Enhancing Teleoperator Performance: Investigating the Impact of Audio Information in Teleoperation Tasks

*

ABHISHEK KAJALA, ANAGHA DESHPANDE, ANANYA BHADAURIA, HAMZA MOHSIN, MUHAMMAD AMIN, USAMA ASIF, VIDHI RAGHVANI, ANDREAS RIENER, and BENGT ESCHER, Technische Hochschule Ingolstadt, Germany

Abstract

This research paper presents the findings on investigating the efficacy of auditory feedback systems within teleoperation contexts, with a focus on enhancing teleoperator performance and situational awareness. Findings revealed nuanced interrelations between sensory modalities, yielding unexpected results wherein minimal differences were observed in response times across varied feedback conditions encompassing audio, audio-to-visual, and audio and visual modalities. Particularly noteworthy was the observation that audio cues alone expedited decision-making processes compared to their audio and visual counterparts, underscoring the pivotal role of real-time auditory cues in dynamic operational scenarios. Furthermore, subjective feedback underscored the intricate balance between sensory inputs and their impact on teleoperator perceptions and preferences. This research contributes to a broader understanding of feedback mechanisms in teleoperation, advocating for continued exploration and refinement of audio and visual integration to optimize operational efficiency and safety within this domain. Future inquiries may delve into alternative feedback methodologies aimed at further augmenting teleoperation efficacy.

Additional Key Words and Phrases: Teleoperation, Autonomous Vehicle, Human Centered Design, Auditory feedback

1 INTRODUCTION

Teleoperation is the capacity to replace employees in hazardous or difficult work environments. It is widely recognized as a crucial strategy and is applied in a variety of situations. Application domains include things like underwater operations, robotic inspection or navigation, space exploration, surveillance, search and rescue missions, and more (Zhu and Gedeon [23]).

Fig. 1. Teleoperation workspace

For situational awareness, the operator needs to be present inside the teleoperation room and monitor the vehicles from a distance, As shown in figure 1. The teleoperation workspace receives information on surroundings and obstacles with the help of sensor-based technology such as lidar (Chamorro et al. [2]), cameras, GPS, and radar sensors on the feedback screen. But

*Authors contributed equally to this research.

Veröffentlicht durch die Gesellschaft für Informatik e.V in L. Hirsch & M. Kurz & N. Wagener (Hrsg.): Mensch und Computer 2023 – Workshopband, 03.-06. September 2023, Rappersweil (SG)

https://doi.org/10.18420/muc2023-mci-src-403

sometimes, in case of emergencies, the teleoperator needs to switch from the autonomous operation of the vehicle to manual control (Zhang [21]). Hence, the operator needs to be physically present inside the teleoperation workspace at all times.

The autonomous shuttle bus industry's long-term plan is to manually operate vehicles from a distance through teleoperation workspace(Riener et al. [16]) . Currently, there are certain limitations that mandate the operator to be present inside the autonomous shuttle bus. During manual control, the operator perceives the situational awareness of surroundings through natural senses like hearing and seeing from the bus. The figure 2 shows the autonomous to manual operation of the vehicle.

Fig. 2. Kexi autonomous shuttle bus operation

Multi-sensory feedback system integration has become an important field of research in teleoperation (de Barros and Lindeman [4]), as decisions made in a split second might impact results. The complex function of audio technology in improving the situational awareness of teleoperators when faced with scenarios that incorporate both visual and auditory inputs. The teleoperation should be effective enough to give complete information of surrounding vehicles including auditory information.Although visual cues have historically dominated teleoperation interfaces, new developments highlight how audio feedback can enhance and supplement operator perception and decision-making.

Evaluations have revealed several major teleoperation issues, such as the sporadic misreading of impediments, cognitive overload, and the constant need for increased attentiveness (Chen et al. [3]). These difficulties underline how important it is for teleoperation systems to have efficient multi-sensory feedback methods (de Barros and Lindeman [4]).

In light of this, our study aims to close knowledge gaps by thoroughly assessing the effectiveness of auditory feedback systems in teleoperation scenarios. We seek to uncover practical insights that can guide the development and implementation of future teleoperation interfaces by closely examining user perceptions and experiences.

2 RELATED WORK

The reviewed literature provides valuable insights into various aspects of teleoperation, ranging from interface design to operator influence and testing methodologies. While existing studies have contributed significantly to our understanding of teleoperation systems, there remains a need for further research to explore the specific role of audio feedback in enhancing teleoperator performance and situational awareness.

Due to its numerous uses in various industries, teleoperation the act of remotely operating a vehicle or system—has attracted a lot of attention. Prior studies have examined many facets of teleoperation, illuminating its obstacles, progress, and possible avenues for enhancement. This section includes a review of important works that are pertinent to our study on the use of audio information in teleoperation tasks.

According to Zhu and Gedeon [23] ,studied camera viewpoint control models emphasizing multitasking scenarios in teleoperation settings. Their research showed how crucial effective camera control systems are to improving teleoperator performance in various activities. Although the primary focus of their research was visual feedback, their findings established a foundation for comprehending the intricacies of teleoperation interfaces and the necessity of feedback systems that are optimal.

According to Schuß et al. [19] investigated operator influence in automated urban shuttle buses, emphasizing the role of human operators in ensuring safe and efficient operations. Their findings underscored the importance of considering operator perceptions and experiences in developing teleoperation systems. While their study did not specifically address audio feedback, it emphasized the holistic approach needed to optimize teleoperation interfaces for real-world deployment.

In their investigation of passengers' information requirements in shared automated vehicles, Flohr et al. [7] Flohr et al. (2024) emphasized the value of user-centered design in considering a range of user preferences and situational scenarios. Their study highlighted the wider implications of human-centered design principles in molding teleoperation interfaces for increased usability and effectiveness, although it concentrated on passenger experience rather than teleoperator performance.

