The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2632-0487.htm

Remote control concept for automated trains as a fallback system: needs and preferences of future operators

Baris Cogan and Birgit Milius
Technische Universitaet Berlin, Berlin, Germany

Received 22 November 2022
Revised 12 May 2023
7 August 2023
Accepted 10 September 2023

Abstract

Purpose – Increasing demand on rail transport speeds up the introduction of new technical systems to optimize the rail traffic and increase competitiveness. Remote control of trains is seen as a potential layer of resilience in railway operations. It allows for operating and controlling automated trains and communicating and coordinating with other stakeholders of the railway system. This paper aims to present the first results of a multi-phased simulator study on the development and optimization of remote train driving concepts from the operators' point of view.

Design/methodology/approach – The presented concept was developed by benchmarking good practices. Two phases of iterative user tests were conducted to evaluate the user experience and preferences of the developed human-machine-interface concept. Basic training requirements were identified and evaluated.

Findings – Results indicate positive feedback on the overall system as a fallback solution. HMI elicited positive emotions regarding pleasure and dominance, but low arousal levels. Train drivers had more conservative views on the system compared to signalers and students. The training activities achieved increased awareness and understanding of the system for future operators. Inclusion of potential users in the development of future systems has the potential to improve user acceptance. The iterative user experiments were useful in obtaining some of the needs and preferences of different user groups.

Originality/value – Multi-phase user tests were conducted to identify and to evaluate the requirements and preferences of remote operators using a simplified HMI. Training analysis provides important aspects to consider for the training of future users.

Keywords Remote driving, User experience, Operator training, Railway automation, Resilience

Paper type Research paper

1. Introduction

Rapid technological development makes the implementation of advanced assistance systems and automated transport systems possible. In various sectors, a shift from manual human control to automatic control under human supervision and ultimately to autonomous automatic control is evident, occurring in gradual steps. High levels of automation in the

Funding: This paper has received funding from the European Union's Horizon 2020 research and innovation program under the grant agreement No. 815001.

Smart and Resilient Transportation
Emerald Publishing Limited
e-ISSN: 2632-0495
p-ISSN: 2632-0487
DOI 10.1108/SRT-11-2022-0018

railway domain are already a reality. Highly automated urban transport systems like driverless metros can be found all around the world. However, the implementation of such systems in mainline or freight operations poses different challenges due to more complex topology and greater risk of external influences.

For the grade of automation (GoA) levels 3 and 4, in which humans only have monitoring tasks or the train runs completely driverless, complex operational implications need to be considered. Until reaching fully autonomous systems, human operators will remain crucial for intermediate levels of automation. On the other hand, human operators may have an advantage over automation in dealing with novel situations due to their problem-solving skills and flexibility. In the railway domain, this is particularly important for degraded mode of operation due to technical problems or external influences. Current railway systems have various procedures for continuing the safe operation of trains or for allowing passengers and goods to change trains. However, increasing demand on rail transport speeds up the introduction of new technical systems to optimize the expanding rail traffic and increase competitiveness. Remote control of trains is seen as a potential layer of resilience in railway operations. It allows for operating and controlling automated trains and communicating with other stakeholders – technical and human – of the railway system. In the event of a failure, a remote operator can interact with the system remotely, thus increasing the flexibility of the response and decreasing the duration of disruptions. Another application of this system is to drive the train remotely after a failure in the autonomous driving equipment. Even though there are still technological limitations and operational challenges for controlling the trains over a distance safely, the potential of teleoperation as a fallback system in mid and long-term is evident.

The European Union (EU) has witnessed an improvement in railway safety through the implementation of safety regulations, as well as the progress of advanced technologies. For example, advanced rail safety technologies such as positive train control are designed to mitigate human error and improve operational safety. Yet, there are still implementation challenges such as interoperability and technological complexity and operational issues such as safety under restricted speed (Zhang *et al.*, 2018). Complex consequences of the impact of assistance systems on human performance also highlight the importance of designing safe and resilient railway systems. The number of significant rail accidents statistically decreased in 2020 compared with the average for 2016–2019 except for derailments and for collision of trains, which increased by 4% (ERA, 2022). A comprehensive risk analysis points to the risks of derailment in both mainline and yard operations with different service options and operation speeds (Di Kang *et al.*, 2023). In 2020, the estimated total cost of significant railway accidents in the EU-27 amounted to approximately EUR 3.2bn. Although this figure is lower than the figures reported in the two preceding years, it is still considered high. Fully automated train operations pose a highly intricate challenge within the domain of mainline traffic in terms of safety and system resilience. Considering the increased operational and technological complexity through automation, fallback systems such as remote operation could provide crucial support for ensuring safe operations.

This paper presents the development concept, optimization phases and the first results of the pilot experiments on remote driving of trains as a fallback system. Technical realization of the data transmission is outside the scope of this study. In phase I, different Human-Machine-Interface options were analyzed and an initial concept for remote operation was created. The phase II simulator experiments were designed to assess the initial concept as well as to evaluate two different driver desk controllers. In phase III, the workstation was optimized based on the results of phase II experiments. A training program was developed to increase awareness of the system. In phase III, the remote driving concept, the optimized workstation and the training program were evaluated by different user groups.

The remainder of this paper is organized as follows: Section 2 of the paper briefly introduces the teleoperation concept in railway domain. Research gaps are identified through a literature review. In Section 3, the materials and methods used in each study phase are presented. In Section 4, the results of the iterative user tests are revealed. The findings and the limitations of the study are discussed in Section 5. Finally, concluding remarks and outlook are provided in Section 6.

2. Related works

In traditional railway operations, the train drivers onboard are the initial point of contact for detecting and responding to any irregularities, acting as a fallback level for the technical system. Therefore, with the absence of driving personnel, solutions to this problem must be found for implementing highly automated railway systems. On the other hand, there might be several aspects that cannot be replaced by automated systems completely, such as evacuations in connection with fire, due to technological challenges or regulatory barriers (Hagemeyer *et al.*, 2021). Operational concepts regarding the fallback level in GoA 3 and GoA 4 systems, thus, gain importance. In GoA 3, the safety responsibility of the routine operation is transferred to the system. Nevertheless, during disruptions or emergency situations, the train attendant may play a role in mitigating failures or facilitating evacuations (Hagemeyer *et al.*, 2021). Additionally, the train attendant could utilize mobile devices like tablets to monitor the train's status on or offboard (Adebahr *et al.*, 2023). The GoA 4 system, on the other hand, envisions a fully automated operation. However, in the short and midterm, it can also be expected that operational control center personnel will supervise the system when necessary. In light of these needs, together with the experiences from other transport domains, remote operation (or teleoperation) concept has gained attention from the railway academia and industry. The remote operation could replace or complement the train attendant in GoA 3, or can be a fallback system in GoA4. The tasks of a remote operator can range from monitoring train status and communicating with third parties to failure mitigation or driving the train remotely. (Brandenburger and Naumann, 2018).

Compared to the aviation and automotive domains, there has been relatively less research related to remote driving (or teleoperation) in railways. In the railway domain, there are several studies and demonstrations on teleoperation. A project aimed at demonstrating the safe operation of remote driving investigated the use of different transmission technologies (Masson *et al.*, 2019) and the interaction and cooperation between remote driver and assistance systems (Gadmer *et al.*, 2021). The former identified several factors that could impact remote driving task, such as the time delay and bandwidth issues, limited field of view of the camera, degraded depth perception and the information loss due to the distance between the remote driver and the train. The latter study presents a two-phase process to support the understanding of the interaction between remote driver and assistance systems by a cooperation framework. There are also attempts to define the operational design of this system and to propose a risk assessment methodology (Tonk *et al.*, 2021), and to combine safety and security risk analyses for remote-driven rail systems (Aktouche *et al.*, 2021).

