, participants completed a demographic characteristics questionnaire and the Karolinska Sleepiness Scale (KSS) [41] on a tablet. Subsequently, participants completed three experimental trials (ETCS track, PZB track, and on-sight driving). The PZB or ETCS tracks were always completed first, with the sequence balanced among participants. Due to technical limitations of the simulator software, the on-sight driving scenario was consistently the final experimental block. Participants were instructed to press (pull or push in the TUB study) the train horn upon seeing an orange or brown cube. Participants were informed that there was no risk of collision or need to alter the train's operation due to the displayed objects. Each participant required approximately 3 hours to complete the three experimental blocks and questionnaires.

3.6. Participants

Qualified train drivers were recruited to participate in the simulator study. In the TUB study, 17 male and 1 female active train drivers participated, with an average age of 33.4 years (age range of 22-57 years). The participants had on average 7.3 years of professional experience (age range of 1-28 years). None of the participants had prior experience with the TUB simulator setup. Participants rated their familiarity with different types of train safety systems on a scale from one (not familiar at all) to ten (very familiar). Familiarity with the PZB system was rated at an average of 8.3, while familiarity with the ETCS system averaged 2.2, with only 2 participants with a rating of 5 or above.

In the DLR study, a total of 25 professional train drivers participated, with an average of 9.92 years of professional experience (range: 1-39 years). All participants were male, with an average age of 33.7 years (range: 22-57 years). None of the participants had previously taken part in a study with DLR's simulator setup RailSET. Participants rated their familiarity with the train safety system PZB at an average of 9.8, while familiarity with the ETCS train safety system was rated at an average of 2.0. All except two participants rated their familiarity with ETCS as less than five.

3.7. Data Analysis

Timestamps for each stimulus occurrence and the activation of the train horn were extracted from the simulator logs. Reaction times were calculated as the difference between these timestamps. Deviations between specified and actual speeds were calculated to ensure train speed matched the independent variable levels (i.e., 40 km/h, 100 km/h, or 160 km/h) at the time of cube appearance. The interquartile range (IQR) of the actual speeds was computed, and data points with deviations exceeding three times the IQR at the time of stimulus appearance were excluded from the analysis (six cases in the DLR study, five cases in the TUB study). After data cleaning, a total of 690 observations from 43 participants remained.

A descriptive analysis of reaction time data was conducted before applying inferential statistics. Reaction times typically exhibit a positively skewed distribution characterized by a minimum bound at just above zero seconds and a long tail of longer reaction times. This pattern was observed in both simulator studies. To address this skewness and facilitate statistical analysis, reaction times were logarithmically transformed, a standard method to normalize data distributions and mitigate the impact of outliers [42, 43].

The impact of independent variables on the log-transformed reaction time was analysed using a linear regression model. A mixed-effects model, incorporating participants nested within simulators (TUB and DLR) as a random effect factor, was employed for the analysis. This model accounts for potential systematic differences between the simulators, enabling the examination of both overall effects of independent variables and variations across simulators [44]. For the linear regression, mixed-effects modelling with the restricted maximum likelihood (REML) approach was employed.

3.8. Research Hypotheses

To evaluate the effect of independent variables on reaction time, several research hypotheses have been determined. The research hypotheses are formulated based on the beta coefficients (β) of the underlying regression model. These coefficients represent the impact of the independent variables on the dependent variable. Research hypotheses can be defined as follows:

It was expected that lower contrast and smaller size would decrease stimulus salience. Therefore, smaller stimuli were expected to be detected slower than larger ones, and slower reactions were expected for low-contrast stimuli compared to high-contrast stimuli.

- H1: Reaction time is longer for small stimuli: $\beta_{1.1}, \beta_{1.2}, \beta_{1.3} > 0$.
- H2: Reaction time is longer for low-contrast stimuli: $\beta_2 > 0$.

Since stimuli appeared on or near the tracks, increased attention focused on the track area at higher speeds would lead to faster reactions to stimuli appearing at higher speeds than those at lower speeds. Additionally, due to optical effects, objects visually enlarge more rapidly at higher speeds, facilitating recognition.

- H3a: Reaction time is longer at slower driving speeds (40 km/h) compared to higher speed conditions: $\beta_{3.1} > 0$.
- H3b: Reaction time is longer at a driving speed of 100 km/h compared to 160 km/h: $\beta_{3.2} < 0$.

It was expected that the use of a train safety system (PZB or ETCS) would lead to longer reaction times compared to on-sight driving, as drivers focus more on track monitoring during on-sight driving. Furthermore, it was expected reaction times would be longer with ETCS cab signaling than with PZB, due to the higher attention demands on the control panel in ETCS compared to PZB.

- H4a: Reaction time is longer when using a train safety system (PZB) compared to on-sight driving: $\beta_{4.1} < 0$.
- H4b: Reaction time is longer when using ETCS compared to PZB: $\beta_{4.2} > 0$.

