Article

Teleoperation of Highly Automated Vehicles in Public Transport: User-Centered Design of a Human-Machine Interface for Remote-Operation and Its Expert Usability Evaluation

Carmen Kettwich *, Andreas Schrank and **Michael Oehl**

German Aerospace Center (DLR), 38108 Braunschweig, Germany; andreas.schrank@dlr.de (A.S.);
Michael.Oehl@dlr.de (M.O.)
* Correspondence: Carmen.Kettwich@dlr.de

Abstract: Paving the way to future mobility, teleoperation of vehicles promises a reachable solution to effectively use the benefits of automated driving as long as fully automated vehicles (SAE 5) are not entirely feasible. Safety and reliability are assured by a human operator who remotely observes the vehicle and takes over control in cases of disturbances that exceed the vehicle automation's skills. In order to integrate the vehicle's automation and human remote-operation, we developed a novel user-centered human-machine interface (HMI) for teleoperation. It is tailored to the remote-operation of a highly automated shuttle (SAE 4) by a public transport control center and based on a systematic analysis of scenarios, of which detailed requirements were derived. Subsequently, a paper-pencil prototype was generated and refined until a click-dummy emerged. This click-dummy was evaluated by twelve control center professionals. The experts were presented the prototype in regular mode and were then asked to solve three scenarios with disturbances in the system. Using structured interview and questionnaire methodology, the prototype was evaluated regarding its usability, situation awareness, acceptance, and perceived workload. Results support our HMI design for teleoperation of a highly automated shuttle, especially regarding usability, acceptance, and workload. Participant ratings and comments indicated particularly high satisfaction with the interaction design to resolve disturbances and the presentation of camera images. Participants' feedbacks provide valuable information for a refined HMI design as well as for further research.

Keywords: human-machine interaction; interface for remote-control; teleoperation; automated driving; usability evaluation; expert evaluation; situational awareness; workload; user acceptance

1. Introduction

The automation of driving tasks is proceeding rapidly [1]. However, until full automation of driving tasks will be achieved, there is still a long way to go [2]. Urban mixed-traffic environments with complex driving maneuvers and the spontaneous encounter of a plethora of traffic participants pose requirements on the automation that are yet to be fulfilled [3]. A possible interim solution to use potential of today's advanced automation without compromising the passengers' safety is outsourcing the monitoring task from the driver's on-board cabin to a higher-level actor, such as a control center that remote-controls and commands the vehicle if necessary [4,5]. Teleoperation is seen as a prerequisite for the introduction of SAE level 4 [6] highly automated driving features on public roads as remote intervention by a human can take over critical situations that the highly automated vehicle cannot handle itself [7]. To achieve efficient and safe teleoperation, not only technological developments such as the data connection need to be considered but also the interaction between operator and the remote-control workstation needs to be designed [8,9].

1.1. Automated Driving and Public Transport

Automated driving is a disruptive technology with the potential to reduce driver stress, energy consumption and pollution while increasing road capacity, granting mobility to those without a driver's license, and, most importantly, raising safety [10,11]. However, fully automated driving (SAE level 5) that can cope with any conceivable scenario is not feasible in the next years [12]. Highly automated vehicles (SAE level 4) need an operator to intervene in certain situations that exceed the system's capabilities [13]. This operator can either be on board of the vehicle [14–16] or somewhere else and remote-control it (see Section 1.2). In the latter case, remote-operation could serve as an interim technology to exhaust the present and near-future potential of automated systems. This paper will therefore focus on SAE level 4 automation.

An area in which automated vehicles (AVs) could have a particularly large impact is the public transport sector. Combining AVs with on-demand mobility may further reduce energy consumption and help to save greenhouse gas emissions [17]. Models suggest that shared AV-based mobility may only require a tenth of the number of cars necessary in case of personal vehicle ownership [18]. Shared AVs bear the potential to extend the existing schedule-based means of transportation by flexible on-demand mobility options [19].

1.2. Teleoperation

Teleoperation of vehicles represents an approach to be able to already use the advantages of automated driving as an interim stage without having to give up the advantages of human information processing. Instead of an attendant on board who monitors the automation on site and intervenes if necessary, the operator in teleoperated driving is located elsewhere and is connected to the vehicle via a communication link. This enables location-flexible control of multiple vehicles at the same time. Based on these characteristics, highly automated vehicles could be used as a complementary on-demand service in public transport and monitored and controlled from a public transport control center. Two main designs for vehicle teleoperation interfaces are conceivable (inspired by Fong and Thorpe's [20] classification): While the *direct approach* mimics the manual driving process and therefore relies on the remote-operator's continuous attention, the *indirect approach*, in turn, only requires an occasional human intervention while using computational power to translate human decisions into driving actions. This does not only make remote-control more efficient but also more adaptive to lagged connections as the remote-operator's input consists of high-level goals that algorithms use to calculate control signals [9]. Kay [21] suggests waypoints to be used as high-level goals. On a map, the remote-operator selects positions that the computer will use to calculate a trajectory which the vehicle will follow if no obstacles are detected on the way. Kim and Ryu [22] demonstrated the capability of algorithms in dealing with problems of time delay in teleoperation. Alternatively, the entire trajectory may be created directly by the operator, such as in the "free corridor" approach proposed by Chen [23]. It targets the issue of connection losses. In this approach, a colored area is laid over the video images which represents a corridor for safe braking even when the connection may be suddenly interrupted. Gnatzig et al. [9] showed that trajectory-based driving is at least fast enough for inner-city traffic. Chucholowski et al. [5] adapted the approach of using a predictive display to anticipate the position of the teleoperated vehicle as well as other traffic participants in mixed traffic environments. By interpolating the trajectory of a traffic participant, the HMI proved capable of mitigating the latency of the connection and presenting the operator an as-if-present view of the traffic situation.

1.3. Psychological Background

Teleoperating vehicles requires the remote-operator to engage in multiple tasks. Sheridan [24] lists five generic functions that supervisory control entails: planning what task to do and how, teaching the computer, monitoring automatic actions, intervening, and learning from experience. This paper focuses on two of them: monitoring the situation, which

requires vigilance, and intervening, that is, taking over the vehicle when the automation is no longer able to do so.

1.3.1. Monitoring

Identifying when to take over control of a vehicle requires continuously monitoring the system, a mentally very taxing task that requires constant vigilance. In the context of AVs, a driving simulator study by Greenlee et al. [24] found slower reaction times and tremendously decreasing hazard recognition during a 40-minutes long virtual ride in a highly automated vehicle. Thus, an underload of cognitive demand can lead to impaired vigilance but the same is true for an overload [25]. Thus, a human-automation interface for teleoperation should impose medium cognitive requirements onto its user. Ideally, the HMI does not depend on the remote-operator's vigilance at all but directs attention quickly and effortlessly to relevant stimuli. And even if the operator manages to maintain attention, its allocation to stimuli is still an issue. Wickens et al.'s SEEV Model predicts the distribution of attention when information is presented on multiple screens [26]. Wickens postulates four factors that influence attention allocation: salience, effort, expectancy, and value. While the first two factors are bottom-up stimuli emanating from the environment, the last two factors stem from existing prior knowledge and are therefore considered top-down factors. In "ideal scanning" [27] (p. 749), attention is only distributed based on the user's assessment of its value and their expectations based on experience, i.e. how often information has changed in the past [28]. "Actual scanning" considers the impact of factors of the environment, such as the effort to move the head to look at another screen [29], for instance, or the physical properties of a stimulus that attract attention, such as contrast and size [30]. For creating an interface, these findings imply that relevant information be presented in a way that not only conforms with user-related factors such as the user's expectations to find the information on a particular screen, for example. Relevant information also needs to be presented in a salient, that is, highlighted, way to reduce the effort of directing one's attention there.

1.3.2. Intervening

When an incident occurs that requires the remote-operator's attention, a smooth and quick *takeover* from the automation to the remote-operator is crucial. An essential factor that determines the takeover is the point when the system's capability is depleted and human action is required. A central prerequisite for takeover is the presentation of stimuli from the environment. They determine the perception of a situation and therefore influence monitoring and controlling the AV. Sensations are integrated into a holistic impression, which is then used to analyze the situation and obtain *situation awareness*, i.e. "the knowledge, cognition and anticipation of events factors and variables affecting the safe, expedient and effective conduct of the mission" [31]. According to this conceptualization, situation awareness is the difference between attentional demand and attentional supply. That is, when the demand imposed on an individual's attention exceeds their resources, situation awareness can no longer be assumed. Merat et al. transferred this concept to the realm of automation-operator transition framed as *"Out-of-the-Loop" Model* [32]. The authors distinguish between the physical control of a vehicle and the monitoring of the situation. They argue that to be "out of the loop," the vehicle driver (1) neither has physical control of the vehicle nor monitors the driving situation or (2) has physical control of the vehicle but does not monitor the driving situation. Therefore, monitoring the situation, rather than direct physical control, is critical for an operator to be considered "out of the loop". Monitoring the situation is therefore a necessary, but not sufficient, for situation awareness. A concept related to this is *telepresence*. It is a prerequisite for the teleoperation of vehicles [33,34] and was defined by Minsky [35] as an amendment to teleoperation. For Sheridan, it is central to the concept that "the operator feels physically present at the remote site" [24] (p. 6).