Human performance, human error, and workload experienced by the user can be influenced by the design of the Human-machine interface (HMI). To reduce human error, it's important to consider user expectations, previous training, and experience. Naumann *et al.* (2013) highlight three key methods of user-centered design: understanding the user, evaluating existing systems, and testing new systems. Simulator and field studies can support efficient HMI development. A study presented an example of remote supervision and manual control by a train operator in a simulator study (Brandenburger and Naumann, 2018). A workplace prototype received positive feedback from train drivers in terms of acceptance ratings and perceived benefit in terms of system resilience. However, the satisfaction while working with the system did not show positive changes. This result highlights the need to further investigate the needs and preferences of future operators in the earlier phases of the development. There were also field trials on remote driving in recent years. A test train without a driver on board was remotely controlled using 5G technology in Germany (Melzer, 2020). The test scenario was to drive the test train remotely via the 5G mobile network from the shunting area to the platform. French railway company SNCF also performed a trial of remote driving in France (Zasiadko, 2019). A simulator study within the EU-funded project Drive2theFuture has been investigating the development and optimization of a remote driving concept with the help of iterative user tests. A previous publication attempted to reveal the challenges and new requirements arising with the increasing level of automation and to set the roadmap for a series of simulator experiments on remote operator's needs and preferences (Cogan and Milius, 2021). The priory risk assessment conducted in the mentioned work identified several risks, such as limited functionality of a remote driving workstation compared to in-cab driving, trust or acceptance issues of future operators and standardization issues. A recent publication based on an online survey investigated the passenger acceptability of teleoperation in railways (Cogan *et al.*, 2022). Potential opportunities for higher acceptance included the increased resilience and reliability in rail service, while safety and security concerns of prospective passengers were highlighted as two of the most important influencing factors.

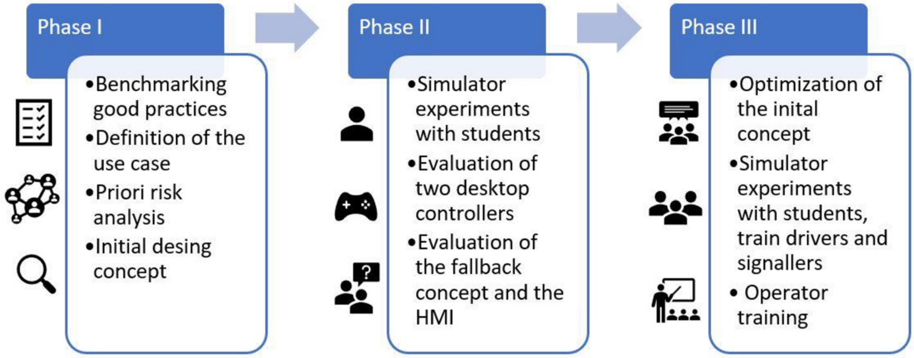
There are many research questions regarding the technological feasibility, the technical layout, and the operational procedures and processes. Research effort on automated transport systems is mostly focused on technical aspects. The human's role in future systems and human operator's needs and preferences are often overlooked, which highlights the importance of the present research. Literature research has demonstrated the significance of involving potential operators in the development of systems. Iterative user testing, such as the one used in the present study, allows us to tailor the developed concept to the users' needs. Deploying digital and automated technology into railway operations raises questions on the roles of the railway personnel. These changes in the characteristics of the tasks of the train drivers and rail traffic operators need to be considered. This change will also result in the need for developing suitable training programs for future operators. Although there are few considerations regarding the task allocation of future systems, literature lacks sufficient effort in training for operators.

3. Materials and methods

This section details the procedure followed for the iterative user tests. The first phase comprises defining use cases and benchmarking of existing practices. In phase II, the design concept is developed into a simple prototype for the iterative user tests. Phase III aims to evaluate the optimized HMI based on the feedback from different user groups (Figure 1).

3.1 Phase I: system definition

First, application cases for evaluating the remote operation concept were determined. Two use cases were selected for the analysis. The first use case includes the development and optimization of an HMI for remote driving. The regular operation is assumed to be a driverless Automatic Train Operation (ATO) with Grade of Automation levels 3 and 4 (UITP, 2018). The operational scenario for this use case is to drive an automated train from a remote-control center after a system failure. The operational procedures need to be defined depending on the type of fault and the time dependency of the disruption. Response to accidents or disruptions that result in an immediate danger to passengers or goods are not



Source: Authors' own work

Figure 1.
Stages of the iterative user testing

considered in the use case. For this defined use case, a simplistic solution for the restricted continuation of the journey is intended to be analyzed to obtain feedback from future users on the remote driving concept and the remote workplace functions. The second use case focuses on the training requirements for future operators. Based on the task characteristics and performance challenges, basic training requisites are identified.

The first step of the design process is to identify existing systems and good practices. A benchmarking method was used within the project to identify good practices amongst pilot demonstrators as well as active or prototypical vehicles of different manufacturers or transport systems (Mathis *et al.*, 2020). Following the identification of related HMIs or systems, best practices are evaluated using expert-rated analysis. An example for a part of the good practices analysis on the automated metro in Nuremberg given in Table 1 below. The full analysis also includes attributes such as the task description, automation levels, feedback on the automated vehicle status as well as audio-visual design properties.

Based on the defined scenario, an initial design of the workplace and tasks of the remote driver was determined. Potentially, the job profile of the operating person can be a mixture of train driver, during manual control, and signal box operator, during the automated operation. Identifying the information the remote operator needs to be always aware of and

Description of HMI element/principle (Brux, 2007)	Indispensable HMI element for function	Not linked to unique situations and specific circumstances
Emergency steering device in vehicle	x	x
Past event memory to estimate a dysfunction		x
Fenced track and safety system for platform lines	x	x
Signal box operator reacts to emergency messages from passengers		x
Vehicle safety devices (fire, obstruction and derailment detectors)	x	x
Daily test runs operated by human drivers	x	x

Source: Authors' own work

Table 1.
Good practices analysis on the automated metro in Nuremberg (excerpt)

to process is crucial for the design of the information interfaces for the developed HMI. Basic information requirements and functionalities of the workplace were identified.

Traditional train driving is a dynamic task that relies on information acquisition and processing. This is achieved by information sources such as infrastructure, environment and train in-cab display. Table 2 summarizes the basic train-driving requirements and information needs. Some of this information is inherent in the task, and some is conveyed by the environment as external cues such as weather conditions. During teleoperation, the operator is physically separated from the vehicle and the local driving environment. A remote workstation lacks some of these external cues or information sources. Physical distance to the vehicle in teleoperation might also cause issues with situation awareness (Linkov and Vanžura, 2021). Therefore, a wide variety of multimodal (audio, voice, gesture, haptic) displays and control elements can be used to provide feedback and information to operate the train safely. Changing task characteristics with the remote operation will require new information sources as well. For example, the allowed speed might be dependent on the latency in the data transmission. Data transmission quality and system reliability could be additional information that operators benefit.

Existing railway regulations were also analyzed to find procedures that can be used directly or can be adapted to the use case. An example is a disruption management (e.g. external disruption) in the train driving regulation in Germany (DB Ril 408), according to which the train driver must clear the affected section of track on sight with a maximum speed of 40 km/h. Some of the potential tasks of a remote operator include start-stop functions, door closure, communicating with passengers and other stakeholders, failure diagnosis and mitigation and remote driving.

The active intervention should take place only when it is requested by the system or by the operator. Continuous supervision by a remote operator for one train wouldn't be efficient. If the operator must monitor the ride without an active task over a longer period, this might cause performance decrements (Onnasch *et al.*, 2014).