Below is the notation of the mixed-effects linear regression model. The reference level is as follows: largest visual angle (13.43'), high contrast, driving speed of 100 km/h and PZB scenario.

$$\begin{aligned}
\text{Log(Reaction Time)} = {} & \beta_0 + \beta_{1.1} \cdot \text{angular.size}_{3.78} + \beta_{1.2} \cdot \text{angular.size}_{6.72} \\
& + \beta_{1.3} \cdot \text{angular.size}_{7.56} + \beta_2 \cdot \text{contrast_low} \\
& + \beta_{3.1} \cdot \text{speed}_{(40\ \text{km/h})} + \beta_{3.2} \cdot \text{speed}_{(160\ \text{km/h})} \\
& + \beta_{4.1} \cdot \text{tpc}_{\text{os}} + \beta_{4.2} \cdot \text{tpc}_{\text{etcs}} + u_{\text{simulator:subject}} + \epsilon
\end{aligned} \tag{1}$$

The term angular.size represents the dummy variable for three levels of stimuli size, with the largest stimuli chosen as the reference level. The term tpc represents the variable for the train protection system. β_0 represents the intercept or base value of the logarithmic reaction time at the population average, assuming all other variables are at their reference values.

The model includes two random components: within-group residual errors (ϵ) and random effects for the covariates ($u_{\text{simulator:subject}}$). The residual errors for the same group are independent of the random effects. The random factor accounts for the differences in the intercepts between the participants, clustered within two simulators, due to inherent differences or unobserved factors. In the random part of the model, the estimated parameters are the variances of the random effect (σ_u^2) and the residual error (σ_ϵ^2). The variance of the random effect captures the variability in reaction times that can be attributed to differences between participants, considering the clustering within simulators, while the variance of the residual error captures the variability in reaction times that cannot be explained by the fixed effects or the random effects [45].

4. Results

4.1. Descriptive Analysis

The distribution of reaction times, depicted in Figure 2, confirms the typical non-normal pattern of reaction time data, with a lower bound just above 0 seconds and a long tail on the right. Given this distribution, the median and the geometric mean are more suitable measures of central tendency than the mean, as they are less influenced by outliers. Reaction times were transformed on a logarithmic scale. The histogram of log-transformed reaction times (Figure 2) and the cumulative distribution function (CDF) of the transformed data indicates near-normality (Figure 3). Figure 4 presents the reaction times for various experimental conditions with their geometric mean values. Black points represent the recorded reaction times, while the orange and brown points indicate the geometric mean of reaction times for each specified condition. To establish causal relationships, a thorough examination using linear regression analysis was conducted. Table 3 shows the geometric mean and standard deviation of reaction times for the examined experimental conditions.

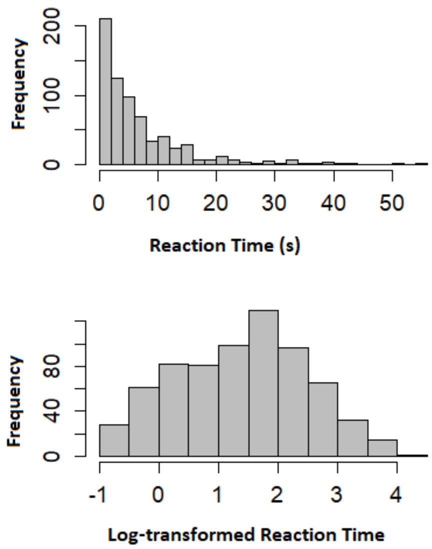

Figure 2: Histogram of raw and log-transformed reaction times

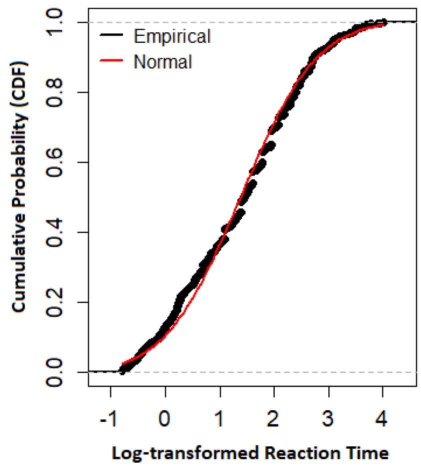

Figure 3: Empirical and theoretical CDF of the transformed reaction times

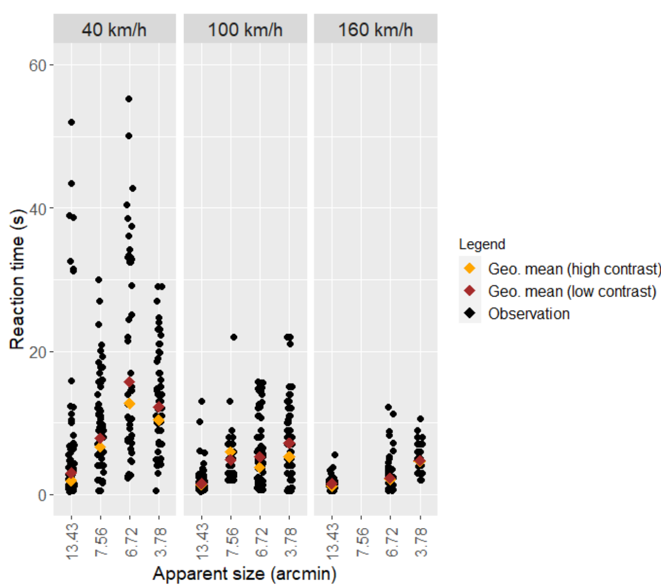

Figure 4: Reaction time at different speed and size conditions and geometric means for different contrast levels

Table 3: Geometric means and standard deviations of reaction times for the examined experimental conditions. Angular sizes are given in arcminutes (′). Rows for high-contrast conditions are in bold font.