1.4. Study Objectives

Although software and hardware solutions for teleoperation of vehicles exist, to the authors' knowledge, no systematic research has been conducted to develop an HMI for teleoperation of highly automated vehicles that follows a human-centered design and usability evaluation process. In particular, there is a gap in research on HMIs for teleoperation of vehicles in the context of public transport that are tailored to the needs, expectations, and operation styles of control centers in this domain. Therefore, the goal of this work is the creation of a human-machine interface for the teleoperation of highly automated vehicles in a user-centered design process and its evaluation regarding usability and other essential concepts mentioned in Section 1.3. by experts in the field of controlling public transport. Usability is crucial in the evaluation process because it determines how well the user is able to "interact with the object of interest", following ISO's conceptualization of usability, according to which it is required to be effective, efficient, and satisfying [36]. User acceptance refers to the evaluation of a system's ergonomics. It is imperative for the success of newly introduced technology that the user embraces it [37]. Resulting from these considerations, the central research question is whether the HMI is suitable to meet the following seven criteria that will be considered during the evaluation study:

1. *Features:* The remote-operation workstation must provide necessary features to monitor the automation, provide disturbance information and support remote-operator with resolving the disturbance.
2. *Information:* The remote-operation workstation must provide necessary information to monitor the automation, provide disturbance information and support them with resolving the disturbance.
3. *Situation Awareness*: The remote-operation workstation must provide a high level of situation awareness to the remote-operator.
4. *Usability*: The remote-operation workstation must have good usability.
5. *User Acceptance*: The remote-operation workstation must have a high user acceptance.
6. *Attention*: The remote-operation workstation must direct the user's attention to information that is currently relevant.
7. *Capacity:* The remote-operation workstation must not overwhelm the user's mental and physical capacities.

2. Materials and Methods

2.1. Prototype

The prototype was developed to provide an HMI concept for the setup of a remote-operation workplace to conduct research studies in the field of teleoperated driving within a variety of research projects at the German Aerospace Center, such as "RealLab Hamburg" [38], "U-Shift33" [39], and "AHEAD" [40]. It is based on ISO's user-centered design process [41]. In an initial step, potential scenarios relevant in teleoperation were defined and analyzed. These were brought about using video analysis of critical scenarios regarding the interaction of highly automated shuttles with other traffic participants, observations of and interviews with control center professionals, as well as brainstorming sessions of experts in the field of automated driving and future mobility. For instance, relevant functional roles for a teleoperation control center in public transport were identified using a card-sorting method [42]. Next, user requirements were derived from the relevant scenarios. A low-fidelity prototype was generated that fulfilled the essential requirements. It was further refined until a click-dummy prototype emerged.

This click-dummy consists of seven monitors, six of them are regular wide-angle PC screens that can be operated with mouse and keyboard. The seventh monitor is a touchscreen integrated in the desk. Figure 1 provides an overview of the prototype. Table 1 presents the distribution of elements across the monitors and the menu structure.

2.1.1. "Video Screens"

The top row consists of three screens that stream video images. Regular screens were chosen for video streaming. In regular mode, the front view is displayed over all the screens, creating a wide scope to see objects in the periphery and assuring situation awareness due a wide visual angle. Other camera views can replace the right and left screens by manual selection or by the automation in case of disturbances. The continuous video stream contributes to telepresence and helps the teleoperator to stay on the loop, even while resolving several disturbances simultaneously. The central screen is always occupied by the front view images. The centrality of this screen, together with the also highly relevant "disturbances screen", assures sustained attentional distribution to the most essential screens, considering the "actual scanning" phenomenon mentioned above.
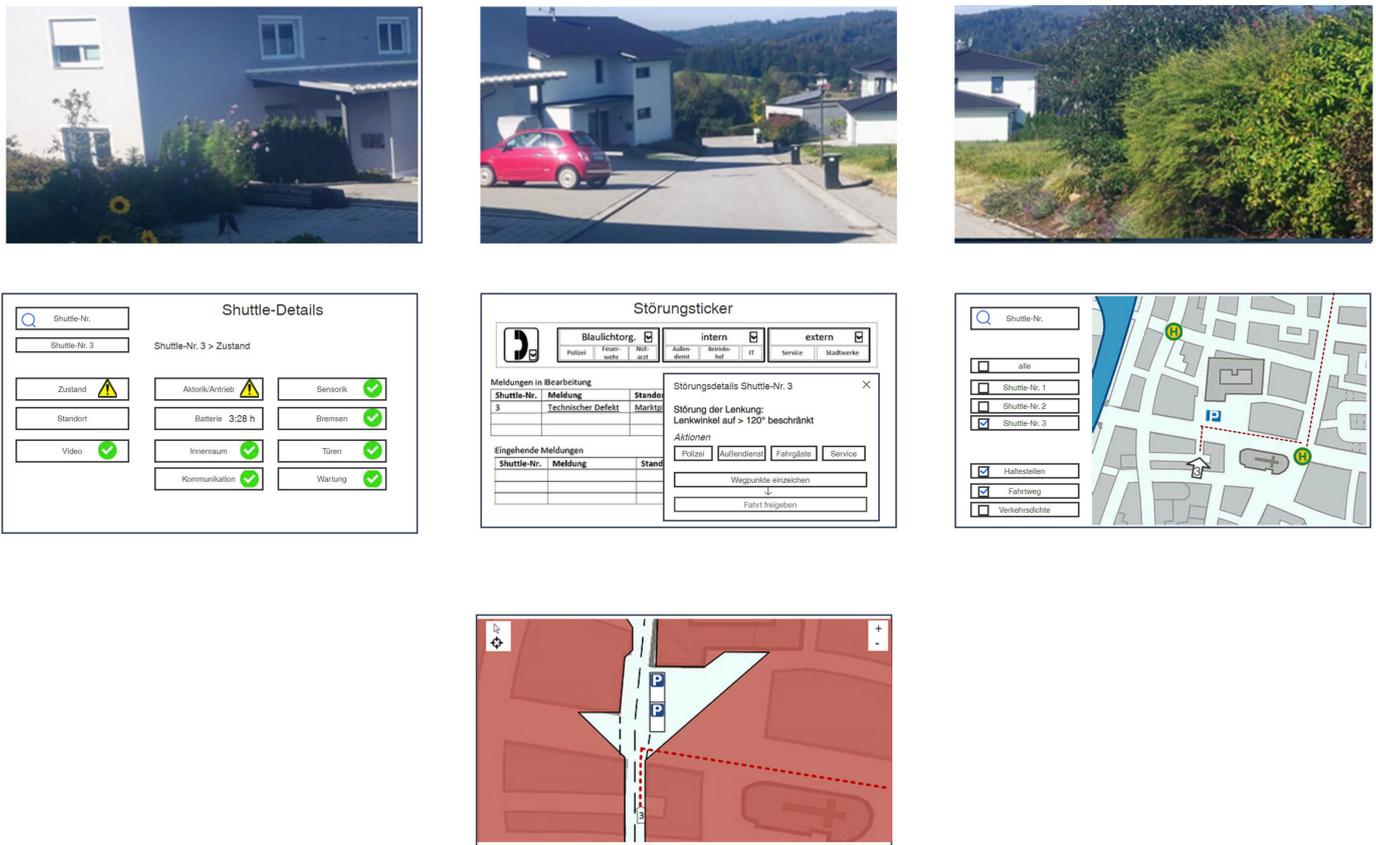

Figure 1. Overview of the prototype evaluated in this study. Each box represents a monitor, six of them are regular computer monitors that are operated with keyboard and mouse. The top row of monitors consists of three Video screens. The central row consists of (from left to right) the "details screen", the "disturbances screen", and the "map screen". The bottom screen is the "touchscreen" embedded in the workstation's desk is therefore operator with fingers or a stylus. The "details screen" contains a search bar on the top left corner and three navigation tabs reading "State", "Position", and "Video" below it. Next to them, different state-related categories are listed, such as "Actorics", "Sensorics", "Battery", and "Brakes". The "disturbances screen" provides a communication bar on top, and two tables below that make up the disturbances ticker. It consists of two sections, "Notifications in Progress" and "Incoming Notifications". The pop-up window on the right shows details on the selected disturbance and steps to resolve it. The "map screen" contains a search bar on the top left corner, checkboxes to select specific shuttles below, and layers to be added to the map, such as "Stops" and "Trajectory". The "touchscreen" shows the original trajectory as a red dotted line. The area where waypoints can be set is shown in white, while no waypoints can be set in the red area.

Table 1. Overview of menu structure and elements of the HMI prototype.

Screen	Elements Level 1	Elements Level 2
"Video screens"	Video images	-
"Details screen"	State	Actuators Sensorics Battery . . .
	Position	Street name Schedule
	Video	Available cameras
"Disturbances screen"	Communication bar	Available communication partners
	Disturbances ticker	Notifications in progress Incoming notifications
"Map screen"	Map	-
	Layers	Stops Trajectories Traffic density . . .
"Touchscreen"	Map	-

2.1.2. "Details Screen"

On the second row, the "details screen" provides an overview of the current state of a single shuttle. The number of the shuttle selected and its position are shown in a navigation menu. A shuttle can be selected either via a search bar or a dropdown menu. Below, three buttons reading "State", "Position", and "Video" are located. For the first and the last one, a colorful symbol represents the overall state of the subsystem: a green checkmark indicates regular operations, a yellow exclamation mark symbolizes a singular disturbance, and a red X sign displays a total breakdown. The color-coding was chosen to increase salience of disorders, following Wicken's SEEV model. "State" opens a list of technical systems, such as "Actuators", "Sensorics", or "Battery". "Position" shows the current location as a street name. Below, a schedule of the following five stops is presented, with the imminent stop highlighted. Scheduled and estimated departure times are provided for each stop. "Video" shows the available camera perspectives presented on the top monitor row.