After identifying basic information and task requirements of remote operation, the initial design concept was developed for iterative user tests.

3.2 Phase II: design concept

In phase II of the study, the initial design concept was developed into a remote workstation. Train operation in a high level of automation (GoA3/4) is remotely controlled from an operation center. In the case of a disruption and without any railway employees on-board, the train must be manually controlled from the Operations Centre.

3.2.1 Phase II experiments. A remote workstation in the EBuEf (Railway Operations and Experimentations Center) was used for the experiments. A sketch of the workstation is given in the left panel of Figure 2. Remote train drivers were provided with a driver desk by

Table 2. Train driving basic information requirements

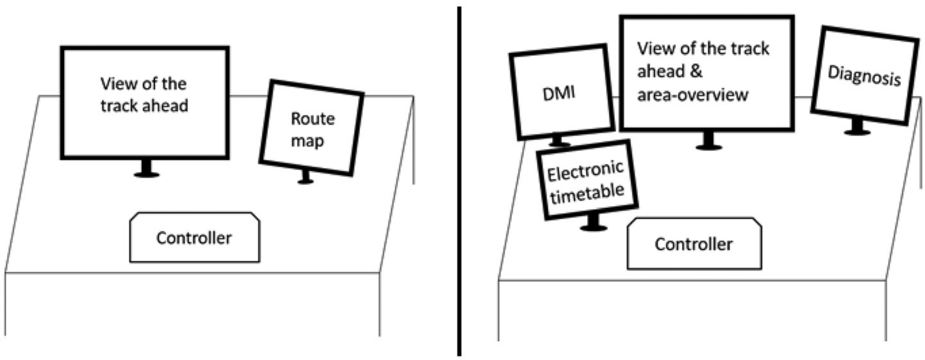
Route knowledge*	Train state indicators*	Environmental state*	Additional info for teleoperation**
Gradients	Speed	Visibility	Multi-angle camera broadcast
Curvature	Passenger/freight	Auditory feedback	Data transmission quality
Signals	Freight type	Aerodynamics	System reliability
Speed restrictions	Fuel usage	Infrastructure condition	Failure diagnosis
Landmarks	Weight	Weather	Past event memory

Note: Adapted from Naweed *et al.* (2013) (*) and extended (**)
Source: Authors' own work

RailDriver – a fully functional model of an American diesel loco – and the front camera view of the tracks on a monitor as well as necessary information such as speed, traction and brake force and a route map. Additionally, a joystick-button controller was built as an alternative controller. Each participant conducted the same ride using two different controllers. The information on traction and brake force was displayed on an LED screen located on the controllers. The visual warning of the dead man's switch was also indicated on the same display. The train ride was simulated on a Berlin suburban train line using in-house developed software (SimMetro). 13 students who have knowledge of railway operations participated in the experiments. Participants were randomly assigned to two experiment groups. The group RD completed the experiment scenario using first the RailDriver and second the joystick controller. The group JS completed the rides vice versa. The participants were informed about the aims of the project. This research complied with the American Psychological Association Code of Ethics or tenets of the Declaration of Helsinki. Informed consent was obtained from each participant. A brief explanation was provided regarding the use of a driving simulator. The experiment started when the participant felt comfortable using the simulator.

In the experiment use case, the automated train is stopped due to a system malfunction. The subjects were asked to drive the train remotely from the place of disruption to a pre-determined location. The route length was approx. 9 km and the participants were asked to maintain a speed of 40 km/h. The tasks included the departure, on-sight driving, attending to the dead man's switch at regular intervals and stopping at a predetermined location. One experimenter was present in the same room for the whole duration of the experiments.

3.2.2 Evaluation. Right after each ride, participants completed the evaluation surveys provided by the experimenter in pen-paper format. The survey included a section for evaluating the user experience for both rides with the help of widely used User Experience Questionnaire (UEQ) designed by Laugwitz *et al.* (2008). UEQ is composed of 26 six items on different aspects of user experience and usability (e.g. understandable, enjoyable) that represent 6 dimensions, namely Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Individual HMI elements (i.e. desktop controller, monitor, seat, information) were rated using a scale of −2 to +2. Another section of the survey included a questionnaire for measuring how persuasive and affective the HMIs are. The questionnaire with a 7-point Likert scale (strongly disagree-strongly agree) is adapted from an earlier study (Thomas *et al.*, 2019). Effectiveness, quality and capability of the workstation in terms



Source: Authors' own work

Figure 2. Sketch of the remote workstation design elements in phase II (left) and phase III (right)

of persuasion are measured by the following statements in the corresponding order; "*After viewing this system, I will make changes in my attitude towards automated technology*.", "*This system is trustworthy*", "*This system has the potential to influence user behaviour*."

3.3 Phase III: optimization and training

In phase III of the pilot study, the remote workstation that was used in the phase II was optimized based on the analysis of the experiment results. Additionally, the tasks of the operator and the functionalities of the workstation were extended. The main goal of the phase III is the evaluation of the teleoperation concept by different user groups as well as the investigation of the impact of training on user acceptance and awareness.

3.3.1 Optimization. The scenario of phase III defines two responsibility areas for the remote operator, namely, supervision and manual control. During the automated operation, the remote operator merely supervises the operation without an active intervention. However, the supervision task does not include continuous monitoring. The remote operator is informed by the automated system when there is a need of manual intervention. According to a pre-defined procedure for certain disruptions, the remote operator takes the manual control of the train. In phase III experiments, the disruption is defined as a sensor failure. The remote operator has to manually drive the train from the point of disruption to the specified location (i.e. next station).

The remote workstation that was used in the phase II was optimized based on the analysis of the iterative user tests as well as on the development process of phase III (Figure 2, right panel). The train movement (Siemens ES64U2 electric locomotive) was simulated using the commercial simulation software Zusi. As the train control panel, the controller that received higher preference scores in the phase II was selected (commercially available desktop controller; RailDriver). Additionally, the workstation was developed to be more interactive, with touchscreens and different user interfaces. The camera view from the head of the train was provided by 32-inch UHD curved monitor. During the supervision mode this screen displayed the area overview of the responsibility area. The area overview includes the tracks, signal and station locations as well as real-time locations of the trains in the area. For the electronic timetable and control command display, the interfaces of the simulation software were displayed on touchscreens. The system mode of the active operation was displayed on a separate monitor (i.e. diagnosis display). The icons and the warning sounds of the three modes were developed within a common framework of the project. There are three icons informing about the system mode: automated mode, manual request and manual mode. The request for each mode change is done by the system via audio-visual warnings and is confirmed by the operator. The same monitor also provided the system diagnosis (e.g. reason of the disruption) as well as instructions for the remote operations (e.g. manual driving until specified location).

3.3.2 Training. In order to identify the basic training needs of future operators, literature research was conducted. The review identified several challenges regarding the technology and human factors of teleoperation. Main findings of the literature review are summarized in Table 3. Basic requirements are determined based on the analysis of operator tasks and performance challenges. A classical presentation-based training material was prepared. First, railway automation and the remote driving concept is introduced. Secondly, advantages and challenges of the remote operation are discussed. Operational challenges and performance implications are detailed under the categories of technical issues and human factors issues. Finally, mitigation strategies for identified issues are discussed. A summary of the contents of the training is given in Table 3. Besides this training, an E-Learning Website was developed within the project. The website offers two additional courses on the rail domain: namely, introduction to autonomous rail, and train traffic safety

Railway automation and remote driving	Technical challenges	Human factors challenges	Human-centred design	Training
Automation levels	Latency and transmission	Limited info availability	Inclusion of future users in development phases	Use of simulators and VR-tools
Remote driving concept and use cases	Data delay and quality	Reduced field of view (e.g., blind spots)	Improving the field of view (e.g., multiple camera views)	Training on emergency management
Remote driving advantages and opportunities		Speed and depth perception	In-cab signalisation	Training on perception and interpretation of new information sources
Instructions on how to use pilot HMI		Risk perception	Advanced driving assistance systems (e. g., obstruction detection)	Non-technical skills training (e.g., problem solving, memory, communication)
Landmarks		Impaired situation awareness	Multimodal information and feedback sources	

Source: Authors' own work

Table 3. Training modules and contents for remote operation by train drivers and signalers

and safety management for train dispatchers/signallers. At the end of each course, there is a quiz for self-assessment.