	40 km/h		**100 km/h**		**160 km/h**	
	Angular size	**Geo. mean (SD)**	**Angular size**	**Geo. mean (SD)**	**Angular size**	**Geo. mean (SD)**
OS	**13.43**	**1.36 (1.58)**				
	7.56	**6.35 (6.34)**				
	3.78	**9.92 (6.80)**				
	13.43	1.51 (1.93)				
	7.56	7.25 (8.89)				
	3.78	13.44 (6.97)				
PZB	**13.43**	**2.90 (8.73)**	**13.43**	**1.22 (0.68)**		
	7.56	**7.02 (5.77)**	**7.56**	**5.99 (2.17)**		
	6.72	**12.68 (11.77)**	**6.72**	**3.18 (1.96)**		
	3.78	**10.80 (5.41)**	**3.78**	**5.60 (4.95)**		
	13.43	6.42 (15.46)	13.43	1.56 (2.64)		
	7.56	8.51 (4.34)	7.56	6.80 (3.90)		
	6.72	15.73 (16.16)	6.72	4.74 (4.81)		
	3.78	11.05 (8.45)	3.78	7.77 (2.50)		
ETCS			**13.43**	**1.46 (2.07)**	**13.43**	**1.14 (0.85)**
			7.56	**4.61 (5.51)**	**7.56**	**2.09 (2.99)**
			3.78	**4.88 (3.68)**	**3.78**	**4.60 (1.17)**
			13.43	1.52 (1.13)	13.43	1.53 (1.03)
			7.56	3.41 (1.01)		
			6.72	5.78 (4.65)	6.72	2.32 (2.51)
			3.78	6.74 (6.95)	3.78	4.78 (2.66)

4.2. Regression Analysis

This study examined the factors that influence reaction time through a mixed-effects linear regression model. The R package lme4 was used for the analysis [46]. The model was applied to a dataset of 690 observations, with log-transformed reaction times as the dependent variable. Fixed and random effects were analyzed to assess their influence on reaction time.

The estimates and model statistics are presented in Table 4. Each fixed-effect coefficient represents the expected change in the log-transformed reaction time for a unit change in the predictor variable. The standard error (SE) estimates the uncertainty of the coefficient, while the t-value and the p-value assess the statistical significance. One-sided p-values were calculated using the Satterthwaite method.

Table 4: Fixed Effects (*p<0.05)

Variable	**Est. (β)**	**S.E.**	**t val.**	**CI low**	**CI high**
(Intercept)	0.34	0.11	3.13	0,13	0,55
angular_size_3.78*	1.45	0.14	10.42	1,18	1,72
angular_size_6.72*	1.07	0.07	14.52	0,92	1,21
angular_size_7.56*	1.09	0.14	7.54	0,81	1,37
contrast_low*	0.22	0.06	3.92	0,11	0,32
speed_40 km/h*	0.86	0.08	10.54	0,70	1,02
speed_160 km/h*	-0.34	0.09	-3.82	-0,51	-0,17
os*	-0.64	0.09	-7.00	-0,82	-0,46
etcs	-0.07	0.08	-0.85	-0,24	0,09

The conditional R^2 was 0.58, indicating that the model explained approximately 58% of the variance in reaction time. The positive and negative signs denote increases or decreases in reaction time compared to the baseline level, respectively. Exponentiated coefficients reveal the multiplicative effect of a unit change in predictor variables. For example, the expected reaction time at a speed of 40 km/h is 136% higher than at a speed of 100 km/h, whereas at 160 km/h, it is 29% lower than at 100 km/h. This confirms the hypotheses H3a and H3b.

Results indicate that the stimuli size had the largest effect on reaction time. Smaller stimuli led to longer reaction times compared to the largest stimulus (H1), with the smallest stimulus size (S1) causing the biggest increase. The two mid-sized stimuli (i.e. S2 and S3 with 6.72 arcmin and 7.56 arcmin, respectively), differing by 12.5% in size, had nearly identical effects on reaction time compared to the largest stimulus (i.e. 197% and 192% increase). A post-hoc Sidak test confirmed that this small difference between the levels of 6.72 and 7.56 was not statistically significant (p=0.99). Low contrast stimuli resulted in a 25% increase in reaction times, supporting hypothesis H2.

On-sight driving resulted in 47% faster reaction times compared to driving under PZB. Contrary to the expectations, the ETCS scenario showed a 7% decrease in reaction times compared to the PZB scenario, but this difference was not statistically significant. Thus, all hypotheses were confirmed except for the relationship between ETCS and PZB, with stimulus size having the most substantial impact on reaction time and contrast having the least (Table 5).

Table 5: Summary of hypothesis testing results

Hypothesis	Result
H1	Confirmed: Reaction time is longer for small stimuli.
H2	Confirmed: Reaction time is longer for low-contrast stimuli.
H3a	Confirmed: Slower driving speeds (40 km/h) resulted in longer reaction times.
H3b	Confirmed: Reaction times at 100 km/h were longer than at 160 km/h.
H4a	Confirmed: Reaction time was longer at PZB than at OS.
H4b	Not confirmed: ETCS did not show an increase in reaction times compared to PZB.