2.1.3. "Disturbances Screen"

The "disturbances screen" consists of a communication bar and a table with incoming notifications about disturbances, the disturbances ticker. The communication bar enables the remote-operator to call relevant actors. The disturbances ticker consists of two sections, "Notifications in Progress" and "Incoming Notifications". The former section lists disturbances currently under review by a remote-operator, the latter one those that are not yet being reviewed. The parameters presented under "Notifications in Progress" are "Shuttle No.", "Notification", indicating the kind of disturbance, e.g., "Technical Malfunction", the parameters "Position", "Next Stop", "Editor", showing which remote-operator is currently editing, and "Action", for reviewing details on the disturbance. The parameters of "Incoming Notifications" are similar but the parameter "Editor" is not shown and pressing "Accept" is the only action available. By clicking on the action "Edit" on the first section, a pop-up window with details on the disturbance appears. A list of potential actions is shown. After successfully implementing them, a window with prerequisites pops up. Only after all the boxes are checked, the ride can be resumed by clicking on "Give Clearance".

2.1.4. "Map Screen"

The last screen on the central row shows a map of all shuttles or a single shuttle's surrounding. The shuttles' trajectories are shown by dashed lines on the map. The shuttles are represented as little arrows on the map, bearing the shuttle number and indicating the direction of travel. On the left margin, there are a navigation column with a search bar and boxes to check shuttles for display. Additional layers, like stops, trajectories and traffic density, can be selected.

2.1.5. "Touchscreen"

The final monitor is the "touchscreen". It is integrated in the table and used to set waypoints with a hairline cross. An adjustable map with the environment around the shuttle is displayed. Areas where no waypoints can be set are indicated by a red layer. The original trajectory is represented by a dashed red line. By touching a point, the remote-operator can set waypoints, which are used to calculate the shuttle's trajectory. The automation's security mechanisms cannot be overruled by the remote-operator.

2.2. Scenarios

In order to assess whether the HMI concept is suitable for scenarios occurring in the teleoperation of self-driving vehicles in public transport, three relevant scenarios were chosen as representatives for monitoring the shuttle, giving clearance, and executing remote control via waypoints. They were selected from the list of relevant scenarios in teleoperation as described in Section 2.1. based on their representativity for the list of scenarios overall. All scenarios contain an irregularity in the operation of the shuttle that cannot be solved by the automation on its own but requires the remote-operator to take action.

In Scenario A, a technical malfunction restricts the steering angle of the shuttle's actuator to max. 120 degrees. Therefore, the shuttle cannot follow the calculated trajectory. Instead, a field engineer needs to be contacted and sent to the shuttle to fix the malfunction. The passengers need to be informed about the incident and the next steps need to be communicated. In order to bring the shuttle to a safe halt and have the passengers alight, the remote-operator needs to determine waypoints for a new trajectory that parks the shuttle at the closest parking lot. The waypoints need to be located in a specified area so that the resulting trajectory does not require steering of more than 120 degrees. After the waypoints have been determined and a list of prerequisites has been checked, the remote-operator gives clearance so the shuttle can drive to the lot, following the updated trajectory, park there, and wait for inspection.

In Scenario B, an unclear detection situation requires the remote-operator to check whether an object is blocking the sensor. The system notifies the remote-operator that an obstacle has been detected by one of the vehicle's sensors. This leads the shuttle to stop and wait for clearance by the remote-operator. On the "details screen", the remote-operator can ascertain that the sensors' hardware and software are intact but that there is uncertainty whether an obstacle blocks the trajectory. After screening the video images and making sure no object is in the way, the remote-operator checks a list of prerequisites and gives clearance so the shuttle can resume its ride.

In Scenario C, the shuttle's doorway is blocked by an object so the door can't be closed. After inspecting the camera images and identifying the object as a suitcase, the remote-operator calls the shuttle's cabin to ask the passengers to remove the suitcase from the doorway. Upon removal, the remote-operator checks a list of prerequisites and gives clearance so the shuttle can resume its ride.

2.3. Participants

As an expert sample, a group of 13 male employees in public transport control centers in Germany with different lengths of experience in control center work where chosen. Their work is compromised of monitoring the operations of public transport within an urban

area, and taking action in case of disturbances, such as accidents, medical emergencies, or technical malfunctions, by deploying alternative means of transport, communicating expected delays to passengers, and ensuring a timely resolution of the disturbance to resume regular operations. To fulfil these tasks, they are in touch with a variety of actors, such as bus drivers and train conductors, blue light organizations, field engineers and technicians, dispatchers, and traffic information services.

One participant had to be excluded due to irregularities during the study and was not considered in the analyses. The remaining 12 participants were between 25 and 64 years old. Of these, 6 participants were between 45 and 54 years and each 2 were in the age groups from 25 to 34, 35 to 44, and 55 to 64 years, respectively. All participants had at least four years of experience as public transport control center professionals. They had average values above the mean of the Likert scale (1 = "not true at all" to 6 = "absolutely true"), 3.5 (M = 4.09, SD = 0.38) on the Affinity for Technology Interaction Scale (ATI [43]), indicating technological affinity slightly above average. Participation was voluntary. All participants were able to abort the study at any time and were debriefed afterward. All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki.

2.4. Study Design

This usability study followed a mixed method design. Due to limitations set by the Covid-19 pandemic, the study was conducted remotely via a conference meeting interface. The HMI prototype was set up as a click-dummy using a website builder. Data was collected to assess whether the criteria for a user-centered design of a remote-operation workplace stated above tended to be fulfilled. The data were collected using a combination of quantitative and qualitative methods, following the idea of method triangulation. The quantitative data are self-report data collected in questionnaires via Likert scales. The qualitative data resulted from a structured interview.

The methods used are suitable for the objective of optimizing the HMI prototype because, on the one hand, they encourage the evaluator to make quantifiable assessments of the design via standardized questionnaires, but on the other hand they also provide sufficient space for individual feedback and suggestions for optimization that cannot be directly quantified, such as a structured interview.

2.5. Dependent Variables

Participants were asked to fill in online well-established standardized questionnaires provided by the interviewer, in order to measure usability, situation awareness, user acceptance and perceived workload. Additionally, a structured interview was used to gain further insights.

2.5.1. Questionnaires

The questionnaires used relate directly to the criteria mentioned above. They are as follows: The NASA Task Load Index (NASA-TLX [44,45]), which measures the degree of task stress; the Post-Study System Usability Questionnaire (PSSUQ [46]), which measures the usability of a system, including its user interface; the Van der Laan Scale (VDL [37]), which quantifies the acceptance of an HMI by its user; the Situation Awareness Rating Technique (SART [47]), which indicates situational awareness; selected questions based on the SEEV model [26] that contain attention-related indicators considering the resolution of disturbances, the presentation of information, and the projection of the future.

2.5.2. Structured Interview

The main goal of the structured interview was to encourage spontaneous recall of particularly liked, disliked, missing and redundant features. It provided a both structured and open format to address shortcomings regarding the HMI's features and presented information parameters, and indicate concrete ideas for improvement. In addition to this

qualitative part, a structured interview assessed the meaningfulness of functions of the HMI and the importance of information presented quantitatively on Likert scales.

2.6. Procedure

After welcoming the participants and documenting their consent to the terms and conditions of the study, basic demographic data was collected in an online questionnaire. Next, participants were introduced to the click dummy realized through a website builder. Setup and basic features were explained in the regular operation mode. Participants were given time to familiarize themselves with the prototype and ask questions about it. Like buttons on a website, participants could click on the essential buttons to explore the basis features of the prototype on every screen, including the "touchscreen". Whenever a button was not clickable, the participant was told so and informed about the intended functionality of this button. Afterwards, participants completed the VDL questionnaire for the first time and were then given the task of resolving three scenarios with disturbances in randomized order using the HMI prototype. Again, they were able to click through the click-dummy to resolve the disturbance. A final notification indicated a successful resolution of the disturbance. After each scenario, they were asked about the particular reason for the disturbance and the suggested steps to fix it. In the online questionnaire, they completed SART, NASA-TLX, and the SEEV-based questionnaire. Next, a structured interview took place. Finally, the participants filled in the PSSUQ questionnaire and the second turn of the VDL questionnaire. The study concluded with a short debriefing. The whole procedure took approx. 75 minutes per participant.