3.3.3 Phase III experiments. In this third phase of the pilot project, three user groups were included, namely train drivers, train dispatchers or signallers and students or members of the railway department of the TU Berlin. These users are referred in following as user groups: train drivers, signallers and others. The first part of the participation included an online training which was conducted with half of the subjects. Participants were randomly assigned to two experiment groups. These participants were also instructed to visit the E-Learning Website developed within the project. This group of participants evaluated the training activities with an online questionnaire. The simulator experiment started with the supervision task during which the remote operator could monitor the area overview until a system disruption occurred. After around 2–3 minutes of supervision, the main monitor and the diagnosis monitor displayed the occurrence of a system failure and the instructions to follow. The purpose of the supervision part was to introduce the participants another potential task that a remote operator could assume. The remote operator then took the manual control of the train, pushing the respective buttons on the interactive user interface and drove the train remotely using the desktop controller and other information interfaces. The ride was conducted on-sight with an allowed maximum speed of 40 km/h. The manual task ended after arriving at a pre-determined station. Each participant conducted two supervision phases (2 min. + 2 min.) and two manual ride phases (5 km + 10km) in one session of total around 45 minutes. The remote operator could confirm the transfer of control (i.e. from manual to automated and vice versa) using the interactive interfaces. The switches between the manual and automated modes were designed as the Wizard-of-Oz system with an experimenter always present in the same room (Plate 1).

3.3.4 Evaluation. An HMI evaluation questionnaire (HMIQ) was developed within the project to collect user feedback on the HMI (Appendix 1). User experience questionnaire that

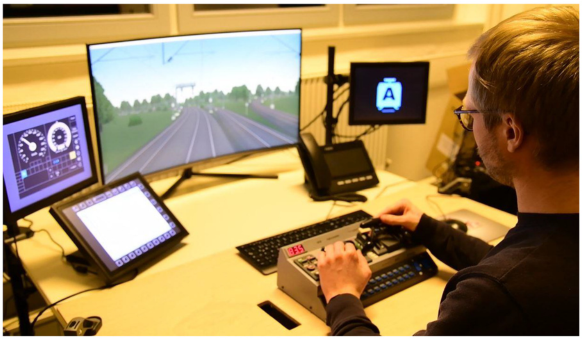


Plate 1.
Phase III experiments on the simulator

Source: Authors' own work

was used in the phase II experiments was also used in the phase III experiments. A non-verbal pictorial assessment technique, called self-assessment manikin (SAM), is used to measure pleasure, arousal and dominance associated with pilot users' affective state during the test (Bradley and Lang, 1994). Two methods were adapted for the measurement of users' situation awareness. First, a freeze-probe technique called SAGAT (Appendix 2) was employed at the end of the first ride and subjects were queried as to their perception of the situation. In total, 7 queries correspond to the three stages of SA (perception, comprehension, and projection). In order to cross-check the SAGAT data, perceived situation awareness was collected post-trial using a subjective assessment technique. A self-rating measure of SA was adapted to this pilot study (Braarud, 2021). Three items were developed to represent each of the three levels of situation awareness. The items were presented as statements, and a 7-point Likert scale was given for the ratings. Additionally, the mental workload was measured using self-reported surveys. NASA-TLX is used for the mental workload assessment, as this is a widely used technique in various research areas (Hart and Staveland, 1988).

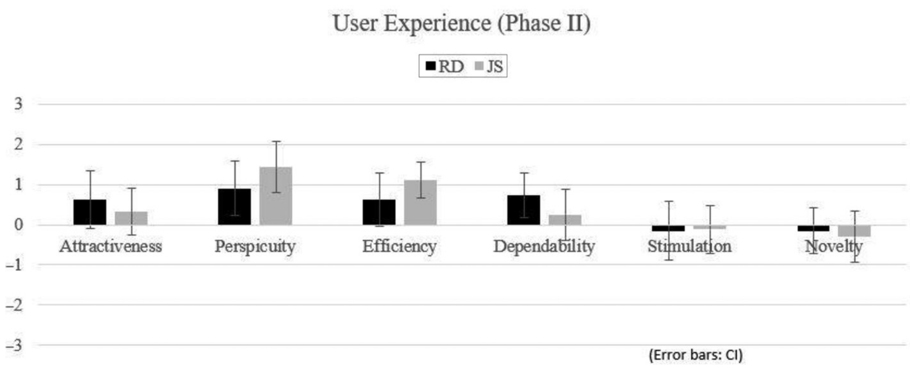
4. Results

4.1 Phase II

13 students which knowledge of railway operations participated in the experiments. All participants had heard of autonomous vehicles but less than half (46%) had experience with automated vehicles or autonomous systems (e.g. driverless metro). In Figure 3, the mean scores of user experience evaluation for each user group are given. Mean values between -0.8 and 0.8 represent neutral evaluation of the corresponding scale. There was no statistically significant difference between the groups for any of the dimensions at the alfa level of 0.05.

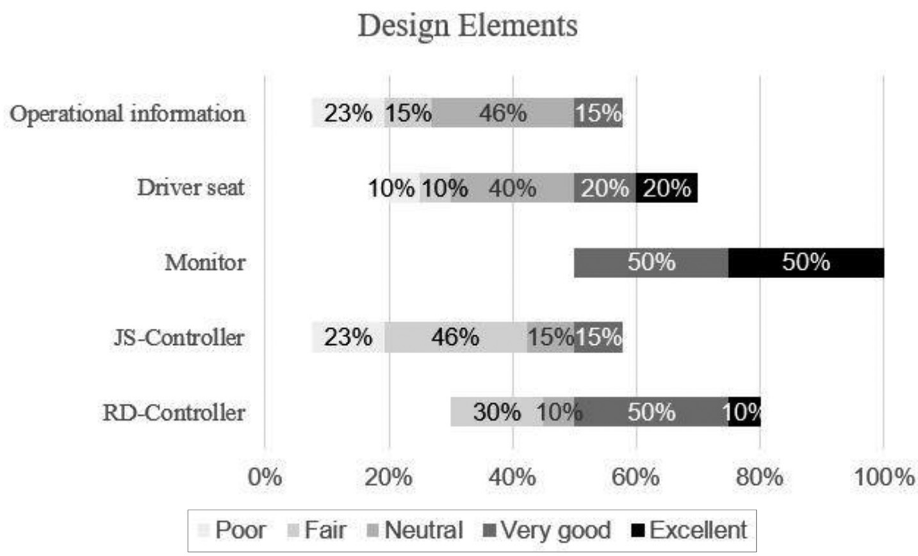
Individual design elements were rated using a scale of -2 to $+2$ (Figure 4). The preferred controller was the RailDriver (alpha $= 0.05$, $p < 0.01$, $t = 2.67$). Optionally, written suggestions for additional functionality recommendations could be given. The most frequently mentioned answer was the need for audio warnings and other audio input from the location of the train. Another frequent answer was the front-view monitor as the preferred location of speedometer instead of the LED indicator on controllers. Some participants recommended including more operational information regarding timetables and travel times.

