Human Factors evaluation of a tablet-based train remote control

Anja Naumann[1] & Justin Adam[2]
[1]Institute of Transportation Systems, German Aerospace Center (DLR)
[2]Technische Universität Berlin,
Germany

Abstract

In the *Automated Regional Trains in Lower Saxony (ARTE)* project, the partners investigated highly automated operation (Grade of Automation 3 and 4, GoA 3 and 4) in today's rail network. The project also focused on the development of a simplified control for fallback mode using a tablet computer. This is the Remote Train Operation (RTO) which is to be used in the event of a malfunction of the Automatic Train Operation (ATO). It offers a lean technical solution for employees to control the vehicle remotely or on site via the user interface on the tablet. In this way, the consequences of a malfunction of the ATO can be reduced, tracks and routes cleared on time and passengers transported to the next station. In three user tests (model railway, test track, public rail infrastructure), the tablet remote control was evaluated from a Human Factors perspective. Results show that for enhancing usability and acceptance, improvements of the user interface and technical preconditions are necessary. They also indicate a lower SA together with a higher cognitive workload, when no direct view from the cab is possible and the video is the only source of visual information.

Introduction

Previous approaches for RTO in the fallback level of ATO mostly consider the development of elaborate workstation environments for the remote operator (RO) (e.g. Brandenburger & Naumann, 2019; Brandenburger et al., 2023; Schöne et al., 2023). This implies the fixed installation of a workstation (e.g., in a control centre) with a type of control panel and usually several screens. In contrast, the ARTE project (for an overview see Specht et al., 2022) has explicitly chosen the approach of developing and testing a minimal and mobile solution for manual remote control which can be used on site as well as from far away and requires little hardware. In the project, the partners ALSTOM (Alstom Transport Deutschland GmbH and Alstom Signal GmbH), the German Aerospace Centre (DLR) and the Technische Universität Berlin jointly decided in favour of implementing RTO on a tablet computer as the most practicable solution for a mobile application. As the remote control only comes into play in the event of a malfunction of the otherwise automated train, the duration of use is rather short (mostly 10-20 minutes). It can therefore be assumed that the personnel entrusted with tablet control in this case are potentially involved in other

In D. de Waard, L. Onnasch, M. Liebherr, A. Dettmann, F. Siebert, K. Karrer-Gauß, S. Winkler, A. Toffetti, and M. De Angelis (2025). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2025 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

activities. For example, a train attendant plus (TA+) would primarily be involved in passenger service and take over this activity only in case of need. Or a Remote Operator (RO), for example situated in a control room or next to the track, is responsible for monitoring and dispatching several trains at the same time and takes over and drives a train manually when there is no staff onboard, e.g. in automated shunting. An initial description of how the use of a remote tablet control in GoA3/4 could look like for TA+ and RO and the interaction between the two roles can be found in Adebahr et al. (2023). A detailed description of the new roles and job profiles and their development with railway operations personnel is documented in Naumann et al. (2024a).

In ARTE, three user tests for the remote control and its development as iterative steps were carried out, with a total of 36 participants. A first test was conducted in a model railway setup and a second test with a real train (ALSTOM Coradia LINT41) on the Alstom test track in Salzgitter (Germany). A third test was carried out with railway staff on the public line between Northeim and Bodenfelde. While single development steps have been reported before (Naumann et al. 2024b; Adam et al. 2025; Adam & Naumann 2025), this paper focuses on psychological measurements obtained from all tests, which give important insight into cognitive strain and workload and allow a conclusion on the usability and acceptance of a tablet-based remote control.

Method

First user test in a model railway layout

A first user test was carried out in highly realistic a model railway layout built by the model railway association Priorter Modulgesellschaft e.V. (PMG). Our aim was to check whether remote control using a tablet computer and the Z21 model railway app (Modelleisenbahn GmbH) is practicable. The user interface with implemented limited function volume is shown in Figure 1. The app and the model railway layout are described in detail in Naumann et al. 2023.

Thirteen participants (all male) took part in the user test. These were ten train drivers, one rail traffic controller and one trainer for rail traffic controllers. Their professional experience totalled 2-40 years (M=8.2 years). A further person had experience in railway, but did not work in any of the aforementioned professions. The distribution across the age groups was as follows: 21-29 years (n=5), 30-39 years (n=3), 40-49 years (n=3), 50-59 years (n=2).

The participants used the Z21 app on the tablet to control a locomotive equipped with a camera from a room separate from the model railway layout, the front video image of which was transmitted to the tablet. After an introduction to the tablet and the app, participants remotely drove a model train according to a given timetable. After that, participants completed the SUS (System Usability Scale; Brooke, 1996) and a self-designed questionnaire regarding the experience with the tablet-based control with the app (e.g., manageability, user experience, suggestions for improvement). The test took in total between 1 and 2.5h depending on route knowledge and Wi-Fi quality.