2.7. Data Analysis

Quantitative data was analyzed using descriptive and basic inferential statistical methods. The primary tests were nonparametric Wilcoxon rank sum tests to investigate whether the empirical means differed significantly from the scale mean in the direction the respective criterion suggested. Qualitative data was analyzed using elements of the qualitative content analysis approach by Mayring [48]. Methods were chosen to summarize and structure the data. To *summarize* the verbal comments of the participants, a summary content analysis was carried out. Its objective is the consideration of all material and to systematically reduce it to the key points. In order to *structure* the verbal data, first scaling structuring was performed as a frequency analysis that counted how often features and information parameters were mentioned. Furthermore, the participants' importance rating of features and information parameters missed were evaluated. A list of features and information parameters organized by decreasing mean importance ratings and the decreasing number of mentions was created. Finally, formal structuring was achieved by classification using a pre-existing categorization system.

3. Results

The following section presents the results of this paper. Mostly, inferential tests were executed in comparison to the scale mean since no other scheme for assessing the results existed. All inferential statistics are the result of nonparametric tests as for some of the scales tested, the normality assumption was violated.

3.1. Criterion 1: Features

According to Criterion 1, the remote-operation workstation must provide necessary features to monitor the automation, provide disturbance information and support remote-operator with resolving the disturbance. Table 2 presents inferential statistics regarding Criterion 1. For all three tested scenarios, the mean of the subscale "Resolving Disturbances" of the SEEV-related questionnaire that measures the HMI's capability to resolve disturbances (1 = "poor" to 5 = "good", or equivalent) was significantly greater (all $p < 0.05$) than the subscale mean, $M_{\text{crit}} = 3.00$ (H$_1$: $\mu > 3$, H$_0$: $\mu \leq 3$), indicating that the experts found the prototype effective to help to resolve disturbances. For all features, the means

of the usefulness ratings (1 = "not useful at all" to 5 = "very useful") were significantly greater (all $p < 0.01$) than the subscale mean, $M_{crit} = 3.00$ (H$_1$: $\mu > 3$, H$_0$: $\mu \leqq 3$), indicating that the features of the prototype were considered useful.

Table 2. Descriptive and inferential statistics regarding the criterion "features".

Construct	M_{emp}	SD_{emp}	95% CI [LL, UL] [1]	M_{crit}	V^2	p (est.) [3]
SEEV: Resolving Disturbances						
Scenario A	3.80	1.16	[3.06, 4.53]	3.00	56.00	<0.05
Scenario B	3.92	1.38	[3.04, 4.79]	3.00	63.00	<0.05
Scenario C	4.67	0.44	[4.38, 4.95]	3.00	78.00	<0.001
Usefulness Rating of Features						
"Map screen"						
Selection of shuttles	4.83	0.39	[4.59, 5.08]	3.00	78.00	<0.001
Adjustment of map	4.92	0.29	[4.73, 5.10]	3.00	78.00	<0.001
Display of following stops	5.00	0.00	[5.00, 5.00]	3.00	78.00	<0.001
Display of driving path	4.58	0.90	[4.01, 5.16]	3.00	76.00	<0.01
Display of traffic density	3.75	0.62	[3.36, 4.14]	3.00	36.00	<0.01
Overall	4.64	0.22	[4.50, 4.78]	3.00	78.00	<0.01
"Disturbances screen"						
Communication bar	4.50	1.00	[3.86, 5.14]	3.00	64.50	<0.01
Display of incoming disturbances	4.92	0.29	[4.73, 5.10]	3.00	78.00	<0.001
Display of disturbances currently processed	4.58	0.67	[4.16, 5.01]	3.00	66.00	<0.01
Pop-up-window to overcome disturbance	4.67	0.49	[4.35, 4.98]	3.00	78.00	<0.001
Pop-up-window to check assumptions	4.17	0.94	[3.57, 4.76]	3.00	36.00	<0.01
Overall	4.61	0.46	[4.32, 4.90]	3.00	78.00	<0.01
"Video screens"						
Overall	4.75	0.62	[4.36, 5.14]	3.00	66.00	<0.001
"Details screen"						
Display of state	4.33	1.07	[3.65, 5.02]	3.00	53.50	<0.01
Display of position	4.25	0.62	[3.86, 4.64]	3.00	66.00	<0.01
Display of next stops	3.92	0.51	[3.59, 4.24]	3.00	55.00	<0.01
Display of estimated times of departure	4.25	0.45	[3.96, 4.54]	3.00	78.00	<0.001
Display of actual times of departure	4.33	0.49	[4.02, 4.65]	3.00	78.00	<0.001
Selection of cameras	4.50	1.00	[3.86, 5.14]	3.00	64.50	<0.01
Overall	4.30	0.45	[4.01, 4.58]	3.00	78.00	<0.01
"Touchscreen"						
Overall	4.58	0.67	[4.16, 5.01]	3.00	66.00	<0.01
Features overall	4.57	0.26	[4.41, 4.73]	3.00	78.00	<0.01

[1] *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. [2] *V* is the test statistic and indicates the sum of positive rank numbers. [3] Parameter *p* can only be estimated since ties exist in the data.

Table 3 provides an overview of liked features named by at least two experts, how many experts mentioned them, the category they are associated with, and a typical utterance related to the respective feature (the complete Table S1 with all mentions can be found in the Supplementary Materials). Categories with most mentions are the "disturbances screen" (14 mentions), design (12), camera view (5), "touchscreen" (2), shuttle details (1), and map view (1). Liked features that were mentioned by one expert only are the adjustment of menus, the number of menu levels, integration of disturbance notification and map view, and the acceptance of disturbance notifications, for example.

Table 4 presents missed features as openly named by at least two experts, categorized by screen, their average importance ratings (1 = "not important at all" to 5 = very important"), and the number of participants who mentioned them (the complete Table S2 with all mentions can be found in the Supplementary Materials). Missed features that were mentioned by one expert only are the display of vehicle type, an option to directly order a substitute vehicle, one click only necessary to accept disturbance notifications, the display of street names, a camera with 360° view inside the vehicle, and the documentation of previous trajectories, for instance.

Table 3. Overview of features liked.

Feature Liked	N	Typical Utterance
Overall Design		
Division of Screens	4	"I like how the information is distributed across several monitors."
Clarity	2	"I liked the interface very much, it is very clear and logically structured."
Display of Relevant Information	2	"I was not disturbed by unnecessary notifications."
"Disturbances Screen"		
Steps to Overcome Disturbance	4	"The process is logical, practicable, and pretty clear."
Communication Link	2	"You can directly get in touch with other people."
Presentation of Disturbances	2	"Whenever something was not in order, the details about it were presented with an exclamation mark."
"Disturbances Screen"	2	"I appreciate that the central screen is reserved for incoming disturbance notifications."
Distribution of Tasks for Processing Disturbances	2	"Accepting a task makes clear who is responsible for what."
"Touchscreen"		
Waypoints	2	"Setting waypoints is useful to get the shuttle away from the road."
"Video Screens"		
Video Images	5	"The video images are very helpful."

Table 4. Overview of features missed.

Feature Missed	N	$M_{\text{Importance}}$	Typical Utterance
"Details Screen"			
Relevant Information about Disturbances	2	4.50	"It would be better to focus on important things"
Less Screens	2	4.50	"There should not be more than five monitors"
"Disturbances Screen"			
Prioritization of Disturbances (Using Color-Coding)	3	5.00	"Priority should be given to passenger emergency calls."
Visual Highlighting of Incoming Notifications	3	4.67	"Incoming notifications should be visually highlighted"
Auditory Highlighting of Incoming Notifications	2	4.50	"I want to hear a sound when a notification comes in"
"Map Screen"			
Clear (Colorful) Display of Trajectory	2	4.50	"Highlight the trajectory by making it bold or using colors. The dashed lines are not clear"
"Video Screens"			
Moving Cameras	3	4.00	"Being able to control the cameras, for example move them or zoom in, would be a meaningful improvement"
Less Screens	2	4.00	"One monitor is enough"
Camera Showing Vehicle from Outside	2	3.50	"Having a bird-view outside camera would help avoiding the blind spot"

3.2. Criterion 2: Information Parameters

According to Criterion 2, the remote-operation workstation must provide necessary information to monitor the automation, provide disturbance information and support them with resolving the disturbance. Overall, the experts assessed the information parameters provided by the prototype as very important ($M = 4.43$, $SD = 0.91$, 1 = "not important at all" to 5 = "very important"). The highest mean ratings were given for the information presented on the "details screen" ($M = 4.54$, $SD = 0.54$), followed by "touchscreen" and "map screen" ($M = 4.58$, $SD = 0.90$; $M = 4.48$, $SD = 0.77$) and, finally, the "disturbances screen" ($M = 4.27$, $SD = 1.08$).

Table 5 presents missed information parameters as openly named by at least two experts, categorized by screen, their average importance ratings, and the number of partici-

pants who mentioned them (the complete Table S3 with all mentions can be found in the Supplementary Materials). Missed information parameters that were mentioned by one expert only are previous stops, the view of the vehicle's bottom, the overall system state, lateral cameras, and numbers of vehicles available, for example.

Table 5. Overview of missing information parameters.