According to Drechsler et al. [5] introduced MiRE, a mixed reality environment for testing automated driving functions, providing insights into advanced testing methodologies for autonomous systems. While their research primarily focused on automated driving, it underscored the critical role of simulation environments in evaluating teleoperation interfaces and ensuring their robustness in real-world scenarios.

In their systematic literature review, Sankar et al. [17] explore the user evaluation of teleoperation interfaces designed for professional service robots. The authors highlight that while these interfaces have primarily been developed for general applications, there is a notable presence of specific implementations within the health and wellness sector. Additionally, the authors note a growing trend towards utilizing intuitive modalities like haptic devices, signaling a shift away from traditional input devices like mouse and keyboard. Sankar et al. advocate for the integration of adaptive interfaces, emphasizing their potential to personalize user experiences and improve task performance. The authors stress the effectiveness of using both qualitative and quantitative subjective measures to derive actionable insights for interface refinement.

Teleoperation of vehicles holds potential for future mobility solutions, bridging the gap between existing automation capabilities and completely autonomous vehicles, especially in public transport settings. Kettwich et al. [10] recently developed a novel user-centered human-machine interface (HMI) specifically designed for the remote operation of highly automated shuttles. The design of this HMI was informed by a methodical examination of various scenarios and specific demands obtained from experts in public transportation control centers. The study evaluated for usability, situation awareness, acceptance, and perceived workload using questionnaires and structured interviews. The outcomes offered strong evidence in favor of the HMI design's efficiency, especially when it came to workload, acceptability, and usability. Also offered insightful feedback for future research activities and HMI design refinement.

Our state-of-the-art research suggests that it can be insightful information on a number of teleoperation-related topics, including operator influence, interface design, and testing procedures. Even if previous research has substantially contributed to our understanding of teleoperation systems, more investigation is still required to fully grasp the precise function that audio feedback plays in improving teleoperator performance and situational awareness. With consequences for the design and development of upcoming teleoperation interfaces, our work attempts to close this gap by objectively evaluating the usefulness of audio information in teleoperation activities. The studied literature offers insightful information on a number of teleoperation-related topics, including operator influence, interface design, and testing procedures. Even if previous research has substantially contributed to our understanding of teleoperation systems, more investigation is still required to fully grasp the precise function that audio feedback plays in improving teleoperator performance and situational awareness. With consequences for the design and development of upcoming teleoperation interfaces, our work attempts to close this gap by objectively evaluating the usefulness of audio information in teleoperation activities.

3 RESEARCH QUESTION AND HYPOTHESES

In this study, we aim to address two fundamental research questions (RQs) pertaining to the integration of audio cues in teleoperation.

3.1 Research Question

RQ1: How does the presence of audio information impact teleoperator performance in teleoperation tasks, compared to audio and visual cues?

Firstly, RQ 1 investigates the influence of audio cues on teleoperator performance, contrasting scenarios where only visual cues are available against those where both audio and visual information are present. This inquiry delves into the extent to which auditory input enhances or alters the teleoperator's execution of tasks.

RQ2: How does the effectiveness of audio-to-visual conversion technology impact teleoperator situational awareness in teleoperation tasks, compared to audio and visual cues?

Secondly, RQ 2 explores the impact of audio-to-visual conversion technology on teleoperator situation awareness. By comparing scenarios utilizing audio and visual cues directly with those employing converted visual representations of audio signals, we seek to understand the effectiveness of such technological solutions in augmenting the teleoperators' comprehension of their operation environment.

Through these inquiries, our study endeavors to contribute insights crucial for designing and implementing effective audio-based interfaces in teleoperation systems, ultimately aiming to enhance operator performance and situational awareness in diverse teleoperation contexts.

3.2 Hypotheses

We propose to investigate the potential performance enhancements associated with integrating audio and visual cues in teleoperation tasks; we propose H1.
H1 states that including both auditory and visual cues will yield significant improvements in teleoperator performance metrics such as response times, task accuracy, and error rates when compared to teleoperation tasks conducted solely with audio cues (Audio condition). We hypothesize that the provision of multimodal sensory inputs, including auditory cues synchronized with visual representations, will afford teleoperators with enhanced perceptual capabilities, leading to more efficient task execution and reduced cognitive load. Through experimentation and analysis, we aim to evaluate the validity of H1 and discern the extent to which the integration of audio and visual cues, supplemented by conversion technologies, contributes to performance enhancement in teleoperation scenarios.

In alignment with the focus of augmenting teleoperator situational awareness, we propose H2 to investigate the impact of audio cues and audio-to-visual conversion on the teleoperator's comprehension of their operational environment. H2 states that the provision of auditory cues, coupled with audio-to-visual conversion technology, will significantly improve the teleoperator's situational awareness during teleoperation tasks. This hypothesis will be substantiated through subjective feedback mechanisms such as surveys and interviews, enabling a comprehensive exploration of the teleoperator's perceptions and experiences. By analyzing qualitative data alongside quantitative performance metrics, we aim to understand the nuanced aspects of situational awareness enhancement facilitated by audio cues and conversion technologies. Through empirical validation of H2, we seek to advance our understanding of the mechanisms underlying perceptual augmentation in teleoperation systems and inform the design of more effective interfaces tailored to enhance teleoperator situational awareness.