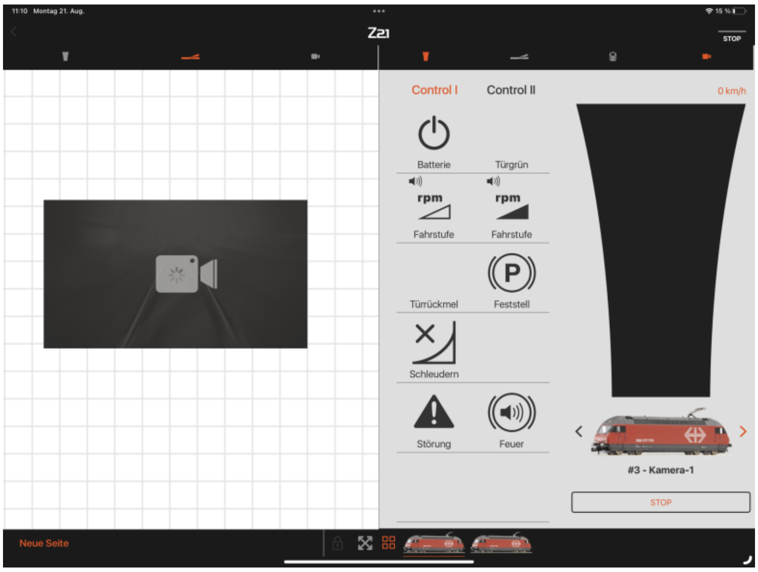

Figure 1. User interface of the tablet control with the adapted Z21 App

Second user test on the Alstom test track (Salzgitter)

A second user test for RTO with a tablet computer and a customised user interface of the Z21 app was carried out with a real train (ALSTOM Coradia LINT41).

User Interface

The tablet interface (Figure 2) had been revised after the first user test. It consisted of two separate windows. The upper one showed a video live stream from two cameras facing in the driving direction – wide angle view left, tele view right. Below the video, current speed and total distance were displayed. System time and a running time stamp for each of the two videos were displayed – in case of delayed video transmission the affected video time stamp turned red. The lower window consists of an adapted control panel of the Z21 app. The left lever controls acceleration. The right lever controls brake power (represented brake pipe pressure as usual). Control buttons between the levers activate door release, while indicators to the right inform of certain trains states relevant for driving and stopping. A detailed description of the technical functionality can be found in Arslan et al. (2025).

Participants

Thirteen participants took part in the second test on the test track of Alstom Transport Deutschland GmbH in Salzgitter. They were 30-59 years old and mostly male (n=11). While only three of them were active train drivers at the time, all the others had basic knowledge of railway operations and were professionally familiar with development and commissioning of new technology in the railway sector. On a seven-point scale, the participants self rated their expertise in rail operations slightly above the middle of the scale which was described as "an average rail worker” (M=4.3) while rating their ability to actually drive a train slightly lower (M=3.6). Only just under half of the participants were familiar with remote control of any kind. The participants rated themselves as having an above-average ability to multitask (M=4.9), and expected to fulfil the tasks well on average (M=4.1). The sample showed a slightly above-average affinity for technology (M=3.68; SD=0.37) (scale of 1-5; measured by the questionnaire *Technology-related attitudes and self-perceptions (TEiSel)*; Hahnel and Stemmann, 2023).