The quality dimension of the persuasiveness potential of the HMI was rated highest by subjects, which indicates high trust to system functions (Figure 5). However, potential effectiveness on changing one's opinion was rated lowest (66% negative), even though the capability to influence others was rated high by around half of participants.



Source: Authors' own work

Figure 3.
Mean values of the user experience scales in phase II experiments



Source: Authors' own work

Figure 4.
Distribution of user preference ratings for individual HMI elements in phase II

4.2 Phase III

26 participants took part in three user groups: 7 train drivers, 11 train dispatchers or signalers and 8 students or members of the railway department of the TU Berlin. All participants had heard of autonomous vehicles before, while only around 46% had personal experience with autonomous vehicles. These were mainly with driverless urban transport systems such as Nuremberg and Paris metro. Only half of the participants had heard of remotely driven trains prior to the pilot experiments. %55 of the participants (except students) had 0–3 years of professional experience. The other participants were equally distributed to the work experience groups of 4–7 years, 8–10 years, and more than 10 years.

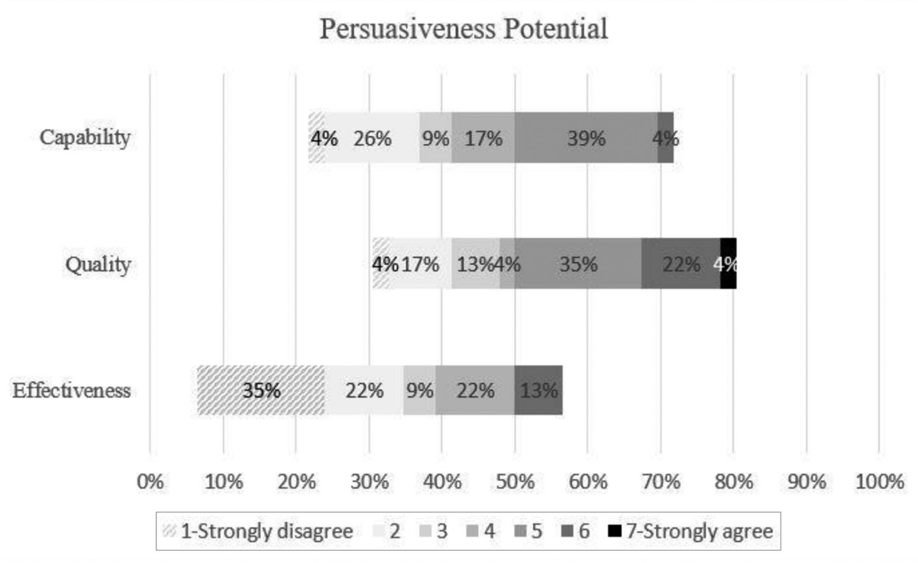


Figure 5.
Persuasive and affective HMI: response distribution in phase II

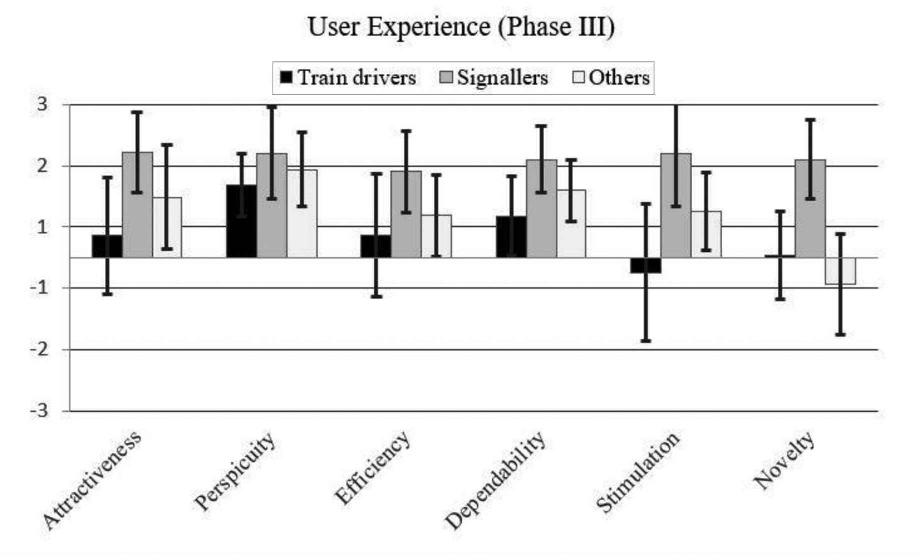
Source: Authors' own work

The HMIQ includes questions for evaluating the user experience (UX) as well as scales for rating the individual design elements. In Figure 6, the mean scores for each user group are given. Mean values between −0,8 and 0,8 represent neutral evaluation of the corresponding scale. Two sample *t*-test indicated that the differences between drivers and signalers in terms of attractiveness ($p = 0.042$), stimulation ($p = 0.018$) and novelty ($p = 0.007$) are significant.

In literature, Likert item data (answers to single items) are often treated as ordinal data whereas Likert scale data (means of the results on sets of items) could be used as interval scale. Therefore, mean values alone might not be sufficient as a measure of central tendency for individual Likert items. For this reason, except for the user experience analysis, the main measure for Likert item data was chosen as the distribution of responses (e.g. % that agree) rather than mean values alone. The mean values and other descriptive statistics should be considered together with the percentage or frequency data when interpreting the results. These types of data are treated as ordinal and non-parametric statistical tests have been employed to test the hypothesis.

Individual responses to the self-manikin scale are plotted on a two-dimensional pleasure-arousal graph based on the model of Russell and Lanius (1984). This model includes only pleasure and arousal dimension, as it regards dominance as a cognitive indicator rather than as indicators of affect. With reference to the model in the upper-left panel of Figure 7, the experimental data concentrates mainly in the pleasant-low arousal area (Figure 7).

For measuring how persuasive and affective the HMI is, the same questionnaire as for the phase II experiments was used. Figure 8 illustrates the distribution of responses to the persuasiveness scale items. The capability dimension has greater positive distribution among all user groups, while the effectiveness has the bigger proportion of negative ratings. Kruskal-Wallis Test (KW) was employed to test whether the differences were statistically significant, with the null hypothesis of equal medians of all user groups. The KW test and a



Source: Authors' own work

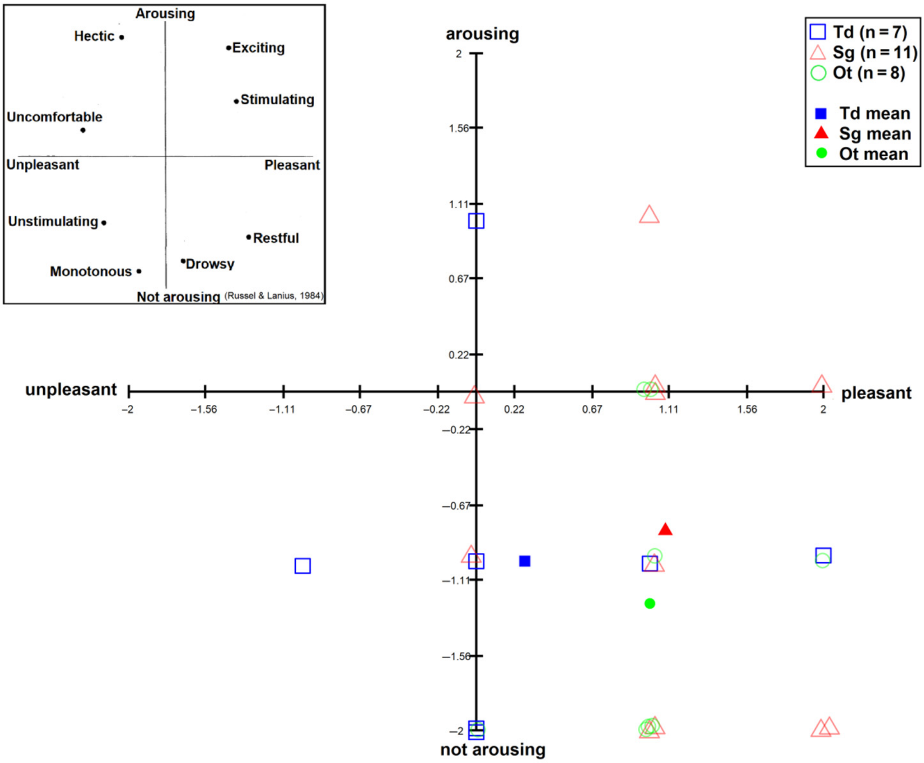
Figure 6.
Mean values of the user experience scales in phase III (error bars: 5% CI)