The likelihood ratio test with 10000 simulated values suggested that the model with random effects provided a better fit than the fixed-effects-only model ($RLRT = 97.7, p < .01$). This indicates that incorporating random effects helps account for variability in the data due to the grouping structure. The random intercept variance for participants grouped within simulators was 0.14, with an estimated standard deviation of 0.38 on the log-transformed scale (Table 6). The intra-class correlation coefficient (ICC) for this grouping variable was 0.22. The intercept given in the Table 4 represents the population average. One intercept value per subject can be calculated to account for the differences between participants.

Table 6: Random Effects

Groups	Variance	Std.Dev.
simulator:subject (Intercept)	0.14	0.38
Residual	0.52	0.72

Marginal predictions estimate the average response time across all levels of random effects, while conditional predictions take into account the specific random effects associated with each case [47]. Assuming that the subject sample in the study is a representative random sample of the real world, the marginal model for the predictions can be used to provide an estimate for those who do not belong to one of the clusters used in the study [48]. The geometric mean of the observed data and the model estimations are shown in Figure 5.

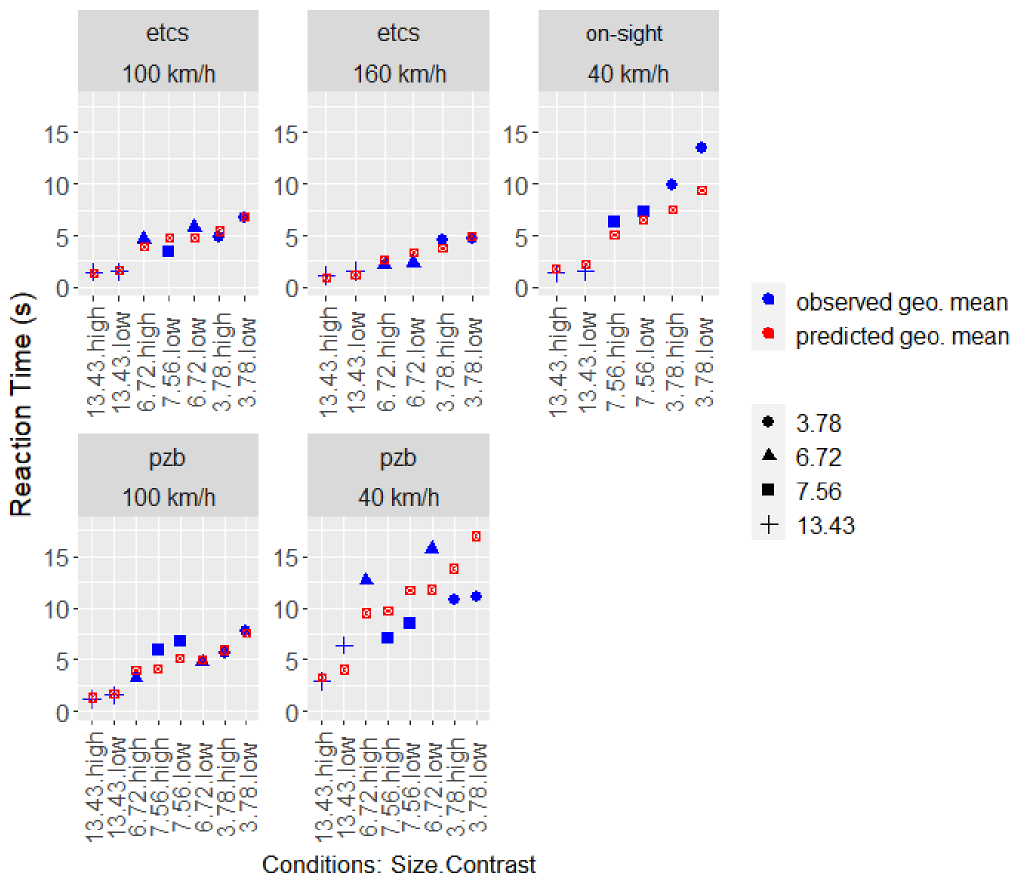

Figure 5: Comparison of the model prediction with the corresponding indicator of the central tendency across conditions. Y-axis: reaction times in seconds. X-axis: Experimental conditions as a combination of angular size and contrast levels.

Sleepiness was assessed using the Karolinska Sleepiness Scale (KSS) before the experiment and after each experimental block. The scale points range from 1 - extremely alert to 9 - very sleepy, great effort to keep alert. Over the course of the experiments, the participants' self-reported sleepiness remained at a similar level, around Alert (3) and Fairly Alert (4), without a discernible pattern.

5. Discussion

Ensuring the safety of increasingly automated railway systems, such as Automatic Train Operation, necessitates robust methods for defining and validating performance requirements. As outlined in the Introduction, a promising approach in safety assurance frameworks is to use the established safety performance of human train drivers as a reference system against which automated capabilities can be benchmarked [4], [5]. Implementing this approach requires a detailed characterization of human performance capabilities and limitations in tasks designated for automation, such as obstacle detection. This study contributes directly to this essential step by providing empirical data on train driver reaction times in perceiving and reacting to target visual cues under various conditions.