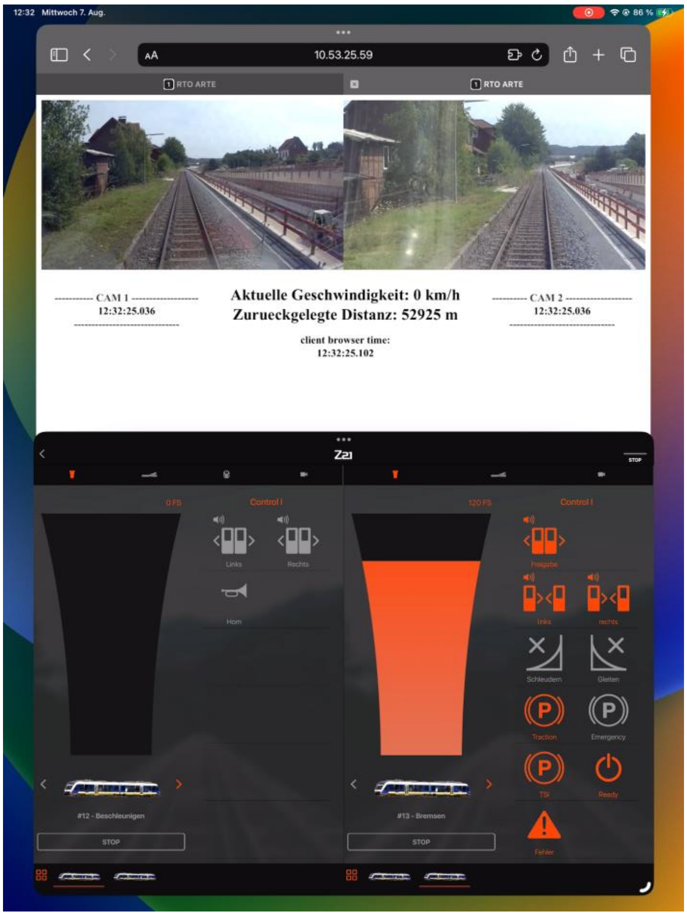

Figure 2. Revised user interface for the second and third user test

Procedure

Four working positions for RTO with a tablet (see Adebahr et al., 2023) were selected for the tests (see Figure 3). The participants controlled the vehicle from the driver's cab (forward view of the tracks), from the passenger compartment (no view of the tracks), from outside the vehicle at the track (direct visual contact to the vehicle) and from a remote control centre (approx. 1 km linear distance, with no direct visual contact to the vehicle).

The usable track length of 1,200 metres was divided into task-relevant sections by coloured flags of c. 40x80 cm size. The first task in each run was to familiarise oneself with the tablet-based controls of the vehicle over a distance of 300 m and to practise accelerating and braking. The participants were then asked to maintain a constant speed of 10 km/h over the next 400 m and of 25 km/h over a further 400 m. At the last flag, they had to stop as precisely as possible. Only when controlling from outside the train on the track, the task was different: the participants were instructed to move the vehicle for approximately 80 m and stop precisely at a step with the first door of the vehicle. In all conditions, driving performance was measured, the results have been reported earlier (Adam et al., 2025).

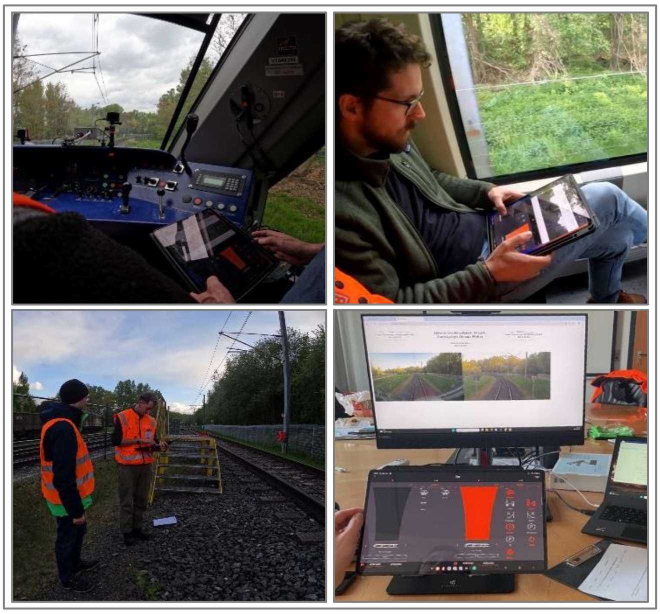

Figure 3. Tablet remote control from: the driver's cab (top left), the passenger compartment (top right), close to the track (bottom left), and the control centre (bottom right)

Demographic data and level of railway knowledge were assessed in a preliminary survey. During the journeys, the participants were observed using a structured protocol and their driving performance was documented using predefined parameters (e.g. deviation from target speed, braking accuracy).

As it was necessary to guide participants for safety reasons, they all completed the tasks in the same order. After completing the tests in and on the vehicle, for which both the video stream and the control system were displayed on the tablet, an initial survey was conducted on participants´ satisfaction with their fulfilment of the task and the technical features of the control system. After completion of the test tasks from the control centre, where an additional external monitor was used to show the video stream and thus the controls on the tablet were displayed in a larger fashion (cf. Figure 3), the survey was conducted a second time.

Participants' *Technology-related attitudes and self-perceptions* were measured (TEiSel; Hahnel & Stemmann, 2023). Four scales were included: positive affect towards technology, expectation of self-efficacy in dealing with technology, experience of competence with technology and attitude towards technology, with a scale from 1 to 5 each.

To measure the perceived usability of the Z21 app, the standard questionnaire SUS (System Usability Scale; Brooke, 1996) was used. It consists of ten statements about usability. Agreement or disagreement with all statements is rated on a Likert scale with five options. The response values are used to calculate the SUS score, which can be interpreted as a percentage value, with 100 % corresponding to a perfect system. In the user test, the SUS was administered twice: once for the tests in and on the train and once for the control centre.

The Van der Laan scale (VDL; Van der Laan et al., 1997) was used to assess the users´ acceptance of the system. The VDL scale consists of nine items, each of which is also rated on a Likert scale with five options. This is used to calculate values for the two subscales of satisfaction and usefulness. Like the SUS, the VDL scale was completed twice in the user test.