post-hoc Dunn's test using alfa of 0,05 indicated that there is no significant difference in any of the dimensions. (Effectiveness $\chi^2(2) = 4.38$, $p = 0.112$, $\eta^2 = 0.1$; Quality $\chi^2(2) = 4.18$, $p = 0.124$, $\eta^2 = 0.1$; Capability $\chi^2(2) = 2.14$, p = 0.343, $\eta 2 = 0.01$).

Similar to the phase II experiments, individual design elements were rated using a scale of −2 to +2 (Figure 9). While each element was rated positively for signalers and other user groups, operational information and the controller interface were rated lower than other elements. The Kruskal-Wallis H test indicated that there is a significant difference in the dependent variable of controller interface between the different user groups, ($\chi^2(2) = 8{,}27$, $p = 0{,}016$) with a mean rank score of 7 for drivers, 17 for signalers, 14,4 for the other group. The post-hoc Dunn's test showed that the difference in the mean rank of train drivers and signalers was significant ($p = 0.004$), while for the pair of train drivers and the others at near significance ($p = 0.049$). For the operational information element, only the driver-signaler pair was statistically significant ($p = 0.042$, $\eta 2 = 0.13$).

Self-assessed mental workload ratings are collected right after the experiments. Unweighted overall workload scores are 26.2 for train drivers, 38.1 for signalers and 26.6 for the other group. Because of the relatively small sample size, determining the distribution of the workload variable was important for choosing an appropriate statistical method. The Shapiro–Wilk test failed to reject the null hypothesis ($p = 0.078$), which indicates that the variable may be normally distributed. One-way ANOVA showed that the differences between the user groups are not significant ($p = 0.398$). For all groups, the results indicate low levels of workload, which are significant compared to the hypothesized average workload score of 50.

Mean scores of correct responses in SAGAT, and mean ratings for the self-assessed SA items are given in Figure 10. Likert-ratings were converted to percentages in order to compare both measurements. Overall, the subjective assessment indicates high SA levels with the ratings of all groups over 70%. The scores of the stage 1 and 2 for train drivers in both measurement methods are consistent. However, SAGAT assessment did not include stage 3. Even though this method provides a global testing of SA, it is subject to memory



Source: Authors' own work

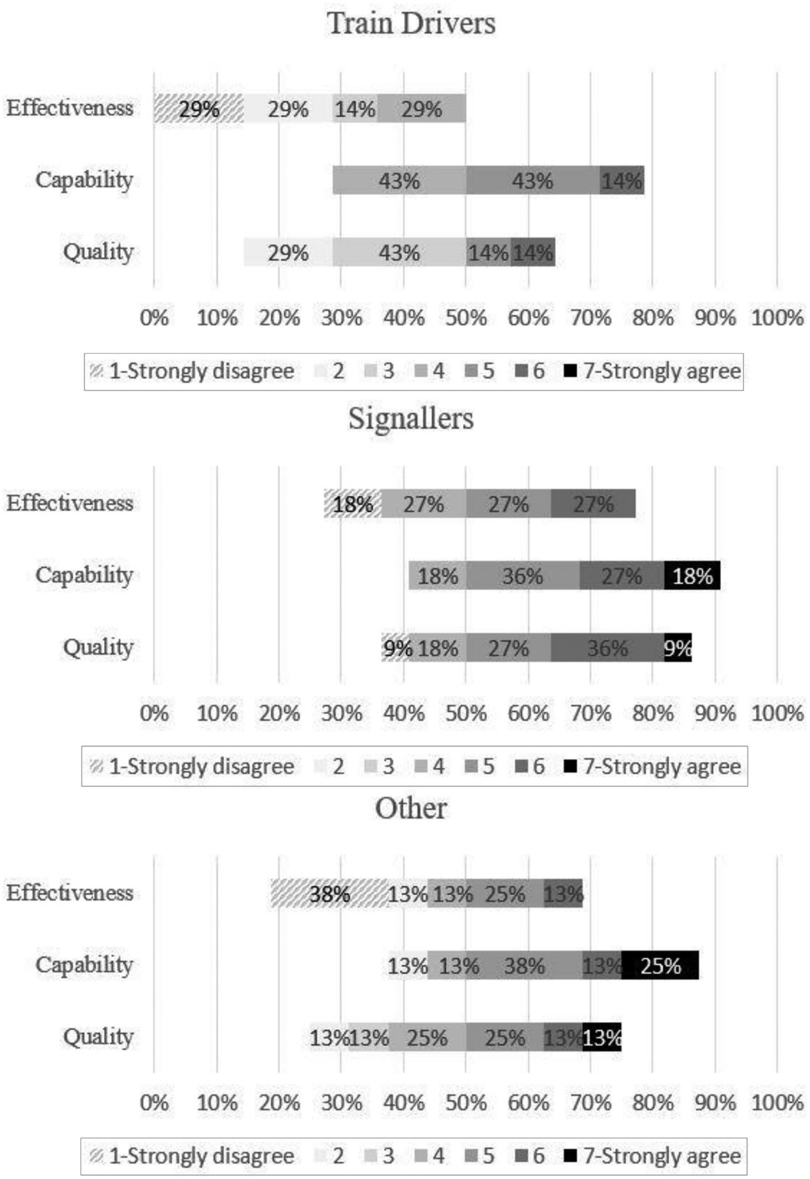
Figure 7.
Responses to SAM plotted on an arousal-pleasure graph, Td: Train driver, Sg: Signaler, Ot: Others, model of the affective appraisal of environments (Russell and Lanius, 1984) (top left corner)

decay. It should also be noted that the queries are made only once during the experiments. This approach minimizes the possible bias of attention for queries in advance, but also reduces the sensitivity of the measurements.

The general opinion on the quality of the training was rated 3.5 on a scale from unacceptable (1) to outstanding (5). This score was statistically significant compared to the hypothesized mean of 2.5. The questionnaire also included specific questions to capture various aspects of the method. These were rated on a scale from strongly disagree (1) to strongly agree (5) (Figure 11). Ease of learning and using as well as increased awareness and understanding of the system were rated high by the majority. However, trust and willingness to use had large distribution of neutral ratings. It is possible that the content of the online training related to the technological challenges of remote operation and their impact on operator performance have caused ambiguous opinions on applicability and willingness to use. The training content as well as the method of asking quiz questions at the end received positive feedback. Additional recommendations include more detailed information on train driving techniques and training in risk management skills regarding personnel and passenger management in cases of emergencies.

5. Discussion

The findings are discussed for phase II and III separately.