This study set out to test the influence of different object properties and operational parameters on train drivers' reaction times to objects on the track. Significant effects were observed for object size, background contrast, and driving speed, with object size having the largest impact. Reaction times were longer for detecting small stimuli compared to large ones. Higher stimulus-background contrast reduced reaction times, consistent with the concept that stimulus intensity, such as size and the differences in brightness or colour between the object and its background, influences reaction time. Driving speed was another significant factor, with faster reactions at higher speeds, which may support the assumptions regarding the upward gaze shift at higher speeds. Additionally, objects appear to increase in size more quickly at higher speeds and are therefore recognized more swiftly.

The findings supported the hypothesis that on-sight driving leads to shorter reaction times compared to PZB and ETCS L2 incab signalling, due to the increased track monitoring during on-sight driving.In this study, although the order of the experimental blocks for PZB and ETCS was equally randomised, on-sight driving was always the last block. Although fatigue did not have a significant effect on reaction times, the order of experimental blocks should be fully randomised to minimise the potential effect of fatigue on one particular system. However, the hypothesis of having longer reaction times at ETCS, compared to PZB, was not supported. The results regarding a comparison between the two train control systems PZB and ETCS L2 were likely influenced by other factors, such as variations in track design between ETCS and PZB routes in the simulator study [49]. Nonetheless, these variations in track design reflect realistic differences in environments where these systems are deployed in the real world. Thus, there is a need for further research into the relationship between reaction time and train protection systems, accounting for various underlying factors e.g. in the track design.

Although separate analyses of both simulators' results revealed similar patterns, a random effect analysis showed a significant clustering effect among participants within simulators. Overall, the model accounted for 58% of the variance in reaction times, explained by both the fixed effects and the random effects (Psuedo-Rˆ2=0.58). Relative validity between simulator results can be shown by independent variables having the same direction of effects. Additionally, the variability in reaction times caused by using different simulators was considered by the random effect structure. It was found that approximately 22% of the total variance in the outcome variable is due to differences between subjects within simulators (ICC= 0.22). The remaining 78% of the variance is due to the residual variability within subjects. This suggests that there is some clustering effect, but most of the variability is within subjects rather than between subjects. Although standardizing variables like apparent object size using arcminutes helped capture some variance between simulators, factors which were not captured in this study such as route geometry could have contributed to the variability between simulators. Nevertheless, this study demonstrated how to account for differences in the simulators resulting from visual setup using a standardization procedure, showing how studies from different simulators can still be compared. Further studies should focus on developing standardization methods for other factors like track design to further delineate which variability results from the participants vs. the simulator setup. Moreover, this study provided valuable insight into the effectiveness of simulator-based research in examining the visual performance of train drivers, providing a basis for future studies to enhance validity through replication.

It is crucial to evaluate the applicability of these findings, before applying them to the development of requirements for future ATO systems. The transferability of results to real-world rail operations depends on personal influencing factors, operational parameters, and physical properties of the stimulus. In simulators, participants knew stimuli would appear, which likely increased visual search behaviours beyond real-world levels. Conversely, one of the central tasks of drivers on a real journey is to monitor the track environment. Therefore, attentive visual monitoring of the infrastructure should also occur during an actual journey. However, the absence of natural risks in simulator settings in the event of inattention may diminish the perceived urgency of visual search tasks.

The study employed a simple reaction task where participants responded to each stimulus without distinguishing whether the object represented a danger. In real-world operations, drivers' responses can vary from emergency stop to activating the train horn or no reaction at all, depending on the situation. Thus, reaction times in practical settings would likely be longer due to the additional time needed to process information and determine an appropriate response. However, at higher speeds, particularly when an obstruction is detected at 800m, such as in this study, the options for intervention become limited. Future research could employ sensitivity analysis or explore different distance ranges to develop a more comprehensive benchmark aligned with safety criteria at higher speeds.

Other influencing factors include journey duration and route familiarity. The short driving periods in the studies minimized the negative effects of fatigue or vigilance loss. However, during extended real-world journeys, such as a seven-hour shift, reaction times could be adversely affected compared to those observed in our studies [50]. Lack of route knowledge may also have hindered effective visual search strategies. In real-world driving, familiarity helps drivers anticipate and react more effectively to objects in expected locations, such as level crossings.

For evaluating the applicability of these operational boundary conditions to real-world scenarios, it is crucial to consider the complexity of the tasks and the route geometry. Participants focused solely on driving, unlike real operations, which include additional tasks like dispatcher communication, timetable checks, and diagnostic monitoring. Theoretically, auditory or verbal tasks such as communication are not expected to negatively affect visual performance [51]. On the other hand, other visual tasks, such as monitoring fault displays or timetables, could impair the driver‘s performance to monitor the infrastructure effectively.

Stimuli were always placed under ideal visual conditions - on straight routes with little or no gradient, with minimal obstructions, allowing participants to detect the objects from 800 meters away. In real-world operations, the drivers often face compromised views due to curves, gradients, or vegetation.