With the DLR-WAT (Brandenburger et al., 2023b), the experienced workload on various dimensions was recorded. The tool comprises a total of eight subscales: information acquisition, knowledge retrieval, decision-making, motor and physical demand, temporal demand, effort, frustration, performance. The perceived optimum is 100 on all dimensions. The DLR-WAT was given after each working position.

Situation awareness (SA) was measured with the Mission Awareness Rating Scale (MARS; Matthews & Beal 2002), a subjective rating scale for assessing participants´ SA content quality and workload in gaining SA. The questionnaire is made up of four scales for each, content and workload. SA content was rated on a four-point scale by expressing the extent of agreement to statements representing the stages of SA, the lower the rating, the higher the subjectice SA quality. SA workload was rated on a four-point scale ranging from very easy to very difficult, again distinguishing between the different stages of SA, with low ratings representing low subjective workload. Participants gave MARS ratings directly after each trial (see Figure 7).

Participants were also asked to rate their rail-related knowledge and skills (e.g., rail-specific knowledge, multitasking, job experience, experience with remote control) before participating in the test, a self-rating of task fulfilment and positive and negative aspects of task execution. They were also asked for certain qualities of the remote control afterwards (size of device, size of buttons, screens etc., missing information, video quality). In addition, personal aspects like motion sickness were asked.

Third user test on public infrastructure (Northeim-Bodenfelde)

Following successful testing at a protected test site, further test runs were carried out on a commercial track.

Participants

Ten participants drove the train on a longer distance (each participant 7 km twice), focusing on exact speed and stopping. The participants were all train drivers or train attendants, with high deviation in years of job experience ranging from two to 37 years and aged 24 to 58 years (M=35). As for the test before, the sample showed a slightly above-average affinity for technology (M=3.16; SD=1.55), a bit less than the sample in the second user test (scale of 1-5; measured by TEiSel; Hahnel & Stemmann, 2023). Further ratings were collected before the test on a seven-point Likert scale. Self-rated knowledge of the track and the specific train was rather low (half of them none, despite info given in test pre-information), all rated above average (= standard rail worker) knowledge of railway operation (M=5.3, SD=1.5). Half of them had pre-experience with any kind of remote control. Self-rated multi-tasking ability was above average, which is in line with job requirements for drivers and attendants (M=5.5, SD=1.0).

Procedure

Before the test, participants received a one-hour online briefing informing them about operational rules, the railway infrastructure, remote control functions, test procedure and tasks, and safety onboard the train. As the remote control is conceived to be used in fallback mode, speed was limited to 40 km/h, with this restriction being mandatory by operational rules. Participants entered the train at Northeim station, with two of them participating in the same test run. For the first task they drove the train via tablet from the driver’s cab, starting at standstill, adhering to signals on the line and stopping the train at the platform of the next station. On the way back, participants drove via tablet from the passenger compartment, completing the task as before. In between, the same questionnaires as in the second test were completed, and afterwards, an interview was conducted. In addition, participants were asked to give their opinion on the future job profiles of *remote train controller* and *train attendant plus* with regard to attractiveness to current and future rail employees.

A comparison of the conditions of the two user tests with the real train is presented in Table 1.

Table 1. Overview of test conditions

	2^{nd} user test: Alstom test track Salzgitter	**3^{nd} user test: test on public infrastructure (Northeim-Bodenfelde)**
Participants	2x train driver 11x from rail sector	8x train driver 2x train attendant
Briefing/ Pre-Questionnaires	On-site	In advance and on-site
Training	300 m drive to practise accelerating and braking	Detailed pre information on remote control and operational test procedure
Task	400m constant speed of 10 km/h 400m constant speed of 25 km/h Precise stopping at defined point Read coloured flags along track	2x 7km with speed limit of 40km/h Read Signals along track Read timetable while driving
Test runs	Driver cabin (on train) Passenger compartment (on train) Control Centre (Remote Operation)	Driver cabin (on train) Passenger compartment (on train)
Separate Test and Task	At the track (at train) **Different task:** move the vehicle for 80 m and stop precisely at defined point	

Results

First user test in a model railway layout

The key findings of the first user test were that route knowledge is also required for remote control and that the video image must be of sufficient size and quality. Also, it must be possible to see when the connection (Wi-Fi) is interrupted and the video image is a still image. Most participants saw the field of application of tablet remote control mainly in shunting and on secondary lines. For the Z21 app with the limited implemented function volume, a SUS-score of 70.8 (optimum = 100) was reached. The majority of test participants found the Z21 app easy to use, not unnecessarily complex and quick to learn. The majority of test participants found the functions in the app to be quite well integrated and indicated a medium level of perceived safety in operation. Around half of the participants felt that they needed further support (e.g. by a manual) to be able to use the app. Further optimization requirements for using this tablet solution for the remote control of a real train were identified (for details see Naumann et al., 2023). For example, more specific functions, such as door control and a separate brake lever, and additional information like train standstill should be integrated.