Information Parameter Missed	N	$M_{\text{Importance}}$	Typical Utterance
"Details Screen"			
Occupancy	4	4.25	"I want to know how many passengers are in the vehicle."
"Map Screen"			
Exact Position with Street Names and House Numbers	2	5.00	"The exact position with street names house numbers should be shown in the map"
Information on Infrastructure and Other Additional Layers	2	3.00	"Elements from the infrastructure, such as cameras on stops, should be visible on the map"

3.3. Criterion 3: Situation Awareness

According to Criterion 3, the remote-operation workstation must provide a high level of situation awareness to the remote-operator. Table 6 presents the statistics of measures related to situation awareness. The overall means of the SART questionnaire (1 = low to 5 = high situation awareness) were not significantly different (all $p > 0.05$) from the scale mean of 3 for any scenario (H_1: $\mu \neq 3$, H_0: $\mu = 3$). This result suggests an average degree of situation awareness perceived in each scenario. Regarding the SEEV-related items' subscale Projection of Future (1 = poor to 5 = high projection of future), the empirical means are significantly larger ($p < 0.05$) than the scale mean of 3 for Scenarios A and C (H_1: $\mu > 3$, H_0: $\mu \leqq 3$), with an insignificant difference for Scenario B ($p > 0.05$).

Table 6. Descriptive and inferential statistics regarding situation awareness.

Construct	M_{emp}	SD_{emp}	95% CI [LL, UL] [1]	M_{crit}	V [2]	p (est.) [3]
SART Overall						
Scenario A	3.18	0.48	[2.87, 3.48]	3.00	54.50	0.24
Scenario B	3.05	0.33	[2.84, 3.26]	3.00	40.50	0.53
Scenario C	2.98	0.37	[2.74, 3.22]	3.00	37.50	0.93
Projection of Future (SEEV)						
Scenario A	3.88	1.03	[3.22, 4.53]	3.00	58.50	<0.05
Scenario B	3.21	1.23	[2.42, 3.99]	3.00	33.50	0.28
Scenario C	4.17	0.58	[3.80, 4.53]	3.00	78.00	<0.01

[1] *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. [2] *V* is the test statistic and indicates the sum of positive rank numbers. [3] Parameter *p* can only be estimated since ties exist in the data.

3.4. Criterion 4: Usability

According to Criterion 4, the remote-operation workstation must have good usability. As Table 7 shows, usability (1 = low to 7 = high usability) is significantly larger (all $p < 0.05$) than the scale mean (H_1: $\mu > 4$, H_0: $\mu \leqq 4$), both overall and for each of the three subscales.

3.5. Criterion 5: User Acceptance

According to Criterion 5, the remote-operation workstation must have a high user acceptance. As shown in Table 8, overall user acceptance (1 = low to 5 = high user acceptance) is significantly greater ($p < 0.01$) than the scale mean (H_1: $\mu > 3$, H_0: $\mu \leqq 3$). This is true both for the assessment immediately after the trial period, as well as after the resolution of the disturbance scenarios. No significant difference between the pre-test and the post-test was reported ($V = 63$, $p > 0.05$). This finding indicates a high

degree of acceptance that did not decrease after the prototype was used by the experts in realistic scenarios.

Table 7. Descriptive and inferential statistics regarding usability.

Construct	M_{emp}	SD_{emp}	95% CI [LL, UL] [1]	M_{crit}	V [2]	p (est.) [3]
PSSUQ Overall	5.51	1.17	[4.77, 6.26]	4.00	63.00	<0.01
System Usefulness	5.74	1.14	[5.02, 6.46]	4.00	66.00	<0.01
Information Quality	5.40	1.10	[4.70, 6.11]	4.00	73.00	<0.01
Interface Quality	5.25	1.70	[4.17, 6.33]	4.00	57.00	<0.05

[1] *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. [2] *V* is the test statistic and indicates the sum of positive rank numbers. [3] Parameter *p* can only be estimated since ties exist in the data.

Table 8. Descriptive and inferential statistics regarding user acceptance.

Construct	M_{emp}	SD_{emp}	95% CI [LL, UL] [1]	M_{crit}	V [2]	p (est.) [3]
VDL Overall						
Pre-test	3.93	0.71	[3.48, 4.37]	3.00	54.00	<0.01
Post-test	4.02	0.84	[3.49, 4.55]	3.00	74.00	<0.01

[1] *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. [2] *V* is the test statistic and indicates the sum of positive rank numbers. [3] Parameter *p* can only be estimated since ties exist in the data.

3.6. Criterion 6: Attention

According to Criterion 6, the remote-operation workstation must direct the user's attention to information that is currently relevant. Table 9 shows that the means of all SEEV scores (1 = low to 5 = high attention or respective construct) are significantly larger than the scale mean, 3, for all scenarios investigated (H[1]: $\mu > 3$, H[0]: $\mu \leq 3$). In addition, the subscale presentation of information that is conceptually linked to attention shows the same result: For every scenario, the subscale mean scores are significantly larger than the scale mean.

Table 9. Descriptive and inferential statistics regarding attention.

Construct		M_{emp}	SD_{emp}	95% CI [LL, UL] [1]	M_{crit}	V [2]	p (est.) [3]
SEEV Overall							
Scenario A		3.75	0.94	[3.15, 4.35]	3.00	67.00	<0.05
Scenario B		3.51	1.26	[2.71, 4.31]	3.00	57.50	0.08
Scenario C		4.34	0.43	[4.07, 4.62]	3.00	78.00	<0.01
SEEV Presentation of Information							
Scenario A		3.67	0.96	[3.06, 4.27]	3.00	55.00	<0.05
Scenario B		3.46	1.35	[2.60, 4.32]	3.00	55.00	0.11
Scenario C		4.27	0.62	[3.88, 4.66]	3.00	66.00	<0.01
SART							
Scenario A	Att.[4] Demand	2.36	0.64	[1.95, 2.77]		132.00	<0.001 [5]
	Att. Supply	3.88	0.62	[3.48, 4.27]			
Scenario B	Att. Demand	2.03	0.89	[1.46, 2.60]		138.00	<0.001
	Att. Supply	3.98	0.76	[3.50, 4.46]			
Scenario C	Att. Demand	1.64	0.76	[1.16, 2.12]		138.00	<0.001
	Att. Supply	3.60	0.58	[3.24, 3.97]			

[1] *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. [2] *V* is the test statistic and indicates the sum of positive rank numbers. [3] Parameter *p* can only be estimated since ties exist in the data. [4] "Att." = "Attentional". [5] Wilcoxon rank sum test between Subscales "Attentional Demand" and "Attentional Supply".

An important aspect of this criterion is the use of attentional resources. For all scenarios, attentional demand is significantly lower (all $p < 0.001$) than attentional supply, both

measured by the respective subscales of the SART questionnaire (1 = low to 5 = high attentional demand or supply, respectively), H_1: $M_{Supply} > M_{Demand}$, H_0: $M_{Supply} \leqq M_{Demand}$. This finding implies that attentional resources are not depleted or even exceeded by using the remote-operation workstation.

3.7. Criterion 7: Capacity

According to Criterion 7, the remote-operation workstation must not overwhelm the user's mental and physical capacities. As indicated in Table 10, for all scenarios, the mean scores for the overall NASA-TLX questionnaire (1 = low to 21 = high workload) are significantly lower ($p < 0.01$) than the scale mean, 11 (H_1: $\mu < 11$, H_0: $\mu \geqq 11$), indicating a lower workload.

Table 10. Descriptive and inferential statistics regarding capacity.

Construct	M_{emp}	SD_{emp}	95% CI [LL, UL] [1]	M_{crit}	V^2	p (est.) [3]
NASA-TLX						
Scenario A	5.63	4.56	[2.70, 8.55]	11.00	5.00	<0.01
Scenario B	6.04	4.76	[3.01, 9.07]	11.00	6.00	<0.01
Scenario C	3.47	2.75	[1.72, 5.22]	11.00	0.00	<0.01

[1] *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. [2] *V* is the test statistic and indicates the sum of positive rank numbers. [3] Parameter *p* can only be estimated since ties exist in the data.

3.8. Additional Improvement Suggestions

As an additional explorative section that did not consider subjective importance, participants could make further suggestions for improvements at the end of the structured interview, as shown in Table 11 (at least two mentions, complete table with all mentions see Table S4 in Supplementary Materials). Most mentions concerned the "disturbances screen" (16 mentions), followed by design (11), "touchscreen", "details screen", "map screen" (4 each), and finally, "video screens" (2). Examples for improvements suggested by one expert only are a customizable distribution of information across screens, the integration of the interface in existing control center operations systems, showing actual departure times only, and the exact position of shuttle.

Table 11. Overview of additional suggestions for improvements.

Improvements Suggested	N
Design	
Only Relevant Information/Clear Presentation	3
Prioritize/Highlight Particularly Important Information	2
Less Monitors/All Video Images on one Monitor	2
Monitor for Overview	2
"Disturbances Screen"	
Only one Click for Accepting and Editing Disturbance Notification	3
Categorizing Incoming Disturbance/Delay Notifications	3
Highlighting/Prioritizing Incoming Disturbance Notifications	3
Improving Check of Prerequisites for Clearance (Making Them Faster/Customizable/Immediately After Each Step in Disturbance Resolution Process)	3
Documenting Prerequisite Checks for Clearance/Disturbances	2
"Touchscreen"	
Marking Bypasses	2

4. Discussion

This study evaluated the usability of a prototype of a novel HMI for the teleoperation of highly automated vehicles (SAE level 4). The quantitative questionnaire indicators as well as the qualitative feedback provided in a structured interview confirm the HMI concept's usability and capability to monitor and control AVs by fulfilling all claimed

criteria and provide valuable insights for refinement of the prototype. The following section discusses implications of the results, delineates potential optimizations of the prototype, transfers the HMI concept to other scenarios as well as other vehicles types, and points out limitations of the study.