4 DESIGN AND IMPLEMENTATION

We designed a 3D simulation using Adobe Illustrator, After Effects, and Blender in which the participants were exposed to a virtual city and highway scenario with moving cars and traffic. The lab study was used to replicate the real world scenario (Ramasundaram

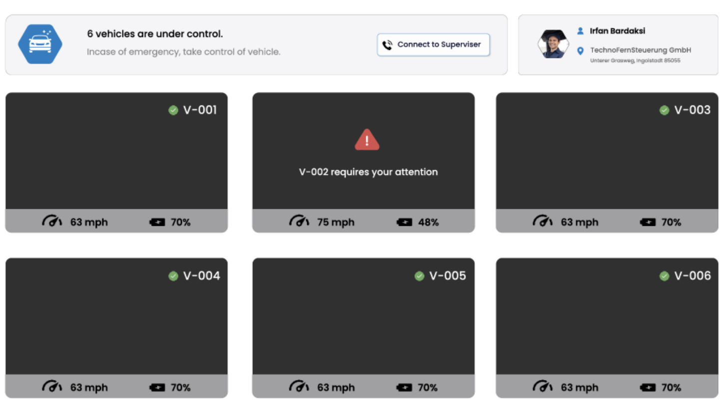

Fig. 3. Autonoumous buses monitoring screen

et al. [15]), since the main purpose of the study was about perception of auditory feedback in teleoperation. The study was designed in a controlled laboratory setting, where emergency vehicles were added in a 3D virtual city to assess the user response. The 3D video simulation was designed in accordance with the triple display monitor setup (Gallagher et al. [8]). A separate virtual environment for familiarization of subjects with this type of 3D simulated environment. The design language and color scheme were implemented considering the real-life sign boards and warnings (Li et al. [11]).

4.1 Vehicle Status Monitoring

Adhering to the European traffic rules, we designed a monitoring screen, which was added to our prototype for anticipation of subjects. The operator can monitor the status of six buses on the main center monitor screen, As shown in figure 3, where they are able to monitor the necessary information like battery and speed of the shuttle bus (Jansen [9]). In case a problem occurs on any bus, the display window related to that bus shows an alert, which indicates that something is wrong with the bus. On clicking, the operator takes control of that bus, and the main screen directs the operator to the 3D virtual city simulation.

4.2 Vehicle Control Take Over By Teleoperators

We designed a three-lane road environment setup where the autonomous bus was moving in the middle lane at the start. On the main screen, the operator can see all the predominant or primary information. The positioning of the rearview mirror, dashboard, and indicators was decided considering the existing mental model(Jansen [9]), which was the result of many iterations on initial paper design concepts. The left and right displays had views of the left and right windows of the vehicle. Hence, the operator could see the approaching vehicles from left, right, front, and behind. The left screen also contained the navigation map information that was considered as secondary information.

Fig. 4. For only audio, no visuals on the simulation screen

Fig. 5. Audio to visual representation

The prototype was designed for three different conditions, and there were four scenarios Zou et al. [24]. The first condition had simple audio, which we perceive conventionally through the ears, As shown in figure 4. The second condition was the representation of the audio in visual form with the help of alerts and other visual information shown in figure 5.

For situational awareness, the alerts were designed in a way they were positioned right below the rearview mirror. Colored borders across the screen were implemented, which appeared on all three displays to show information on the distance of the approaching emergency vehicles from the surroundings. The border appears green initially, and as the distance from the emergency vehicles decreases, the color changes to bright red to denote caution. The additional signal visual information also enhances the user experience of the teleoperation (Triantafyllidis et al. [20]), As shown in Figure 5. Along with these visual representations, another approach is used in which we have a bird-eye view of our vehicle in the bottom right of the primary middle screen. It shows the emergency vehicle's directional information. If the emergency vehicle is approaching from behind, it indicates the alert symbol from behind, As shown in Figure 6.

Fig. 6. Bird-Eye View for directional information

During the scenario the participants were exposed to a virtual city environment with the aim to measure the response time and task completion rates of the participants. Additionally, for each scenario, we designed a takeover that requires the subjects to press the button on the keyboard and make the right decision based on their auditory perception. The buttons were "W, A, S, D keys," which are used commonly for movement [18]. For "W," the bus is supposed to move in the same lane, and "A" for moving in the left lane. "S" is for stop, and "D" is for moving in the right lane. In each scenario, when some emergency situation arises, the subject needs to make wise decisions based on what they actually perceive about auditory information.

4.3 Code integration for Response Time Calculation

In order to make response times for teleoperation activities easier to calculate, we integrated HTML with JavaScript in our research. For this integration, an interactive interface with four unique scenarios—each intended to mimic a different teleoperation challenge—had to be created. In order to efficiently elicit responses from participants, we included pre-established timestamps in every scenario that indicated the intervals of time during which participants had to act.

Using keyboard inputs to carry out predetermined actions that matched the teleoperation situations, participants interacted with the teleoperation interface during the experiment. The 'W' key, for instance, might be used to travel straight forward, the 'A' key to move left, the 'S' key to stop, and the 'D' key to move right. Participants were able to interact with the scenarios and reply appropriately thanks to these keyboard inputs.

We captured timestamps for each participant's actions within predetermined time periods to ensure precise measurement of response times. In order to assess participant reactions precisely, these intervals, which stand for time frames, were carefully chosen to capture crucial periods during the teleoperation activities.

The recorded intervals, along with their corresponding time duration, were as follows: - Response Time S1(17-24 seconds),S2(36-42.5 seconds),S3(74-86 seconds),S4(119-130 seconds). where "S" represents scenarios.

Every interval functioned as a benchmark for recording the participants' reaction times to particular tasks in the teleoperation scenarios. Our objective was to gather comprehensive insights into the performance and behavior of participants in teleoperation situations by gathering data on their response times inside these intervals and linking them to the appropriate actions.

Every interval functioned as a benchmark for recording the participants' reaction times to particular tasks in the teleoperation scenarios. Our objective was to gather comprehensive insights into the performance and behavior of participants in teleoperation situations by gathering data on their response times inside these intervals and linking them to the appropriate actions.