Second and third user test: test track and public infrastructure

Although test conditions vary and statistical comparisons are not possible for the tests on test track and public infrastructure, the results are placed side by side in the following figures. Firstly this allows for getting a first impression how measures evolve from driving a short distance on a closed and protected test track to a longer distance on a real track. Secondly both tests together represent the variety of possible tasks future TA+ and RO might have to solve via a tablet-based remote control, which is why results shall be considered together. The perceived usability (SUS questionnaire) for the Z21 app on the tablet in the second user test on the test track (see Figure 4) amounted to a mean value of 61.9 (SD = 14.4) for the 'on/at train short' condition and a mean value of 60.4 (SD = 15.7) for the 'Remote operation' condition, both in the range of medium to good usability (optimum = 100). The condition 'on/at train short' is rated slightly more positively, but the conditions do not differ significantly.

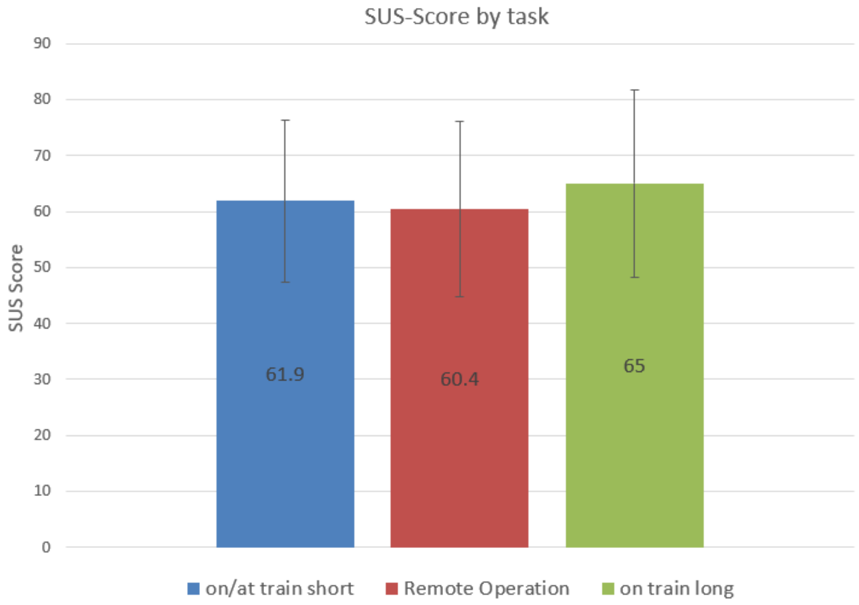

Figure 4. Usability (SUS; 100 = optimum) for the three conditions: on/at train short (on test track), remote operation (in control centre), and on train long (on public infrastructure)

For the third user test on public infrastructure ('on train long'), the usability (M = 65; SD = 16.7) is slightly higher than in the test on the test track and in the control centre. This might be due to a more detailed participant pre-test information, a longer test and therefore usage duration and less actions and incidents. In both field tests, the usability is slightly below that of the model railway test (M = 69.4; SD = 10.2). This could be explained by the more complex control system (two controllers: one for acceleration and one for deceleration), the more complex camera view (two viewing angles instead of one) and the generally greater responsibility of controlling a real vehicle.

For the acceptance (VDL scale) of the tablet control system with the Z21 app in the second user test on the test track (see Figure 5), the mean value for the satisfaction with the system scale was 0.23 (SD = 0.94) for the ‘on/at train short’ condition and -0.13 (SD = 0.87) for the ‘Remote Operation’ condition. For the usefulness scale, the mean value for the ‘on/at train short’ condition was 0.58 (SD = 0.62) and for the ‘Remote Operation’ condition 0.35 (SD = 0.87). Overall, usefulness was rated slightly higher than satisfaction, but the values do not differ significantly when comparing the two dimensions or any of the conditions. In the third user test on public infrastructure, positive values were obtained for satisfaction (M = 0.19; SD = 0.73) and usefulness (M = 0.42; SD = 0.62), which lie between those for in the ‘on/at train short’ and ‘Remote Operation’. Remote Operation in the control centre generally got the lowest rating, mainly caused by long latencies and problems with the video transmission.