Source: Authors' own work

Figure 8.
Persuasive and affective HMI: phase III response distribution for train drivers (top), signalers (middle) and others (bottom)

5.1 Phase II

The phase II experiments showed positive feedback from subjects on the overall system as a fallback solution. Despite the preference ratings that favored the RailDriver, the evaluation did not show significant differences between the rides with two different controllers in terms

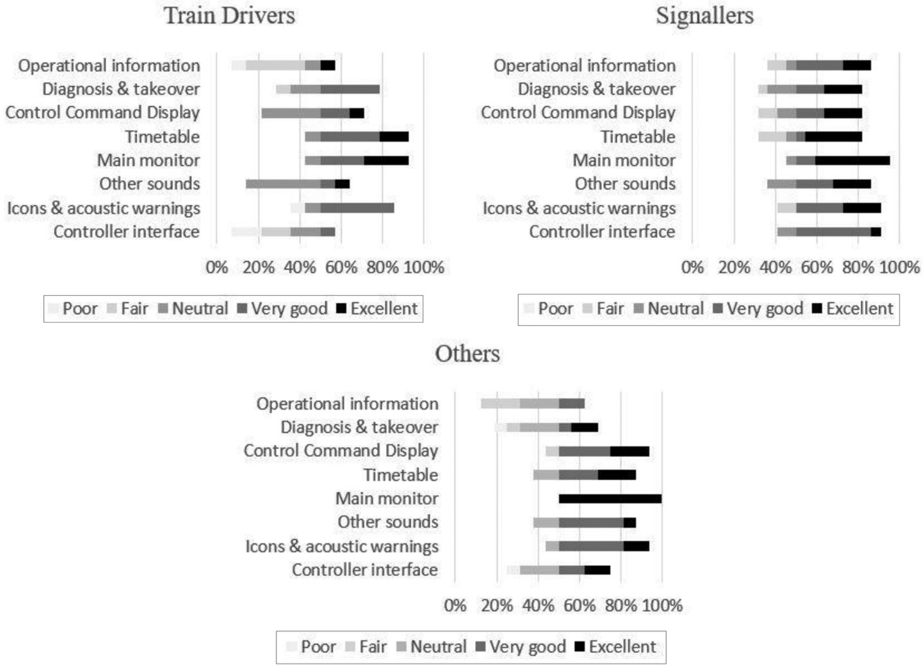


Figure 9. Distribution of user preference ratings for individual HMI elements in Phase III

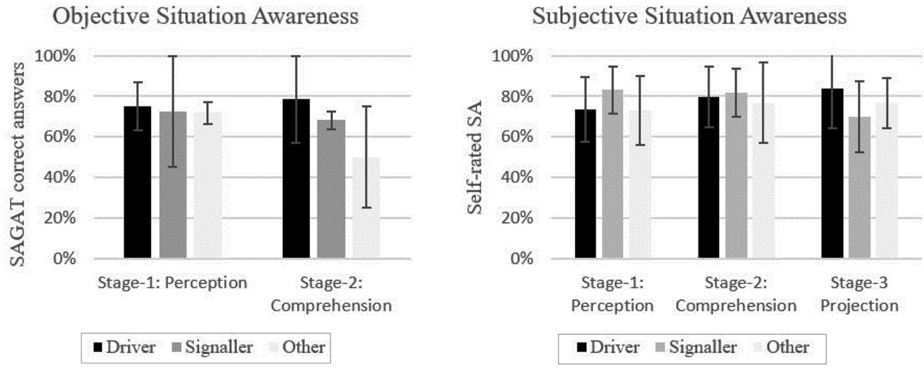
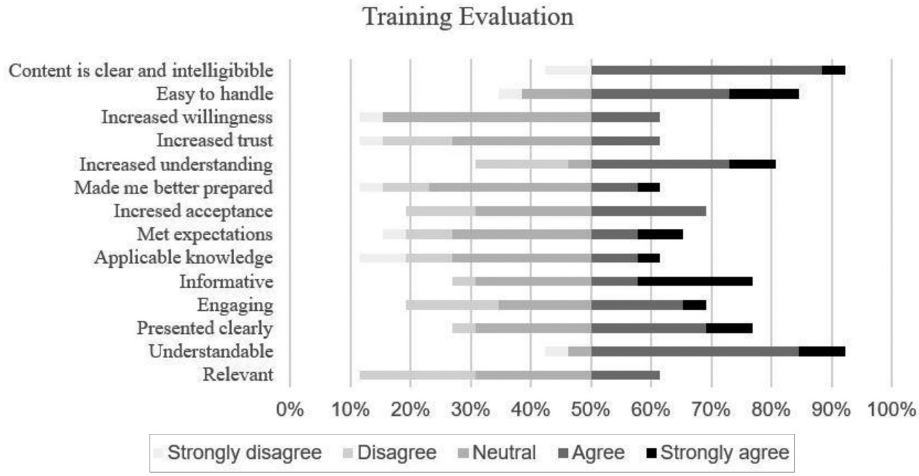


Figure 10. Objective (left) and subjective (right) situation awareness (error bars: SD)

Source: Authors' own work

of user experience. The flexibility of hardware choice could be an advantage in future implementations. However, the interviews also suggest participants would be inclined to use RailDriver rather than the custom-made Joystick-button controller. The persuasion quality of the HMI based on the system trust was rated positive by the subjects. Subjects recommend using more detailed and varied operational information. Since the participants were not experienced drivers, the feedback provided might not be representative of the functional expectations of users. Nevertheless, the findings of the experiment were useful in



Source: Authors' own work

Figure 11.
Training evaluation questionnaire response distribution

obtaining some of the needs and preferences of users, which are used to further optimize the system in the next phase of experiments.

5.2 Phase III

The phase 3 experiment results indicate positive feedback on the overall system as a fallback solution. Overall, it can be interpreted that train drivers rated pragmatic quality (i.e. task-related aspects) of the system higher than the hedonic quality (i.e. stimulation, originality), which falls into the neutral area. UEQ results suggest that train drivers have more conservative views on this technology, with lowest mean scores on nearly all scales. The HMI elicits positive emotions regarding pleasure and dominance, but low arousal levels. This finding is parallel with the low subjective workload scores reported by the participants. Lower than average workload ratings could indicate free capacity for additional tasks. In real world applications, the remote operator would have a higher variety of tasks and responsibilities, which would consequently increase the task demand of the operators. Analysis on operator's situation awareness indicated that the workplace functions and provided information were sufficient for building awareness. However, it is important to note that these results are based on the specified use case only.

The training activities achieved increased awareness and understanding of the system for future operators. The findings of the user experiments were useful in obtaining some of the needs and preferences of different user groups. Ideally, the remote operator training to be provided should be adjusted based on the knowledge and experiences of different user groups. While an improved driving desk and enhanced information sources would increase the acceptance and trust to the HMI, the simplified solution for the fallback system received positive ratings.

6. Conclusion

The objective of this project was to determine the needs and preferences of future users of automated transport systems. This paper presented the first results of the multi-phased simulator study on the development and optimization of remote train driving concepts from the operator's point of view. An HMI prototype was developed in order to provide remote manual intervention in cases of system disruption for the grade of automation (GoA) 3 and 4. The study also explored the requirements of the new job profile. The initial design concept, defined in phase I, was optimized by iterative user tests. Phase II experiments helped evaluate the remote driving concept and to select a preferred train control panel. The workstation and its functionalities were expanded in the third phase. Additional to non-expert users in phase II, train drivers and signallers were also included in the third phase. Results showed that inclusion of potential users in the development phase of future systems has the potential to improve user acceptance. Basic training requirements were identified, and the developed training program received positive ratings from users.

As the next step, the performance of the developed system will be evaluated based on related Key Performance Indexes defined by the project consortium. Recommendations on policy regulations and automated transport acceptance roadmap will be developed.

References

Adebahr, F.-A., Milius, B. and Naumann, A. (2023), "Flexible arbeitsumgebungen für die ATO-Rückfallebene", *Der Eisenbahningenieur*, Nos 1/23, pp. 39-41.