We saved the recorded timestamps and related response time data in JSON format in the browser's local storage to make data gathering and analysis easier. Throughout the trial, this method made it possible for us to effectively store and retrieve data, which made it possible to track participant answers and analyze teleoperation performance later on.

Overall, the use of keyboard inputs and predefined timestamps in conjunction with HTML and JavaScript integration made it easier to measure and record response times accurately for teleoperation tasks, giving researchers important information about participant behavior and performance. As shown in figure 7.

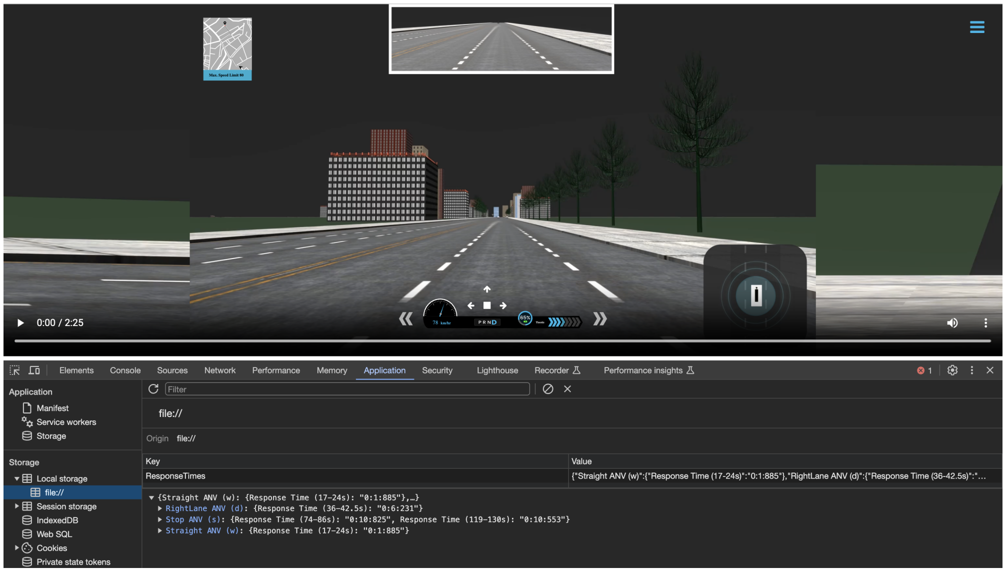

Fig. 7. Response time Calculation

5 STUDY DESIGN

This study aims to systematically compare and characterize how different feedback modalities (audio, audio-to-visual conversion, audio and visual) influence teleoperator understanding of the emergency situation based on response time, error rate, and situational awareness metrics.

5.1 Participants Criteria

For the participant selection process in our study, we aimed to ensure a balanced representation across gender demographics. Participants were required to fall within the age range of 22 to 38 years and demonstrate familiarity with the usage of standard input devices such as keyboards, mice, etc.

To maintain the integrity of our sample, exclusion criteria were applied to filter out individuals who did not meet specific qualifications. Notably, participants lacking a valid driver's license and/or possessing less than 1.5 years of driving experience were excluded from the study. This criterion was implemented to ensure that participants had sufficient practical exposure to vehicular operations.

The demographic composition of participants involved in the study reflects an intentional effort to promote diversity while ensuring adequate representation across gender identities/ Among the participants, a majority of 13 were male, while two were female, indicating a slight gender imbalance within the sample pool.

5.2 Scenarios and Tasks

To assess the efficacy of various sensory feedback modalities in teleoperation, participants engaged in four different driving scenarios crafted to simulate real-world challenges. Each scenario was designed to examine the participant's responses under different sensory input conditions, thereby facilitating a comprehensive evaluation of teleoperation performance.

Audio only: Participants rely solely on auditory cues (e.g., sirens) to identify and respond to hazards.

Audio to visual conversion: Auditory cues will be translated into visual representations (e.g., symbol alerts) on the interface (Li et al. [11]).

Audio and visual: Participants will receive both auditory and visual information simultaneously.

The four scenarios are:

Ambulance Approaching from Behind in Left Lane: Participants encountered an ambulance approaching from behind in the left lane, requiring prompt identification and appropriate response based on the sensory cues provided.

Ambulance Approaching from Behind: An ambulance approached the participant's vehicle from behind at high speed, excluding the additional complexity of lane positioning present in the previous scenario.

Ambulance Crossing in Front: Participants faced an ambulance crossing in front of their vehicle at an increased velocity, necessitating rapid assessment and response to the dynamic situation.

Bicycle Crossing in Front: A bicycle crossed in front of the participant's vehicle at a heightened speed, offering a comparison with non-emergency vehicles to evaluate the effectiveness of sensory feedback modalities in diverse scenarios.

By combining the three sensory feedback conditions with the four driving scenarios, we established a comprehensive framework to assess participant performance and situational awareness across various teleoperation tasks. This systematic approach allowed us to examine the nuanced effects of different sensory feedback modalities on teleoperator behavior and decision-making in dynamic driving environments.