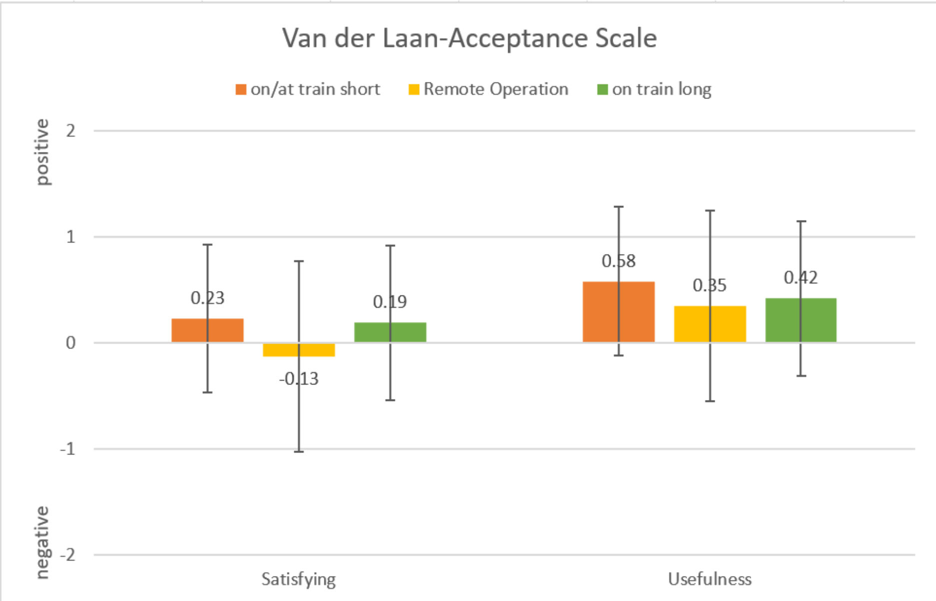

Figure 5. Acceptance (VDL-Scale) for the three conditions: on/at train short (on test track), remote operation (in control centre), and on train long (on public infrastructure)

For cognitive workload (DLR-WAT, see Figure 6), the three conditions on/at the train in the second user test on the test track show the values closest to the optimum of 100 (driver cabin/short: M= 105.10, SD=16.86; passenger compartment/short: M=103.18, SD=14.62; at train/short: M=103.7, SD=17.38). In comparison to that, a slight cognitive underload is shown for the third user test on the public infrastructure (driver cabin/ long: M=84.49, SD=18.65; passenger compartment/long: M=83.16, SD=23.52). This might be explained by the long driving task with a lower number of events and less calls to action compared to the second user test on the test track with a short driving time and high density of events (signal flags that had to be observed and reacted to). The condition 'Remote Operation' got the highest workload ratings, a bit above the optimum (M=110.34, SD=17.66). In this condition, the distance from the operator to the real train was the biggest. In addition, long latencies and disruptions in video transmission made the task demanding, also seen before in the usability/acceptance ratings.

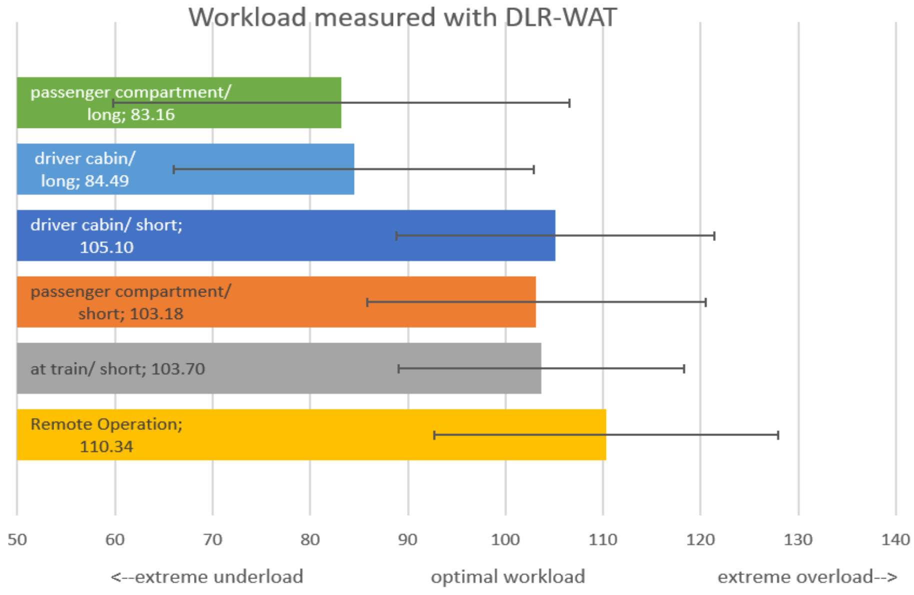

Figure 6. Cognitive workload measured with DLR-WAT (optimum = 100)

The results from the MARS questionaires are shown for the second and third test in Figure 7 and 8. The lower the rating, the higher the subjective quality of the SA content or the easier it was to either identify, understand or predict relevant information and decide on the next action. For the second user test, we assume that, the majority of participants not being train drivers, a learning effect between the first run on the cab and the second run from the passenger area to have affected the results.

Overall results indicate a lower situational perception together with a higher workload on perceiving and understanding information, when no direct view from the cab is possible and the video is the only source of visual information. This is even more visible with the test run on the public line, where most participants had professional experience in train driving. Also, additional feedback on the video quality suggests that difficulties in reading signals on the video and thereby uncertainities in decision-making resulted in higher workload for perceiving, comprehending and predicting. Also, participants reported mental strain from continuous and focussed visual attention to the tablet.