Visual stimuli represent a reference without explicitly defining parameters such as shape and pattern, which might influence reaction time. The decision to use a cube was a compromise between using a human-sized object and maintaining an abstract form to prevent traumatic experiences. The influence of specific shapes and patterns of the stimuli on reaction time is outside the scope of this study. At higher speeds, the distortion of visual cues—such as motion blur—can impede the driver's ability to quickly and accurately detect these objects. This raises the question of whether there is a threshold speed beyond which faster detection becomes impractical due to perceptual limitations and object characteristics. Future studies should explore this aspect by testing different object characteristics at a higher range of speeds. The colour difference in RGB colour space was used to calculate the colour contrast between the stimuli and their background. Differences in luminance, glare, and contrast between simulator screens and actual conditions can further impact object perceptibility.

The analysis produced average reaction time estimations between 1.51 and 15.73 seconds across different conditions. The least favourable conditions in terms of reaction times were small and low-contrast objects at 40 km/h under the PZB system.

This study focused on operationalizing a critical aspect of human perception performance relevant to tasks designated for automation: driver response quantified as simple reaction time to visual stimuli. The complex process of formally deriving, validating, and applying these performance characteristics as definitive safety benchmarks for ATO systems constitutes a significant area for future research. The empirical reaction time values obtained in this study provide a foundational dataset that can be directly utilized in future work to derive specific human-referenced benchmarks or parameters for such risk criteria. For instance, by combining these human reaction times with factors like train speed and available stopping distance, metrics such as safe detection range, minimum required obstacle size detection capabilities, or collision probabilities based on human limits could be estimated.

In summary, this study provided insights into specific aspects of visual perception. Future research could benefit from exploring additional parameters such as more complex tasks and driving situations, dynamic objects or longer travel times on familiar routes to comprehensively assess drivers' visual perception performance.

6. Conclusion

This paper presented findings from two simulator studies investigating factors influencing train drivers' reaction times to objects along the track. The results revealed significant effects of object size, object contrast, and train speed on train drivers' reaction times. Larger and more contrasting objects were associated with faster reaction times, while stimuli were detected more quickly at higher speeds. The study produced average reaction time predictions between 1.14 and 15.73 seconds across different conditions. The least favourable condition based on observed values was small low-contrast stimuli (S1 and S2) approached at 40 km/h while using the PZB system. The visual performance values obtained in this study may be used for deriving safety metrics that can serve as a benchmark for developing future automated train operation systems, taking into account the limitations described above. The results provide insights into factors shaping train driver performance and guide future studies for establishing criteria for effective implementation of ATO systems.

Conducting such experiments on actual tracks is impractical and hazardous, highlighting the invaluable role of simulator studies in understanding parameters influencing train driver performance. The study provides information about aspects influencing the comparability of results obtained from different simulators in similar experiments while demonstrating a way to standardize differences between simulator setups.

7. Acknowledgements

This work is closely linked to the ATO-Sense project, which is part of a project family commissioned by the German Center for Rail Transport Research on the topic of automated train operation (ATO) [52].

References

[1] IEC, Bahnanwendungen - automatischer städtischer schienengebundener personennahverkehr (augt) sicherheitsanforderungen (2010).

[2] X2Rail-1, Ato (up to goa3/4) system requirements specification: (x2r-t4.7-d-als-001-06) (2019).

[3] Die Europäische Kommission, Durchführungsverordnung (eu) nr. 402/2013 der komision vom 30. april 2013 über die gemeinsame sicherheitsmethode für die evaluierung und bewertung von risiken und zur aufhebung der verordnung (eg) nr. 352/2009: 402/2013 (2013).

[4] L. Quante, M. Zhang, K. Preuk, C. Schießl, Human performance in critical scenarios as a benchmark for highly automated vehicles, Automotive Innovation 4 (3) (2021) 274–283. doi:10.1007/s42154-021-00152-2.

[5] S. Rosić, D. Stamenković, M. Banić, M. Simonović, D. Ristić-Durrant, C. Ulianov, Analysis of the safety level of obstacle detection in autonomous railway vehicles, Acta Polytechnica Hungarica 19 (3) (2022) 187–205.

[6] C. Harrison, J. Stow, X. Ge, J. Gregory, H. Gibson, A. Monk, At the limit? using operational data to estimate train driver human reliability, Applied ergonomics 104 (2022) 103795.

[7] N. Brandenburger, H.-J. Hörmann, D. Stelling, A. Naumann, Tasks, skills, and competencies of future high-speed train drivers, Proceedings of the Institution of Mechanical Engineers, Part F:

Journal of Rail and Rapid Transit 231 (10) (2017) 1115–1122. doi:10.1177/0954409716676509.

[8] Verordnung über die erteilung der fahrberechtigung an triebfahrzeugführer sowie die anerkennung von personen und stellen für ausbildung und prüfung (triebfahrzeugführerscheinverordnung - tfv), Nichtamtliches Inhaltsverzeichnis, ausfertigungsdatum: 29.04.2011, Vollzitat: "Triebfahrzeugführerscheinverordnung vom 29. April 2011 (BGBl. I S. 705, 1010), die zuletzt durch Artikel 1 der Verordnung vom 30. November 2023 (BGBl. 2023 I Nr. 345; 2024 I Nr. 177) geändert worden ist", Stand: Zuletzt geändert durch Art. 1 V v. 30.11.2023 I Nr. 345; 2024 I Nr. 177 (Apr. 2011).