The study design is based on a between-group methodology. These 15 participants will be divided into three groups with 5 participants each, and each participant of the group will be tested with either of the three conditions with all four scenarios.As shown in Table 1

Group Division	**Conditions**	**Scenarios**
Group A (5 participants)	Condition A (Audio)	S1: Ambulance Approaching from behind in left lane
		S2: Ambulance Approaching from behind in Same lane
		S3: Ambulance Crossing
		S4: Cycle Crossing
Group B (5 participants)	Condition B (Audio To Visual)	S1: Ambulance Approaching from behind in left lane
		S2: Ambulance Approaching from behind in Same lane
		S3: Ambulance Crossing
		S4: Cycle Crossing
Group C (5 participants)	Condition C (Audio And Visual)	S1: Ambulance Approaching from behind in left lane
		S2: Ambulance Approaching from behind in Same lane
		S3: Ambulance Crossing
		S4: Cycle Crossing

Table 1. Study Design Table

The study design is based on Between group methodology. These 15 participants will be divided into three groups with 5 participants each, and each participant of the group will be tested with either of the three conditions with all four scenarios

5.3 Procedure

The study commenced with an Introduction and Onboarding phase, where participants were introduced to the prototype and guided through a simple task to acquaint themselves with the interface. Informed consent documents and task instructions were provided to ensure participants' comprehension of the study protocol. Subsequently, participants were divided into three groups based on their assigned feedback modality: audio, audio-to-visual, and audio-and-visual, facilitating controlled experimentation. The familiarization phase was followed by testing, where each group completed three pre-defined scenarios simulating teleoperation tasks. These scenarios required participants to assess and respond to dynamic environmental stimuli within the teleoperation interface, with measurements recorded, including response time, errors, and situational awareness using the SART scale. Post-scenario interviews were then conducted to gather subjective feedback on participants' experiences, capturing their perceptions and challenges encountered. The End Session included open-ended questions and a User Experience Questionnaire to comprehensively evaluate overall user satisfaction, usability, and perceived effectiveness of the teleoperation prototype.

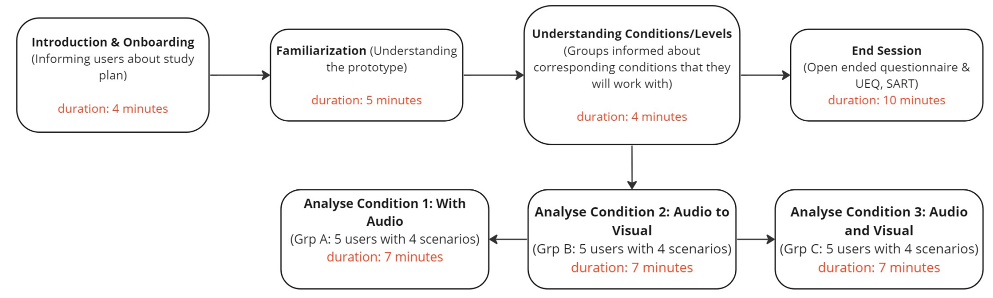

Fig. 8. Study procedure flow

5.4 Test environment

The study was conducted in a controlled laboratory (Ramasundaram et al. [15]) setting on a computer simulation with minimum external interference, triple screens display for separate camera views of all sides of the vehicle, and headphones ensure controlled sound delivery as shown in figure 9

Fig. 9. Test Setup

5.5 Evaluation metrics

Implementing the mixed-method approach, the metrics are divided into Qualitative and Quantitative analysis methods. As shown in table2

Quantitative	Qualitative
Response time (the time it takes for the user to respond to the feedback after it is triggered)	Post-test interview questions
Number of errors (the number of times the user gives an incorrect response to the situation, for example- does not stop for the bicycle to cross in scenario 4, leads to failure of the feedback to make the user aware of the situation)	Probing questions
SART (Situation Awareness Rating Technique) Scale	Prototype-related specific questions
User Experience Questionnaire	Questions related to the overall experience

Table 2. Feedback Evaluation Metrics

Reasoning:

Collection of Response Times: To analyze what kind of multi-sensory feedback system is more appropriate and quick for the teleoperator to perform an action.

Identifying Number of Errors: To analyze what the users are not able to grasp in the prototype and eventually what prompted them to make an incorrect decision.

6 RESULTS

6.1 Quantitative Analysis

In this section, we present the quantitative analysis of the data collected during the study. Our analysis aims to provide statistical insights into the performance metrics and user responses across different experimental conditions.

6.1.1 Performance Metrics. We examine the performance metrics, including response times and number of errors, recorded during the teleoperation tasks. Statistical analyses, such as Analysis of Variance (ANOVA) and post-hoc tests (Mrkvicka et al. [12]), were conducted to assess the significance of differences among the experimental conditions. Results indicate no significant differences in response times (F=0.27, p=0.76) across the three conditions.As Shown in Table3

ANOVA: Single Factor						
Summary						
Groups	**Count**	**Sum**	**Average**	**Variance**		
Audio	20	35.634	1.7817	5.216751		
Audio & Visual	20	48.66	2.433	8.77461		
Audio to Visual	20	42.637	2.13185	9.503026		
ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	4.24992	2	2.12496	0.271336	0.763339	3.158843
Within Groups	446.3934	57	7.831462			
Total	450.6433	59				

Table 3. ANOVA: Single Factor

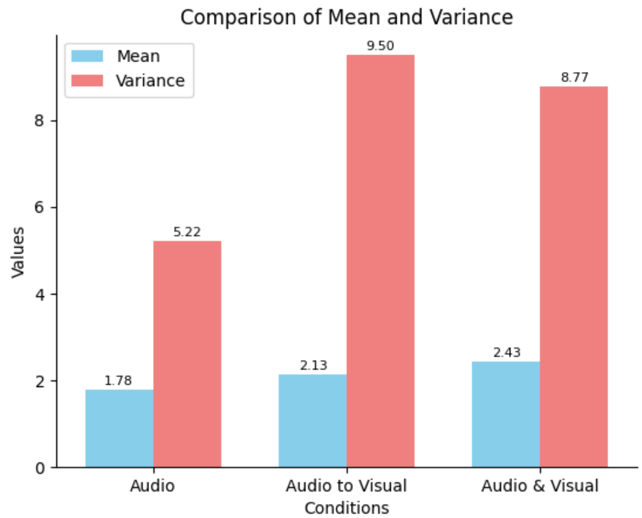

Fig. 10. Mean Comparison

In the conducted experiments, participants exhibited varying response times across different experimental conditions. Notably, participants demonstrated faster response times in both the 'Audio' and 'Audio to Visual' conditions compared to the 'Audio and Visual' conditions. This observation suggests that the inclusion of auditory cues, either alone or in conjunction with visual feedback, facilitated quicker responses from participants in completing teleoperation tasks.