4.1. Interpretation of Results

Overall, our findings support the presented HMI concept and establish it as a suitable interface design for the teleoperation of highly automated vehicles in public transport. All criteria were fulfilled to a satisfying extent regarding its early stage in the design process with some potential for further optimization in the following iterations. All the *features* the prototype contains were considered highly relevant by the participants (Criterion 1). The "video screens" and the "map screen" received particularly high ratings. Watching the video stream, getting notified about disturbances, and being guided through the disturbance resolution process can therefore be considered inevitable. In a similar vein, the features the expert evaluators liked were related to the display of video images, the process to overcome disturbances, and the distribution of information across the screens. A common feature missed by the participants was the prioritization of important disturbance notifications, such as emergency calls, and increasing their salience by using color-coding. Highlighting incoming notifications in general, both visually and acoustically, was another feature that was mentioned multiple times and obtained high importance ratings. The results regarding the prototype's *information parameters* (Criterion 2) are similar to those of the features (Criterion 1). Multiple participants missed information about occupancy of each shuttle and its exact position including numbers of the closest building. Information on the infrastructure were mentioned twice but regarded less important.

A number of constructs was investigated using quantitative measures. Regarding *situation awareness* (Criterion 3), the results were around the scale mean. Thus, a medium extent of situation awareness was reached. This finding either concerns the HMI concept, which would make it relevant for optimization, or it is a contingency of the way the prototype was implemented, its lack of visual sophistication resulting from its early developmental stage as an early prototype. Another argument in favor of the latter explanation is the subscale Projection of Future of the SEEV-related questionnaire. Situation awareness is defined by anticipating upcoming developments in the environment, based on accurately perceiving and comprehending the situation. Thus, if the future can be projected, the fulfillment of the prior steps can be assumed. For Scenarios A and C, this subscale's score is significantly above the scale mean, supporting the presence of situation awareness.

A key construct investigated in the study is *usability* (Criterion 4). The system's usefulness is particularly appreciated by the participants but also the other two subscales information quality and interface quality were given mean scores ratings significantly above the scale mean. Of similar importance are the results obtained for *user acceptance* (Criterion 5), a construct related to usability and an indicator for the satisfaction with the HMI concept. The participants assigned rating values significantly above the scale mean. Hence, the concept can be considered user-friendly. The construct *attention* (Criterion 6) was investigated in two different aspects: On the one side, the HMI concept's capability to support the remote-operator to maintain sustained attention, or *vigilance*, was measured following Taylor's approach [47]. For all scenarios, average subscale scores for attentional supply outweighed those for attentional demand, indicating a surplus of attentional resources. On the other side, attention can be examined as the distribution of attention across monitors. The questionnaire following the SEEV Model [26], particularly its subscale Presentation of Information, provided evidence that information was spread out across screens in a way that assured attention was directed where it was needed, particularly in Scenarios A and C, both of which showed subscale means significantly above the theoretical scale mean. This finding might have been influenced by the setup of the monitors as classic screens instead of head-mounted displays (HMDs). HMDs were not found to improve driver's performance or controllability but even increased error rates

under some circumstances [33,49]. Finally, the physical and mental *capacities* (Criterion 7) as operationalized by workload was significantly below the scale mean in all scenarios examined. This implies a rather low workload. This is an ambivalent finding since in monitoring automated processes, an underload of mental demand can lead to poorer performance when taking over control from the system. However, other findings came to the opposite conclusion, blaming an overload for a decrease in performance [25]. This implies that a future optimization of the prototype might come with additional workload to achieve optimum performance.

All in all, the findings provide considerable value for the evaluation of the HMI concept for several reasons. First, the expert sample was highly suitable for assessment since it consisted of the selective group of experienced control center professionals from public transport services across Germany. They will be the primary users of the interface so taking their needs into account is pivotal for the acceptance of the interface. Even with an affinity for technology that was found to be only slightly above the scale mean, the novel interface was openly accepted, with acceptance ratings remaining stable after it was put to the test by having the experts resolve three relevant scenarios with disturbances. Second, the methodology chosen was thorough and tailored to the assessment of the current status of the prototype so that the HMI concept's suitability to monitor shuttle operations and intervene, when necessary, could be confirmed. Concrete ideas for modifications could be obtained. To fulfil the former objective, quantitative evaluation methods were used. The latter objective was reached by conducting in-depth structured interviews that provided both the time and the framework needed. Open mentions helped avoid imposing ideas onto the participants and thus restrain their creative thinking, existing categories helped structure the participants' thoughts and encouraged them to explore areas they might not have thought about by themselves.

4.2. Refinement of the Prototype

From the most frequent mentions with the highest urgency ratings, a list of improvements was extracted. The most concrete, substantial, and feasible ones were selected to provide a list of solutions that can be applied in the following iterative cycle. In order to raise situation awareness, the HMI concept should display realistic video stream instead of static images after the next iteration of the design process. Additional camera perspectives should be provided to rule out the possibility of blind spots. A 360° view might not significantly improve situation awareness, though. Instead, camera views should be adjustable, for example by physically moving the camera. Situation awareness could be further improved by providing more exact information on the current position that relates to the remote-operator's pre-existing knowledge as a control center professional, such as the display of street names, building numbers, important landmarks, and intersections. On the "disturbances screen", a visual signal should be provided to highlight incoming disturbance notifications. An acoustic signal could be added for notifications with top priority only to prevent an inflation of notifications and a distraction of coworkers in the control center. Disturbances should be prioritized depending on their severity and the need for immediate action. Different categories of disturbances should be distinguishable by color to raise salience of particularly urgent notifications. The colors should be unequivocal for users suffering from color vision deficiencies. The disturbance resolution process should be sped up by combining the "Accept" and "Edit" commands. Checking the assumptions to continue the shuttle's ride should not be at the very end of the disturbance resolution process but immediately after the respective step. The actions taken should be documented in a system to support reporting procedures. On the "details screen", shuttle operation features should only be highlighted in color when a malfunction exists to provide a better overview. Instead of many information parameters on each shuttle's state, a quick overview of the aggregated state of the whole fleet should be provided. Finally, since the current workload proved to be rather low, bearing potential negative implications for vigilance as described above, it is conceivable to increase the number of shuttles for which a remote

operator is responsible to increase the remote-operator's workload so that a medium level of mental demand is assured.

4.3. Transfer of the Prototype

This section investigates the transferability of the HMI concept both to other scenarios as well as other vehicle types.

4.3.1. Transfer to Other Scenarios

In this evaluation study, the proposed HMI concept was assessed based on three relevant scenarios. However, this HMI concept can cope with a wide range of scenarios beyond the tested ones. It should still be considered that there are situations that can be handled from the control center to a limited extent only, such as in the following scenarios. First, a person gets *injured* in an accident the shuttle is involved in. The proposed HMI can only partially assist to resolve this issue: For summoning medical assistance, an immediate call can be placed via the communication bar on the "disturbances screen". However, first aid cannot be provided since no driver or operator is on board of the shuttle. This can be particularly fatal in case no other passengers are on board to assist the injured one. A loudspeaker through which the remote-operator could address the immediate surrounding of the shuttle could help to communicate with passersby to seek immediate medical aid.

Second, weather events or construction sites could emit *disturbing objects* such as slush or dirt that block both the sensors and the cameras. This would result in diminished situation awareness, preventing the remote-operator from checking the local situation and giving clearance to continue the ride. Collecting additional information from other local sources, such as intelligent road-site units or additional sensors that are not susceptible to blocking objects or sending staff to the location to clean sensors and cameras, could prevent this. To request a substitute vehicle, an automated or at least visualized process dialogue could be added to the interface in addition to simply using the communication bar. In case that curious passersby block the doorway, a loudspeaker to address them could help the remote-operator resolve the case, just as in the scenario outlined above.

Third, the communication link to the passengers could be impaired, resulting in *noise* that could prevent the remote-operator from understanding important passenger calls. This would be particularly troublesome in case of emergencies. Installing a backup system such as a second communication mode to put through high-priority calls could remedy this shortcoming. Again, intelligent road-site units that are themselves linked to a stable cable-based connection could work as a relay, maintaining a connection to the shuttle when wireless connection fails. For the interface, this would imply channeling all communication modes into the existing system, the communication bar, and automatically select the most stable connection to assure optimal connectivity.

Fourth, a scenario is conceivable in which the GPS receiver on board of the shuttle is out of order, failing to report the shuttle's *position* to the remote-operator. To overcome this problem, an algorithm that interpolates the shuttle's trajectory based on the latest velocity data could provide an estimation of the current position. The interpolated trajectory could be presented on the "map screen" using a dashed line in another color, provided that the GPS-based trajectories are presented by solid lines.

Fifth, in case of a software issue that disables the calculation of *trajectories* based on waypoints that were set by the remote-operator, the semi-automated process of parking the shuttle in a designated area is not available any longer. In this scenario, adding manual teleoperation to the set of features could take over. This option would require adding manual driving facilities such as for example a steering wheel and pedals to the interface and integrate the suggestion of manual teleoperation into the established disturbance resolution process. The process itself, however, could be directly transferred from the existing solution.