[9] H. Summala, Brake reaction times and driver behavior analysis, Transportation Human Factors 2 (3) (2000) 217–226. doi:10.1207/STHF0203_2.

[10] M. Dozza, What factors influence drivers' response time for evasive maneuvers in real traffic?, Accident Analysis & Prevention 58 (2013) 299–308.

[11] T. Luke, N. Brook-Carter, A. M. Parkes, E. Grimes, A. Mills, An investigation of train driver visual strategies, Cognition, Technology & Work 8 (1) (2006) 15–29. doi:10.1007/s10111-005-0015-7.

[12] A. Naweed, G. Balakrishnan, Understanding the visual skills and strategies of train drivers in the urban rail environment, Work 47 (2014) 339–352. doi:10.3233/WOR-131705.

[13] B. Guo, Y. Mao, A. Hedge, W. Fang, Effects of apparent image velocity and complexity on the dynamic visual field using a high-speed train driving simulator, International Journal of Industrial Ergonomics 48 (2015) 99–109. doi:10.1016/j.ergon.2015.04.005.

[14] K. Wada, M. Hataoka, Effects of the angle between objects of gaze and a visual target when driving a train, Japanese Journal of Applied Psychology 46 (SpecialEdition) (2020) 19–28.

[15] C. Becker-Carus, M. Wendt, Allgemeine Psychologie: Eine Einführung, Springer, Berlin, 2016.
URL `http://www.springer.com/`

[16] C. Bonnet, J. Gurlekian, P. Harris, Reaction time and visual area: Searching for the determinants, Bulletin of the Psychonomic Society 30 (5) (1992) 396–398. doi:10.3758/BF03334099.
URL `https://link.springer.com/article/10.3758/bf03334099`

[17] H. Piéron, Ii. recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices, L'année psychologique 20 (1) (1913) 17–96.

[18] D. Pins, C. Bonnet, On the relation between stimulus intensity and processing time: Piéron's law and choice reaction time, Perception & Psychophysics 58 (3) (1996) 390–400. doi:10.3758/BF03206815.

[19] H.-J. Schmidt-Clausen, A. Freiding, Sehvermögen von Kraftfahrern und Lichtbedingungen im nächtlichen Straßenverkehr, Berichte der Bundesanstalt für Straßenwesen, 2004.

[20] M. Land, J. Horwood, Which parts of the road guide steering?, Nature 377 (6547) (1995) 339–340. doi:10.1038/377339a0.

[21] G. Weller, B. Schlag, G. Gatti, R. Jorna, M. van de Leur, Human factors in road design, State of the art and empirical evidence-Delivery RIPCORD-ISEREST EU-project within the 6th framework. Brussels: European Commission (2006).

[22] A. Cohen, Blickverhalten des fahrzeuglenkers als komponente des verkehrssicherheitsverhaltens, Fortschritte der Verkehrspsychologie, Bd 2 (1987).

[23] T. Miura, Visual search in intersections–an underlying mechanism, IATSS research 16 (1) (1992) 42–50.

[24] B. Schlag, R. Heger, M. Baier, B. STEI NAUER, Empfehlungen zur berücksichtigung physiologischer und psychologischer fähigkeiten und grenzen der kraftfahrer bei der straßenplanung in brandenburg, Stufe: Systematisierung und Evaluation vorhandener Kenntnisse. Unveröffentlichter Bericht. Dresden: Technische Universität (2002).

[25] Y. Cao, J.-H. Wang, A driving simulation based study on the effects of road marking luminance contrast on driving safety, in: Institute of Transportation Engineers, Vol. District 1 Annual Meeting, Burlington, Vermont, 2004.

[26] J. Rogé, T. Pébayle, E. Lambilliotte, F. Spitzenstetter, D. Giselbrecht, A. Muzet, Influence of age, speed and duration of monotonous driving task in traffic on the driver's useful visual field, Vision Research 44 (23)

(2004) 2737–2744. doi:10.1016/j.visres.2004.05.026.
URL https://www.sciencedirect.com/science/article/pii/S004269890400272X

[27] D. Suzuki, K. Yamauchi, S. Matsuura, Effective visual behavior of railway drivers for recognition of extraordinary events, Quarterly Report of RTRI 60 (4) (2019) 286–291. doi:10.2219/rtriqr.60.4_286.

[28] W. Dong, W. Fang, X. Jiang, H. Bao, H. Qiu, Y. Li, Railway safety under increasing speed: Train drivers' hazard perception of foreign object intrusion on railway tracks, International Journal of Industrial Ergonomics 105 (2025) 103684.

[29] X. Zhao, Z.-R. Xiang, Z. Zhang, T.-C. Ding, H.-N. Liu, H.-B. Wang, R. Zou, Y. Wang, Factors affecting the visual ergonomics of train drivers in vr simulation driving: Snow and ice line environment and train speed, Safety Science 185 (2025) 106806.

[30] G. Li, W. I. Hamilton, G. Morrisroe, T. Clarke, Driver detection and recognition of lineside signals and signs at different approach speeds, Cognition, Technology & Work 8 (2006) 30–40.