The Tukey's HSD test was conducted at an alpha level of 0.05 (Nanda et al. [13]), indicating a willingness to accept a 5 percent chance of making a false positive error. The test compared the means of three groups denoted as T1, T2, and T3.As shown in Table4

Pairwise Comparisons	**HSD (.05 = 2.1296, .01 = 2.6851)**	**Q (.05 = 3.4032, .01 = 4.2910)**
T1:T2 M1 = 1.78, M2 = 2.43	0.65	Q = 1.04 (p = .74324)
T1:T3 M1 = 1.78, M3 = 2.13	0.35	Q = 0.56 (p = .91743)
T2:T3 M1 = 2.43, M3 = 2.13	0.30	Q = 0.48 (p = .93825)

Table 4. Tukey's HSD Post-Hoc Test

Results from the Tukey's HSD test revealed no statistically significant difference between any of the pairs of groups. The p-value for the comparison between groups T1 and T2 was found to be 0.74, which exceeds the alpha level of 0.05. This indicates that we cannot reject the null hypothesis that the means of these two groups are equal. Similar results were observed for all other pairwise comparisons.

The post hoc Tukey's HSD test was employed to compare the means of three groups following a significant outcome in the one-way ANOVA analysis (Nanda et al. [13]). However, the test did not identify any statistically significant differences between any pair of groups. These findings suggest that despite observing significant differences among the overall group means, the specific pairwise comparisons did not yield statistically significant disparities.

6.1.2 User Satisfaction. Participants were asked to rate their experience using the teleoperation prototype on a Likert scale (Nemoto and Beglar [14]) ranging from 1 (Strongly Disagree) to 7 (Strongly Agree).

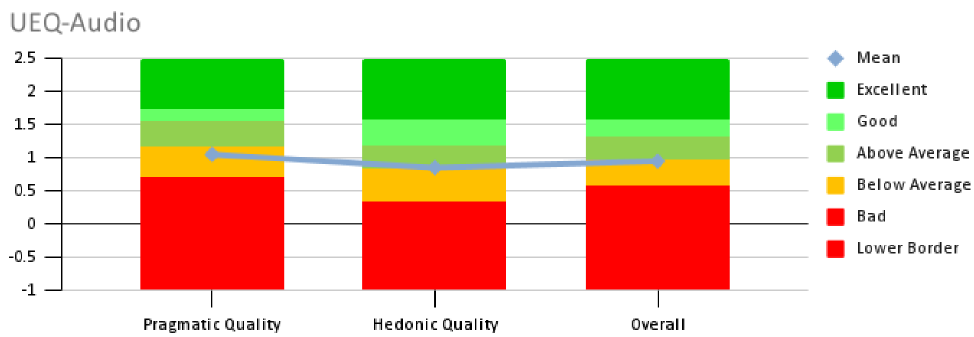

Fig. 11. Audio

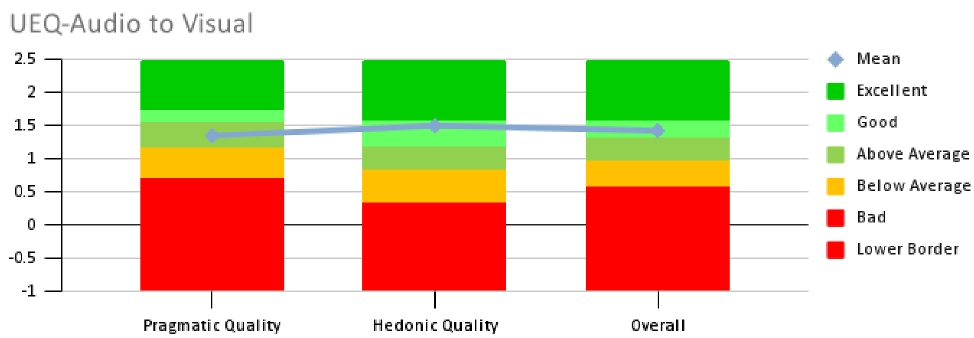

Fig. 12. Audio To Visual

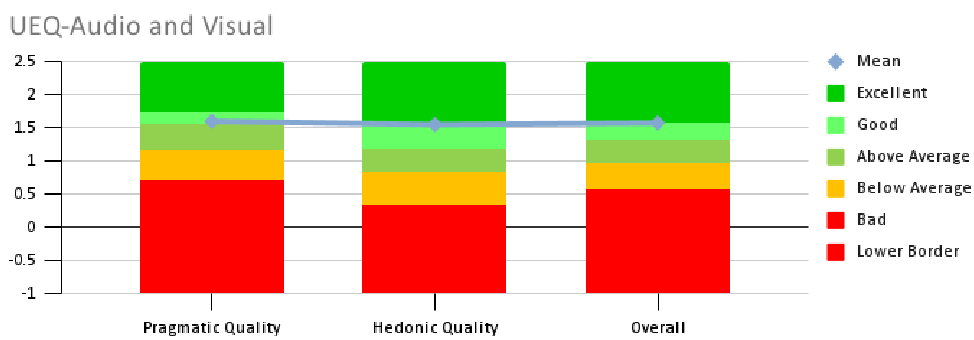

Fig. 13. Audio And Visual

Across the three conditions, Hedonic Quality scores consistently rank highest, followed by Overall and Pragmatic Quality scores. A noteworthy distinction emerges in the distribution of data points among the graphs, with the first condition exhibiting the highest data points overall. This suggests a prevailing perception that both audio and audio and visual qualities were superior to the audio and visual combination. These insights underscore the nuanced interplay between sensory modalities and their impact on user perceptions and preferences in teleoperation interfaces.