Participants found it especially difficult to gain SA and were rather unsatisfied with the respective quality when they controlled the train from remote. While still feeling the train movement and hearing the train (motor, movement etc.) in the passenger area, visual information from the video stream and status indicators were now the drivers´ only source of information and feedback to their actions.

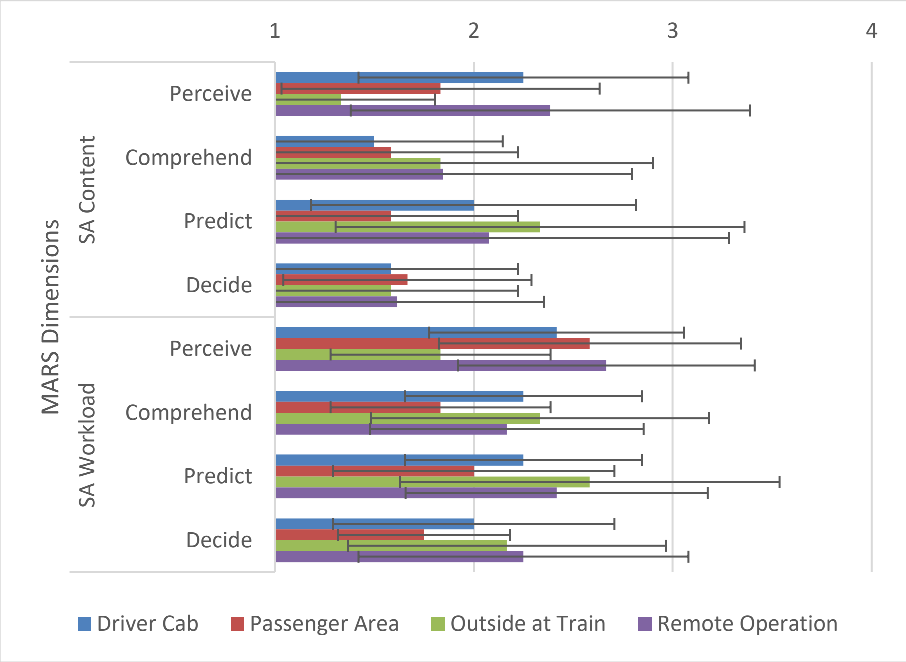

Figure 7. MARS subjective SA ratings from the Salzgitter test site remote driving test, SA Content: 1= fully aware 4= not aware at all, SA Workload: 1= very easy 4= very difficult

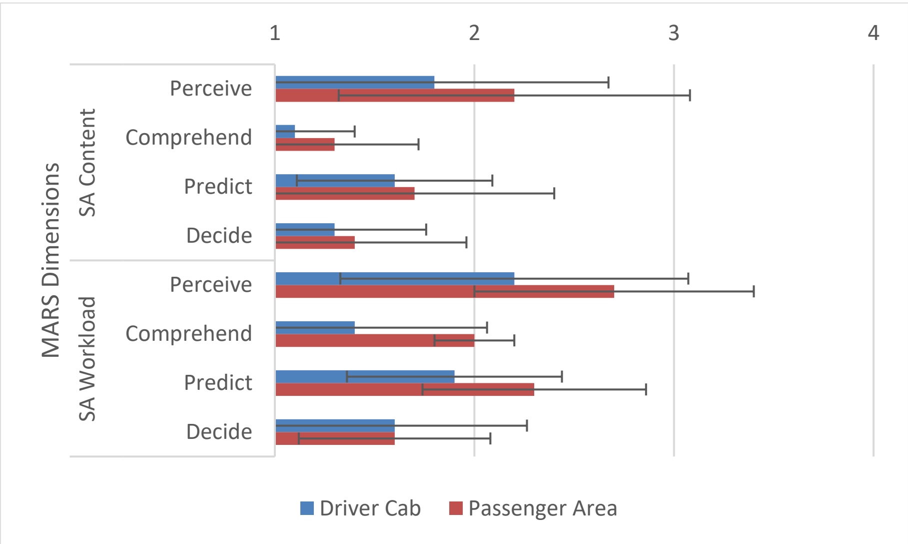

Figure 8. MARS subjective SA ratings from the public line remote driving test, SA Content: 1= fully aware 4= not aware at all, SA Workload: 1= very easy 4= very difficult

Discussion

The findings of all three user tests in ARTE indicate that using a tablet computer for remotely controlling a train (RTO) is a practible solution, also from a Human Factors perspective, especially when personnel is still on board the train. However, for enhancing usability and acceptance, improvements of the user interface and technical preconditions are necessary. This concerns above all enhancement of video quality, a smooth control, and direct feedback. Sufficient mobile data or Wi-Fi coverage with low latencies is a mandatory requirement for low latency of feedback. The results also indicate a lower SA together with a higher cognitive workload, when no direct view from the cab is possible and the video is the only source of visual information. Driving remotely from a control centre was rated as the most difficult condition, caused by long latencies as well as the furthest distance to the train itself, with the RO not being able to see and feel the train. Further research is needed on identifiying suitable additional feedback for using RTO in the control room.