[31] H. Wasle, A. Goralzik, B. Thomas-Friedrich, D. Schackmann, E. J. Bosch, When something is in the way: Parameters of perception and reaction speed in train drivers, in: 5th International Conference on Human Systems Engineering and Design: Future Trends and Applications (IHSED 2023), 2023.

[32] Helias Marinkos, Thomas Sheridan, and Jordan Multer, Effects of supervisory train control technology on operator attention, Tech. rep., John A. Volpe National Transportation Systems Center (US) (2005).
URL `https://api.semanticscholar.org/CorpusID:107842952`

[33] M. C. G. Hely, T. Shardlow, B. Butt, R. Friswell, A. S. McIntosh, A. Williamson, Effects of automatic train protection on human factors and driving behaviour, 2015.
URL `https://api.semanticscholar.org/CorpusID:59485921`

[34] R. Van der Weide, D. De Bruijn, M. Zeilstra, Ertms pilot in the netherlands–impact on the train driver, in: International Human Factors Rail Conference, London, UK, 2017.

[35] A. Naghiyev, S. Sharples, M. Carey, A. Coplestone, B. Ryan, Ertms train driving-incab vs. outside: an explorative eye-tracking field study, in: Contemporary Ergonomics and Human Factors 2014: Proceedings of the International Conference on Ergonomics & Human Factors 2014, Southampton, UK, 7–10 April 2014, CRC Press Boca Raton, FL, 2014, p. 343.

[36] T. Luke, N. Brook-Carter, A. M. Parkes, E. Grimes, A. Mills, An investigation of train driver visual strategies, Cognition, Technology & Work 8 (2006) 15–29.

[37] J. Edquist, T. Horberry, M. Regan, I. Johnston, 'visual clutter'and external-to-vehicle driver distraction, in: INTERNATIONAL CON-

FERENCE ON THE DISTRACTIONS IN DRIVING, 2005, SYDNEY, NEW SOUTH WALES, AUSTRALIA, 2007.

[38] C. Hölscher, Zusi bahnsimulatoren, braunschweig, Germany.
URL https://www.zusi.de/

[39] M. Johne, M. Busse, Railsite® (rail simulation and testing), Journal of large-scale research facilities JLSRF 2 (2016). doi:10.17815/jlsrf-2-144.

[40] W. S. Mokrzycki, M. Tatol, Colour difference De - a survey, MG&V 20 (4) (2011) 383–411.

[41] A. Shahid, K. Wilkinson, S. Marcu, C. M. Shapiro, Karolinska sleepiness scale (kss), in: A. Shahid, K. Wilkinson, S. Marcu, C. M. Shapiro (Eds.), STOP, THAT and One Hundred Other Sleep Scales, Springer-Verlag, s.l., 2012, pp. 209–210. doi:10.1007/978-1-4419-9893-4_47.

[42] M. Czamolewski, An empirical validation for the natural log transformation of reaction time, Proceedings of the Human Factors and Ergonomics Society Annual Meeting 40 (24) (1996) 1287. doi:10.1177/154193129604002471.

[43] R. H. Baayen, P. Milin, Analyzing reaction times, International Journal of Psychological Research 3 (2) (2010) 12–28.
URL https://dialnet.unirioja.es/servlet/articulo?codigo=3405162

[44] A. Gelman, J. Hill, Data analysis using regression and multilevel/hierarchical models, 10th Edition, Analytical methods

for social research, Cambridge Univ. Press, Cambridge, 2009. doi:10.1017/CBO9780511790942.

[45] J. J. Faraway, Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models, second edition Edition, Texts in statistical science, CRC P, Boca Raton, Florida, 2016. doi:10.1201/9781315382722.

[46] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, Journal of Statistical Software 67 (1) (2015). doi:10.18637/jss.v067.i01.

[47] S. Welham, B. Cullis, B. Gogel, A. Gilmour, R. Thompson, Prediction in linear mixed models, Australian & New Zealand Journal of Statistics 46 (3) (2004) 325–347. doi:10.1111/j.1467-842X.2004.00334.x.

[48] M. Pavlou, G. Ambler, S. Seaman, R. Z. Omar, A note on obtaining correct marginal predictions from a random intercepts model for binary outcomes, BMC Medical Research Methodology 15 (1) (2015) 59. doi:10.1186/s12874-015-0046-6.
URL `https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-015-0046-6`

[49] D. Schackmann, E. Bosch, Automated detection of train drivers' head movements: A proof-of-concept study, Automation 5 (1) (2024) 35–48.

[50] Eric T. Greenlee, Patricia R. DeLucia, David C. Newton, Driver vigilance in automated vehicles: Hazard detection failures

are a matter of time, Human factors 60 (4) (2018) 465–476. doi:10.1177/0018720818761711.

[51] C. D. Wickens, Multiple resources and performance prediction, Theoretical Issues in Ergonomics Science 3 (2) (2002) 159–177. doi:10.1080/14639220210123806.

[52] B. Milius, B. Cogan, B. Thomas-Friedrich, E. Bosch, H. Last, U. Metzger, D. Leinhos, Funktionale anforderungen an sensorik und logik einer ato-einheit.
URL https://www.dzsf.bund.de/SharedDocs/Fachmitteilungen/DZSF/2023/08_2023_Forschungsbericht_40_2023.html