Aktouche, S.R., Sallak, M., Bouabdallah, A. and Schön, W. (2021), "Towards reconciling safety and security risk analysis processes in railway remote driving", *paper presented at 2021 5th International Conference on System Reliability and Safety (ICSRS)*.

Braarud, P.Ø. (2021), "Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human–machine work", *International Journal of Industrial Ergonomics*, Vol. 86, p. 103233.

Bradley, M.M. and Lang, P.J. (1994), "Measuring emotion: the self-assessment manikin and the semantic differential", *Journal of Behavior Therapy and Experimental Psychiatry*, Vol. 25 No. 1, pp. 49-59.

Brandenburger, N. and Naumann, A. (2018), "Towards remote supervision and recovery of automated railway systems: the staff's changing contribution to system resilience", *2018 International Conference on Intelligent Rail Transportation (ICIRT), 12.12.2018 – 14.12.2018, IEEE, Singapore*, pp. 1-5.

Cogan, B. and Milius, B. (2021), "Development and optimization of an HMI for the remote control of automated trains in case of a system disruption: work-in-progress", *paper presented at 7th Humanist Conference, 26-27 October 2021*, Rhodes.

Cogan, B., Tandetzki, J. and Milius, B. (2022), "Passenger acceptability of teleoperation in railways", *Future Transportation*, Vol. 2 No. 4, pp. 956-969.

Di Kang, Zhao, J., Tyler Dick, C., Liu, X., Bian, Z., Kirkpatrick, S.W. and Lin, C.-Y. (2023), "Probabilistic risk analysis of unit trains versus manifest trains for transporting hazardous materials", *Accident Analysis and Prevention*, Vol. 181, p. 106950.

ERA (2022), "Report on railway safety and interoperability in the EU", Publications Office of the European Union, available at: https://data.europa.eu/doi/10.2821/28376

Gadmer, Q., Pacaux-Lemoine, M.-P. and Richard, P. (2021), "Human-automation – railway remote control: how to define shared information and functions?", *IFAC-PapersOnLine*, Vol. 54 No. 2, pp. 173-178.

Hagemeyer, F., Preuß, M., Meyer zu Hörste, M., Meirich, C. and Flamm, L. (2021), *Automatisiertes Fahren Auf Der Schiene*, Springer Fachmedien Wiesbaden, Wiesbaden.

Hart, S.G. and Staveland, L.E. (1988), "Development of NASA-TLX (task load index): results of empirical and theoretical research", in Hancock, P.A. and Meshkati, N. (Eds), *Human Mental Workload, Advances in Psychology*, Aufl., Elsevier textbooks, s.l,., Amsterdam, Vol. 52 No. 1, pp. 139-183.

Laugwitz, B., Held, T. and Schrepp, M. (2008), "Construction and evaluation of a user experience questionnaire", in Holzinger, A. (Ed.), *HCI and Usability for Education and Work: 4th*

Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008; Proceedings, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, Vol. 5298, pp. 63-76.

Linkov, V. and Vanžura, M. (2021), "Situation awareness measurement in remotely controlled cars", *Frontiers in Psychology*, Vol. 12, p. 592930.

Masson, É., Richard, P., Garcia-Guillen, S. and Adell, G.M. (Eds) (2019), "TC-Rail: railways remote driving", *In Proceedings of the 12th World Congress on Railway Research, Tokyo*, Vol. 28.

Mathis, L.-A., Diederichs, F., Widlroither, H., Ruscio, D., Napoletano, L., Zofka, M.R., Viehl, A., Fröhlich, P., Friedrich, J., Lindström, A., Thorslund, B., Litke, A., Papadakis, N., Bakalos, N., Hernandez-Jayo, U., Espié, S., Cavallo, V., Panou, M., Gaitanidou, E. and Bekiaris, E. (2020), "Creating informed public acceptance by a user-centered human-machine interface for all automated transport modes", *In Rethinking Transport. 8th Transport Research Arena TRA 2020*, Helsingfors, Finland, April 27-30, 2020, p. 9.

Melzer, T. (2020), "Konzeption neuer technologien für autonomes fahren", Der Eisenbahningenieur, February.

Naumann, A., Grippenkoven, J., Giesemann, S., Stein, J. and Dietsch, S. (2013), "Rail human factors-human-centred design for railway systems", *IFAC Proceedings Volumes*, Vol. 46 No. 15, pp. 330-332.

Naweed, A., Dorrian, J. and Rose, J. (2013), *Evaluation of Rail Technology: A Practical Human Factors Guide*, Bearman, C. (Ed.), 1st ed., CRC Press, doi: 10.1201/9781315581217.

Onnasch, L., Wickens, C.D., Li, H. and Manzey, D. (2014), "Human performance consequences of stages and levels of automation: an integrated meta-analysis", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 56 No. 3, pp. 476-488.

Russell, J.A. and Lanius, U.F. (1984), "Adaptation level and the affective appraisal of environments", *Journal of Environmental Psychology*, Vol. 4 No. 2, pp. 119-135.

Thomas, R.J., Masthoff, J. and Oren, N. (2019), "Can I influence you? Development of a scale to measure perceived persuasiveness and two studies showing the use of the scale", *Frontiers in Artificial Intelligence*, Vol. 2, p. 24.

Tonk, A., Boussif, A., Beugin, J. and Collart-Dutilleurl, S. (2021), "Towards a specified operational design domain for a safe remote driving of trains", in *Proceedings of the 31st European Safety and Reliability Conference*.

UITP (2018), World report on metro automation, Statistics Brief, available at: www.uitp.org/publications/world-report-on-metro-automation/ (accessed 1 June 2021).

Zasiadko, M. (2019), "SNCF tests its first autonomous train", RailTech.com, 7 December, available at: www.railtech.com/digitalisation/2019/07/12/sncf-tests-its-first-autonomous-train/?gdpr=deny (accessed 4 August 2022).

Zhang, Z., Liu, X. and Holt, K. (2018), "Positive train control (PTC) for railway safety in the United States: policy developments and critical issues", *Utilities Policy*, Vol. 51, pp. 33-40.

Appendix 1. HMI evaluation questionnaire

(1) Thinking about yourself and how you felt while using the system: Observe the following figures and mark the one which represents how you felt.

[For Self-Assessment Manikin figures, see Bradley and Lang (1994)].

(2) User experience questionnaire items.

(3) Consider now the overall interaction and usage of the automated system and how it affected your attitude and overall impression of the system/product. Please rate the following statements on a scale from 1 (strongly disagree) to 7 (strongly agree):

- After viewing this system, I will make changes in my attitude towards automated technology.
- This system is trustworthy.
- This system has the potential to influence user behaviour.
- I felt like I was actually driving the train from the cab.

(4) Situation awareness scale (7-Likert-Scale):

- All needed critical information was available and easily accessible.
- I had complete understanding of everything that was happening.
- I could look ahead and foresee what was going to happen.

(5) User preference for the design parameters (rating of −2 to +2):

- Haptic and tactile control elements.
- Symbols and acoustic warnings.
- Other sounds (locomotive, driving environment etc.).
- Main monitor.
- Timetable.
- Control command display.
- Diagnosis display.
- Operational information.

Appendix 2. SAGAT queries

(1) What is the train number? (Multiple-Choice question).
(2) What is the reason for the system disruption?
(3) At which stop have you just stopped? (Multiple-Choice question).
(4) How is the weather? (Multiple-Choice question).
(5) At which km have you taken manual control of the train? (Show it on the sketch given in the next page).
(6) The head of the train drives. . .

- Up the hill.
- Down the hill.
- Level ground.

(7) In which direction is the train heading? (Multiple-Choice question).

Corresponding author
Baris Cogan can be contacted at: baris.cogan@tu-berlin.de