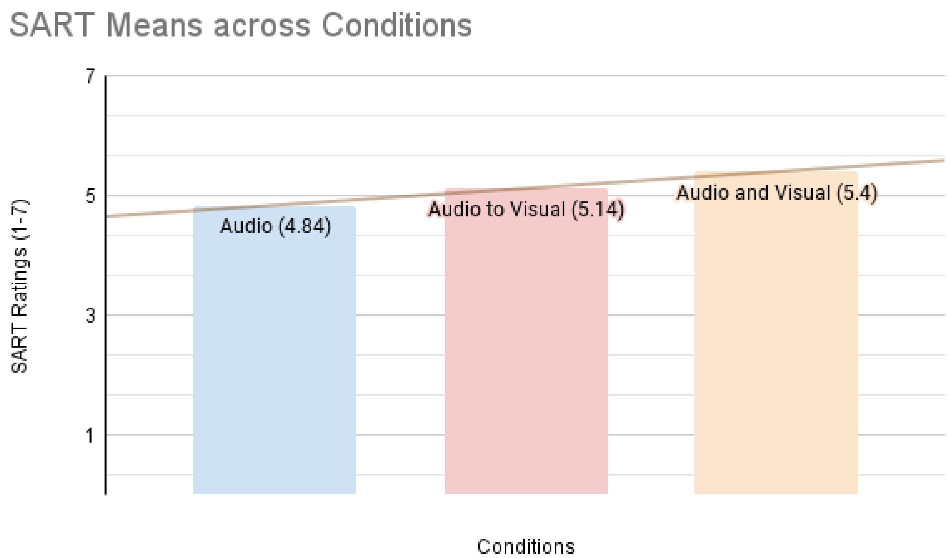

Fig. 14. Comparison of means between 3 conditions

Additionally, participants' situational awareness during the tasks was assessed using the Situation Awareness Rating Technique (SART) scale (Endsley et al. [6]). The SART scale comprises items measuring the participants' perception of their awareness of the operational environment, including their ability to detect and respond to critical stimuli.

The graph shows that the average score for the Audio and Visual condition was the highest (5.4) compared to the other two conditions. This suggests that people had the highest level of situational awareness when they were exposed to both audio and visual cues, as compared to the other two conditions, where they were exposed to only either of the conditions.

6.2 Qualitative data analysis

The post-scenario interview phase involved a combination of open-ended and closed-ended questions (Baburajan et al. [1]) aimed at understanding qualitative insights from participants regarding the impact of audio and visual feedback in teleoperation. Each condition was specifically addressed to discern the influence of auditory and visual cues on user perception. The participants provided nuanced feedback regarding the effectiveness of different multisensory feedback modalities: Participants remarked that auditory cues alone were sufficient for situational awareness, with one participant noting, "Audio is enough to detect the situation of the surroundings." Another participant highlighted the utility of audio in perceiving the direction of approaching vehicles. They also reported challenges when both audio and visual feedback were simultaneously presented. Some participants indicated occasional diversion of attention towards the driving environment, potentially overlooking feedback cues. Additionally, participants expressed concern over information overload due to the concurrent presentation of audio and visual stimuli. In addition to feedback specific to each condition, participants offered general observations regarding teleoperation tasks. One participant highlights the need for consideration of additional environmental factors, such as pedestrians. Onboarding processes were deemed beneficial by one participant, while another emphasized the importance of extensive training to comprehend visual cues effectively. Moreover, participants recognized the influence of scenario complexity on feedback modality preferences and cautioned against potential misinterpretation of auditory cues.

7 DISCUSSION

The central finding of this study is that teleoperators using only audio information made faster decisions compared to those receiving both audio and visual cues. This result underscores the critical role of real-time audio in facilitating rapid and effective decision-making in dynamic teleoperation scenarios.

Several potential explanations can be offered for this observation:

Focus and attention: Processing visual information alongside audio can be cognitively demanding, potentially dividing attention and slowing down decision-making. Audio alone might allow for more focused processing of critical task-relevant auditory cues, leading to faster responses.
Information overload: In dynamic situations, the visual environment can be cluttered and overwhelming, making it difficult to extract (Zhang [21]).
Cognitive processing speed: Auditory information is processed faster by the brain than visual information. This inherent advantage of audio might translate to quicker decision-making, especially in time-sensitive tasks.
Modality-specific expertise: Teleoperators might develop heightened expertise in interpreting and responding to auditory cues through experience. This expertise could enable them to leverage audio-only information more effectively for rapid decision-making (Zhao et al. [22]).
This study underscores the significant role of real-time audio information in expediting decision-making during the teleoperation of dynamic scenarios. The findings imply that audio-only interfaces could offer a viable and potentially advantageous option for specific teleoperation tasks.

8 CONCLUSION

Our study aimed to understand how different feedback conditions affect teleoperator performance and situational awareness in teleoperation tasks. We conducted experiments and gathered subjective feedback to analyze the effectiveness of audio, visual, and audio-visual cues. Contrary to our expectations, we found minimal differences in response times across conditions. Surprisingly, audio alone led to quicker decision-making compared to audio-visual feedback. This highlights the importance of real-time audio information for assessing and responding to dynamic scenarios. Subjective feedback provided valuable insights into teleoperator situational awareness. Participants' experiences shed light on the effectiveness of different feedback modalities.