4.3.2. Transfer to Other Vehicle Types

It is generally feasible to upscale the HMI to either other categories of vehicles and/or a context of use other than the shuttle. The following section is a non-exhaustive list of vehicle and context types, how an HMI for teleoperation could be used with them, whether adjustments to the current design would need to be made for this, and how. First, cars that are part of a *corporate-owned vehicle fleet* could benefit from teleoperation. In addition to taking over control to resolve disturbances, teleoperation could serve as an integral part of the fleet. Teleoperation could be utilized to supply vehicles to employees or customers as well as maintenance facilities, so the ride or walk to and from the pickup facility would become obsolete. The basic setup and functionality of the HMI proposed here would also work in this context. However, additional elements and interfaces with disposition, planning and payment tools are conceivable. A likely difference is the size of the geographical area the vehicles would cover. A solution could be shared responsibilities for subareas with a separate remote-operator in charge of each subarea.

Second, vehicles could be used in the context of *delivering* goods. For this scenario, the current HMI would fit as well. However, some features would need to be adapted. The option to speak to the vehicle's passengers would not be necessary since only goods would be transported. Likewise, emergency calls from the passengers would not play a role. However, the safe securing of cargo would be key. The HMI could enable the remote-operator to monitor the goods by viewing a video stream from the cargo section of the vehicle. The "details screen" could show additional information on the customer, the route, and the goods transported. Instead of stops to pick up and drop off passengers, logistic centers and customers would be displayed on the "map screen". In the event of disturbance, no passenger can be consulted, thus all the information required to resolve the disturbance must be provided by the interface.

Third, a *combination* of multiple purposes could be controlled by this HMI. An example for such a vehicle is DLR's research project series "U-Shift", an automated driving module that can be used to transport both passengers and goods using different types of capsules [39]. In this case, all the features described above could be combined to a "toolbox". Just like the intermodal setup of the vehicle, a set of modules that could be included or excluded, based on the current context's requirement. The current HMI provides a solid basis for this idea of a modular purpose-oriented HMI as it comes with a plethora of features already.

Fourth, the HMI could be applied to *crisis intervention* vehicles that deliver goods in dangerous terrain with poor infrastructure without risking attacks on the intervention staff. The automated vehicle would need to be operated in difficult terrain, where standardized automatable use cases are hard to implement due to a variety of events that cannot be anticipated. To recognize particularly different terrain that is covered with land mines, for instance, a "heat map" could be added as an additional layer on the "map screen" of the HMI. Within the project Autonomous Humanitarian Emergency Aid Devices (AHEAD), a collaboration of DLR with the World Food Programme, remote-controlled trucks will deliver supplies to their destinations without risking the staff's physical integrity [40].

Fifth, mobile objects other than vehicles are also conceivable to work with this HMI. Examples are delivery bots that use the walkways for delivering small goods, such as food or orders from retailers [50,51], and ships, both for off-shore and inland shipping. Since the context of use differs considerably from the original one, particularly for the latter, the HMI would require more fundamental adjustments to comply with the common control practices and kinds of data needed to monitor and control the respective object. However, setting waypoints might provide a feasible solution in these contexts as well. Additional requirements could be met by enhancing the HMI with modules, e.g., a display to present radar data and intercom connections to water police and other ships, in the latter case.

4.4. Limitations

There is the need to emphasize that the prototype evaluated in this paper is still on an early stage of development. Its main purpose is getting a first impression on whether the development is in the right direction, particularly whether its overall setup is valid. Compared to the final interface, it possesses some shortcomings: Not all of its features are fully implemented and clickable. The interface design is not yet fully developed and thus not visually appealing. This criticism had been made by some of the participants. It cannot be ruled out that this might have biased their impression of the prototype and therefore the evaluation feedback they gave. In addition, the HMI prototype was designed for control center professionals within public transport in urban mobility networks in Germany. Even though using it in other contexts of use appears feasible with minor adaptations, this must be accompanied by further research. Since control centers of public transport and the tasks of their employees differ considerably by country, further studies to validate the HMI concept in other countries need to be conducted. Task and skill analyses of control center professionals may help to adjust the prototype to the respective country. It is also conceivable that a distinct group of professionals specifically trained for teleoperations needs to be deployed, as their tasks will differ substantially from traditional work in control centers.

Moreover, the generalizability of the evaluation is limited since the participants that served as evaluators were experts for control center operations. That means they know the context of use well and represent the typical users that are the benchmark of the design process as their skills and workflows need to be considered to assure a smooth transition from the work as a regular control center professional to a remote-operator of automated vehicles. However, due to their professional background, they are only able to focus on the limited context of use in a public transport control center and are not experts for usability. Therefore, they do not come with a structured understanding of the requirements that need to be met in order to identify potential downfalls of using the interface, particularly those that did not become evident in the evaluated prototype.

Regarding the quantitative data, only basic statistical analyses could be conducted due to the low number of participants. The failure to meet assumptions needed to use tests with a higher power might have resulted in not detecting significant differences between groups. Instead, the theoretical scale means were mostly used to provide a crude estimation of whether a criterion was met or not. In the next step, several HMI designs could be developed and systematically tested against each other or against a baseline, such as an existing remote-operation solution from a different context of use. This procedure could help assess suitability of the HMI for its future context of use more precisely.

Furthermore, future research could consider the specific requirements for monitoring and operating an entire fleet of vehicles by a small number of remote-operators. This helps to make operations more efficient and thus cost-effective but likely poses additional cognitive demands onto the remote-operators. Also, a wider range of scenarios could be implemented in the prototype and evaluated. Finally, an alternative or updated HMI concept could be created, e.g., considering the results of this evaluation study, and empirically tested against the presented one.

5. Outlook

Not an abundancy of HMI concepts exists for teleoperation, at least not in the context of public transport. To the authors' knowledge, the presented concept is the first HMI tailored to the context of public transport control centers, equipped with a guided step-by-step disturbance resolution process and semi-automated steering by setting waypoints. Thus, unlike many other HMIs, it does not rely on direct control. Following the notion of user-centered design, this paper presents an early stage of the HMI that was evaluated by a highly selective group of experts, its future users. The results of this evaluation study support the design and delineate approaches to further optimize it. The next steps in this research project will therefore be the refinement of the prototype and a repeated evaluation

process that keeps its future users on the loop but at the same time involves usability experts to determine its overall user-friendliness. This could turn out as crucial when it comes to the transfer of the HMI to other scenarios and/or vehicle types. Subsequently, a physical, fully operable remote-operation workstation will be developed and set up. It will then be able to be tested extensively in a setting that is closer to the real world and may therefore provide clearer and more detailed insights into its usability, suggesting further potential for optimization and eventually paving the way to pilot studies in actual control centers of public transport. Hence, teleoperation combined with a user-friendly, widely accepted interface could open the door for automated driving in public transport, enabling a wide range of people to benefit from the possibilities of system automation.

References

1. IHS. *Autonomous Cars: Not If but When*; IHS Automotive: Northville, MI, USA, 2014.
2. Bengler, K.; Dietmayer, K.; Farber, B.; Maurer, M.; Stiller, C.; Winner, H. Three Decades of Driver Assistance Systems: Review and Future Perspectives. *IEEE Intell. Transport. Syst. Mag.* **2014**, *6*, 6–22. [CrossRef]
3. Bhavsar, P.; Das, P.; Paugh, M.; Dey, K.; Chowdhury, M. Risk Analysis of Autonomous Vehicles in Mixed Traffic Streams. *Transp. Res. Rec.* **2017**, *2625*, 51–61. [CrossRef]
4. Gnatzig, S.; Chucholowski, F.; Tang, T.; Lienkamp, M. A system design for teleoperated road vehicles. In Proceedings of the 10th International Conference on Informatics in Control, Automation and Robotics, Reykjavík, Iceland, 29–31 July 2013; pp. 231–238. [CrossRef]
5. Chucholowski, F.; Gnatzig, S.; Tang, T.; Hosseini, A.; Lienkamp, M. Teleoperiertes Fahren-Aktuelle Entwicklungen. In Proceedings of the 6. Tagung Fahrerassistenz, Munich, Germany, 28–29 November 2013.
6. Society of Automotive Engineers. *SAE J 3016-2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*; SAE International: Warrendale, PA, USA, 2018.
7. Neumeier, S.; Wintersberger, P.; Frison, A.K.; Becher, A.; Facchi, C.; Riener, A. Teleoperation: The Holy Grail to solve problems of automated driving? Sure, but latency matters. In Proceedings of the 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, The Netherlands, 21–25 September 2019; pp. 186–197. [CrossRef]
8. Sheridan, T.B. Teleoperation, telerobotics and telepresence: A progress report. *Control Eng. Pract.* **1995**, *3*, 205–214. [CrossRef]
9. Gnatzig, S.; Schuller, F.; Lienkamp, M. Human-machine interaction as key technology for driverless driving—A trajectory-based shared autonomy control approach. In Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, Paris, France, 9–13 September 2012; pp. 913–918. [CrossRef]
10. Litman, T.A. *Autonomous Vehicle Implementation Predictions: Implications for Transport Planning*; Victoria Transport Policy Institute: Victoria, BC, Canada, 2020.
11. Schoitsch, E. Autonomous Vehicles and Automated Driving Status, Perspectives and Societal Impact. In Proceedings of the 24th annual IDIMT conference, Poděbrady, Czech Republic, 7–9 September 2016; pp. 405–423.
12. Milakis, D.; Snelder, M.; van Arem, B.; van Wee, B.; Homem De Almeida Correia, G. Development and transport implications of automated vehicles in the Netherlands: Scenarios for 2030 and 2050. *Eur. J. Transp. Infrastruct. Res.* **2017**, *17*, 1. [CrossRef]
13. Lutin, J. Not If, but When: Autonomous Driving and the Future of Transit. *JPT* **2018**, *21*, 92–103. [CrossRef]