In general, such minimal workplaces as investigated here, are getting more and more important as a fallback solution for ATO. In ARTE, also implications for new tasks, roles and job descriptions are investigated (Naumann et al., 2024 and 2025). Further research in this direction in follow-up projects is recommended.

Funding information

The Federal Ministry of Economics and Climate Protection (BMWK) is funding the work presented here as part of the *New Vehicle and System Technologies* programme. This content was created as part of the *ARTE - Automated Regional Trains in Lower Saxony* project.

References

Adam, J., Arslan, B., & Naumann, A. (2025). Entwicklung und Realerprobung einer Steuerung über Tablet für Remote Train Control. In: *7th Scientific Railway Signalling Symposium*, 2024-11-28, Berlin, pp. 7-22.

Adam, J. & Naumann, A. (2025). Performance and usability testing of a tablet based train remote control. *6th German Conference on Rail Human Factors, 2025-02-18 - 2025-02-19*, Berlin, Germany. Available from: https://elib.dlr.de/213190/

Adebahr, F.-A., Milius, B., & Naumann, A. (2023). Flexible Arbeitsumgebungen für die ATO-Rückfallebene. *EI - Der Eisenbahningenieur, 01/23*, pp. 39-41. Tetzlaff Verlag.

Arslan, B., Ramsdorf, T., & Herholz, H. (2025). ARTE – Eine Zugfahrt mit Z21. Mit einer Modellbahn-App ein echtes Triebfahrzeug steuern. *EI – Der Eisenbahningenieur, 01/25*, Tetzlaff Verlag.

Brandenburger, N., Bier, I., Busse, M., Gärtner, J., & Melzer, T. (2023). A technical demonstration of remote train operations using 5G mobile communications. *SIGNAL + DRAHT, 06*, pp. 34-40. DVV Media Group.

Brandenburger, N., Dreßler, A., & Grippenkoven, J. (2023). *DLR Workload Assessment Tool (DLR-WAT) - Official English Version*. Available from: https://elib.dlr.de/201450/

Brandenburger, N. & Naumann, A. (2019). Der Arbeitsplatz des Train Operator im Fokus. *EI – Der Eisenbahningenieur, 11/19,* pp. 13-17. Tetzlaff Verlag.

Brooke, J. (1996). SUS: A ‘quick and dirty’ usability scale. In P. Jordan, B. Thomas, B. Weerdmeester & I.L. McClelland (Eds.) *Usability Evaluation in Industry*. S. 189-194, Taylor & Francis.

Hahnel, C. & Stemmann, J. (2023). Entwicklung eines Fragebogens zur ökonomischen Erfassung technikbezogener Einstellungen und Selbstwahrnehmungen. *Diagnostica*.

Matthews, M.D. & Scott, A.B. (2002). A field test of two methods for assessing infantry situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Naumann, A., Adam, J., & Oehme, A. (2025). *ARTE - Neue Rollen und Berufsbilder im automatisierten Fahren*. *EI – Der Eisenbahningenieur, 06/25*. Tetzlaff Verlag.

Naumann, A., Adam, J., Fritsch, M., & Hofstädter, R. (2024a). ARTE: Neue Aufgaben und Rollen für Betriebspersonal. *EI – Der Eisenbahningenieur, 08/24*, pp. 42-45.

Naumann, A., Arslan, B., Herholz, H., & Schöne, S. (2024b). Evaluation of train remote control via tablet in a railway model layout. *HFES Europe Annual Meeting, 2024*, Lübeck, Germany.

Naumann, A., Arslan, B., Herholz, H., & Schöne, S. (2023). Evaluation einer Zugfernsteuerung mit Tablet im Modellbahnaufbau. *EI - Der Eisenbahningenieur, 12/24*, pp. 10-15. Tetzlaff Verlag.

Schöne, S., Mönsters, M., Käthner, D., & Brandenburger, N. (2023). ATO-Cargo: Operating procedures for the fallback levels of highly automated railway operation. *SIGNAL + DRAHT, 115 (10)*, pp. 18-25. DVV Media Group.

Specht, F., Michels, A., Adebahr, F.-A., Meirich, C., Hofstädter, R., Milius, B., & Naumann, A. (2022). Automatisiertes Fahren in Niedersachsen - ARTE. Automated driving in Lower Saxony – ARTE. *SIGNAL + DRAHT (114) 09/22*, pp. 10–16.

Van der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research - Part C: Emerging Technologies, 5, 1-10*.