14. Sherry, L.; Shortle, J.; Donohue, G.; Berlin, B.; West, J. Autonomous Systems Design, Testing, and Deployment: Lessons Learned from The Deployment of an Autonomous Shuttle Bus. In Proceedings of the 2020 Integrated Communications Navigation and Surveillance Conference (ICNS), Herndon, VA, USA, 8–10 September 2020. [CrossRef]

15. Rehrl, K.; Zankl, C. Digibus©: Results from the first self-driving shuttle trial on a public road in Austria. *Eur. Transp. Res. Rev.* **2018**, *10*. [CrossRef]

16. Backhaus, R. Automatic Shuttle Buses—From the Test Track to Scheduled Services. *ATZ Electron. Worldw.* **2020**, *15*, 34–39. [CrossRef]

17. Greenblatt, J.B.; Shaheen, S. Automated Vehicles, On-Demand Mobility, and Environmental Impacts. *Curr. Sustain. Renew. Energy Rep.* **2015**, *2*, 74–81. [CrossRef]

18. Fagnant, D.J.; Kockelman, K.M. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transp. Res. Part C Emerging Technol.* **2014**, *40*, 1–13. [CrossRef]

19. Hyland, M.; Mahmassani, H.S. Operational benefits and challenges of shared-ride automated mobility-on-demand services. *Transp. Res. Part A Policy Pract.* **2020**, *134*, 251–270. [CrossRef]

20. Fong, T.; Thorpe, C. Vehicle Teleoperation Interfaces. *Auton. Robots* **2001**, *11*, 9–18. [CrossRef]

21. Kay, J.S. STRIPE: Remote driving using limited image data. In Proceedings of the Conference on Human Factors in Computing Systems, Denver, CO, USA, 7–11 May 1995; pp. 107–108.

22. Kim, J.S.; Ryu, J.H. Shared teleoperation of a vehicle with a virtual driving interface. In Proceedings of the International Conference on Control, Automation and Systems, Gwangju, Korea, 20–23 October 2013; pp. 851–857, ISBN 9788993215052.

23. Chen, T.L.T. Methods for Improving the Control of Teleoperated Vehicles. PhD Thesis, Technische Universität, Munich, Fermany, 2015.

24. Sheridan, T.B. *Telerobotics, Automation, and Human Supervisory Control*; MIT Press: Cambridge, MA, USA, 1992; ISBN 0262193167.

25. Thomson, D.R.; Besner, D.; Smilek, D. A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspect. Psychol. Sci.* **2015**, *10*, 82–96. [CrossRef] [PubMed]

26. Wickens, C.D.; Goh, J.; Helleberg, J.; Horrey, W.J.; Talleur, D.A. Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors: J. Hum. Factors Ergon. Soc.* **2003**, *45*, 360–380. [CrossRef] [PubMed]

27. Wickens, C.D. Noticing Events in the Visual Workplace: The SEEV and NSEEV Models. In *The Cambridge Handbook of Applied Perception Research*; Hoffman, R.R., Hancock, P.A., Scerbo, M.W., Parasuraman, R., Szalma, J.L., Eds.; Cambridge University Press: New York, NY, USA, 2015; pp. 749–768.

28. Senders, J.W. The Human Operator as a Monitor and Controller of Multidegree of Freedom Systems. *IEEE Trans. Hum. Factors Electron.* **1964**, *HFE-5*, 2–5. [CrossRef]

29. Sanders, A.F.; Houtmans, M. Perceptual processing modes in the functional visual field. *Acta Psychol.* **1985**, *58*, 251–261. [CrossRef]

30. Boot, W.R.; Kramer, A.F.; Becic, E. Capturing Attention in the Laboratory and the Real World. In *Attention: From Theory to Practice*; Kramer, A.F., Wiegmann, D.A., Kirlik, A., Eds.; Oxford University Press: Oxford, NY, USA, 2007; pp. 27–44. ISBN 9780195305722.

31. Taylor, R.M.; Selcon, S.J.; Swinden, A.D. Measurement of Situational Awareness and Performance: A Unitary SART Index Predicts Performance on a Simulated ATC Task. In *Human Factors in Aviation Operations, Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP), Dublin, Ireland, March 1994*; Fuller, R., Johnston, R., McDonald, N., Eds.; Ashgate Publishing: Aldershot, UK, 1995.

32. Merat, N.; Seppelt, B.; Louw, T.; Engström, J.; Lee, J.D.; Johansson, E.; Green, C.A.; Katazaki, S.; Monk, C.; Itoh, M.; et al. The "Out-of-the-Loop" concept in automated driving: Proposed definition, measures and implications. *Cognit. Technol. Work* **2019**, *21*, 87–98. [CrossRef]

33. Chellali, R.; Baizid, K. What maps and what displays for remote situation awareness and ROV localization? *Lect. Notes Comput. Sci.* **2011**, *6772 LNCS*, 364–372. [CrossRef]

34. Neumeier, S.; Walelgne, E.A.; Bajpai, V.; Ott, J.; Facchi, C. Measuring the feasibility of teleoperated driving in mobile networks. In Proceedings of the 3rd Network Traffic Measurement and Analysis Conference, Paris, France, 19–21 June 2019; pp. 113–120. [CrossRef]

35. Minsky, M. Telepresence. Available online: https://web.media.mit.edu/~{}minsky/papers/Telepresence.html (accessed on 4 May 2021.).

36. International Organization for Standardization. ISO Standard No. 9421-11:2018. Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts. Available online: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en (accessed on 4 May 2021).

37. van der Laan, J. Acceptance of Advanced Transportation Telematics. *Transp. Res.* **1997**, *5*, 1–10.

38. RealLab Hamburg. Autonomes Fahren. Available online: https://reallab-hamburg.de/projekte/autonomes-fahren/ (accessed on 25 March 2021).

39. Weimer, J. U-Shift—A Novel on-the-Road Modular Vehicle Concept. Available online: https://verkehrsforschung.dlr.de/public/documents/2019/DLR_U-Shift_Presentation-EN_V1.6_0.pdf (accessed on 4 May 2021).

40. German Aerospace Center. DLR Robotic Vehicles Will Support Deliveries in Difficult Areas for the World Food Programme. Available online: https://www.dlr.de/content/en/articles/news/2020/04/20201021_remote-controlled-access-in-crisis-regions.html (accessed on 4 May 2021).

41. International Organization for Standardization. ISO Standard No. 9421-210:2019. Ergonomics of Human-System Interaction—Part 210: Human-Centered Design for Interactive Systems. Available online: https://www.iso.org/standard/77520.html (accessed on 4 May 2021).

42. Kettwich, C.; Dreßler, A. Requirements of Future Control Centers in Public Transport. In Proceedings of the 12th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications (Virtual Event), Washington, DC, USA, 21–22 September 2020; pp. 69–73. [CrossRef]

43. Franke, T.; Attig, C.; Wessel, D. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *Int. J. Hum. Comput. Interact.* **2019**, *35*, 456–467. [CrossRef]

44. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*; Hancock, P.A., Meshkati, N., Eds.; Elsevier: Amsterdam, The Netherlands, 1988.

45. Hart, S.G. NASA-Task Load Index (NASA-TLX); 20 Years Later. In Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, San Francisco, CA, USA, 16–20 October 2006; pp. 904–908.

46. Lewis, J.R. Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *Int. J. Hum. Comput. Interact.* **2002**, *14*, 463–488. [CrossRef]

47. Taylor, R.M. *Situation awareness rating technique (SART): The development of a tool for aircrew systems design. Situational Awareness in Aerospace Operations*; Routledge: London, UK, 1990.

48. Mayring, P. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*, 7th ed.; Deutscher Studien Verlag: Weinheim, Germany, 2000.

49. Georg, J.M.; Feiler, J.; DIermeyer, F.; Lienkamp, M. Teleoperated Driving, a Key Technology for Automated Driving? Comparison of Actual Test Drives with a Head Mounted Display and Conventional Monitors. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Maui, HI, USA, 4–7 November 2018; pp. 3403–3408. [CrossRef]

50. Kiwibot. Launch a Robot Delivery Service in Days with a One Stop Solution. Available online: https://www.kiwibot.com/ (accessed on 31 January 2021).

51. Postmates. Postmates Unveils Latest Addition To Its Fleet: Meet Serve—The Autonomous Delivery Rover. Cision PR Newswire. Available online: https://www.prnewswire.com/news-releases/postmates-unveils-latest-addition-to-its-fleet-meet-serve---the-autonomous-delivery-rover-300764897.html (accessed on 4 May 2021).