In conclusion, our research contributes to understanding feedback systems in teleoperation. Further research is needed to refine audio and visual cues integration for optimal performance. Future studies could explore alternative feedback methods to enhance teleoperation efficiency and safety.

8.1 Limitations

We are aware of the limitations of our experiment. Firstly, the recorded simulation prototype does not accurately capture the dynamic nature of real-world scenarios. Since the participants do not actually control the vehicle, there may be a difference in the level of decision-making by the participants. The controlled environment might not represent the challenges experienced by operators. The degree of unpredictability, user preferences, and perceptions might differ in a real driving scene. In addition, our study involved only a limited population, mainly young college students and staff in Europe. Other users might behave and perform differently. We would also like to point out that participants' preferences and feedback might vary over time as new technologies and people's attitudes and acceptance towards them are introduced. Therefore, we suggest that future studies explore our proposed approach with individuals of different ages, education, technophilia, or cultural backgrounds.

8.2 Future Work

In future studies, based on the results of this research, we intend to incorporate and evaluate the addition of real-time video simulations instead of altered prototypes. With the help of this modification, the users will have more control over the bus throughout the simulation. This will be helpful in providing a realistic bus-driving experience.

Moreover, the participant experience can be improved by integrating multi-modal feedback mechanisms (de Barros and Lindeman [4]). This means integrating haptic input with visual and auditory cues to simulate a more engaging experience. We also plan to implement Eye tracking to assess the visual gaze of participants to identify focal points or areas of focus in the prototype. This can help us identify some important factors, like the aspects that lead to decision-making while driving.

REFERENCES

[1] Baburajan, V., e Silva, J. d. A., and Pereira, F. C. (2021). Open-ended versus closed-ended responses: A comparison study using topic modeling and factor analysis. *IEEE Transactions on Intelligent Transportation Systems*, 22(4):2123–2132.

[2] Chamorro, S., Collier, J., and Grondin, F. (2021). Neural network based lidar gesture recognition for realtime robot teleoperation. In *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 98–103.

[3] Chen, J. Y. C., Haas, E. C., and Barnes, M. J. (2007). Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1231–1245.

[4] de Barros, P. G. and Lindeman, R. W. (2013). Performance effects of multi-sensory displays in virtual teleoperation environments. In *Proceedings of the 1st Symposium on Spatial User Interaction*, SUI '13, page 41–48, New York, NY, USA. Association for Computing Machinery.

[5] Drechsler, M. F., Peintner, J., Reway, F., Seifert, G., Riener, A., and Huber, W. (2022). Mire, a mixed reality environment for testing of automated driving functions. *IEEE Transactions on Vehicular Technology*, 71(4):3443–3456.

[6] Endsley, M. R., Selcon, S. J., Hardiman, T. D., and Croft, D. G. (1998). A comparative analysis of sagat and sart for evaluations of situation awareness. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 42, pages 82–86. SAGE Publications Sage CA: Los Angeles, CA.

[7] Flohr, L. A., Schuß, M., Wallach, D. P., Krüger, A., and Riener, A. (2024). Designing for passengers' information needs on fellow travelers: A comparison of day and night rides in shared automated vehicles. *Applied Ergonomics*, 116:104198.

[8] Gallagher, K. M., Cameron, L., De Carvalho, D., and Boule, M. (2021). Does using multiple computer monitors for office tasks affect user experience? a systematic review. *Human Factors*, 63(3):433–449.

[9] Jansen, B. J. (1998). The graphical user interface. *ACM SIGCHI Bulletin*, 30(2):22–26.

[10] Kettwich, C., Schrank, A., and Oehl, M. (2021). Teleoperation of highly automated vehicles in public transport: User-centered design of a human-machine interface for remote-operation and its expert usability evaluation. *Multimodal Technologies and Interaction*, 5(5):26.

[11] Li, H., Sarter, N. B., Sebok, A., and Wickens, C. D. (2012). The design and evaluation of visual and tactile warnings in support of space teleoperation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 56, pages 1331–1335. SAGE Publications Sage CA: Los Angeles, CA.

[12] Mrkvicka, T., Myllymaki, M., Jilek, M., and Hahn, U. (2016). A one-way anova test for functional data with graphical interpretation. *arXiv preprint arXiv:1612.03608*.

[13] Nanda, A., Mohapatra, B. B., Mahapatra, A. P. K., Mahapatra, A. P. K., and Mahapatra, A. P. K. (2021). Multiple comparison test by tukey's honestly significant difference (hsd): Do the confident level control type i error. *International Journal of Statistics and Applied Mathematics*, 6(1):59–65.

[14] Nemoto, T. and Beglar, D. (2014). Likert-scale questionnaires. In *JALT 2013 conference proceedings*, pages 1–8.

[15] Ramasundaram, V., Grunwald, S., Mangeot, A., Comerford, N. B., and Bliss, C. (2005). Development of an environmental virtual field laboratory. *Computers & Education*, 45(1):21–34.

[16] Riener, A., Schlackl, D., Malsam, J., Huber, J., Homm, B., Kaczmar, M., Kleitsch, I., Megos, A., Park, E., Sanverdi, G., et al. (2021). Improving the ux for users of automated shuttle buses in public transport: investigating aspects of exterior communication and interior design. *Multimodal Technologies and Interaction*, 5(10):61.

[17] Sankar, G., Djamasbi, S., Li, Z., Xiao, J., and Buchler, N. (2023). Systematic literature review on the user evaluation of teleoperation interfaces for professional service robots. In *International Conference on Human-Computer Interaction*, pages 66–85. Springer.

[18] Schreiber, S. L. (2021). *Application of the User Centered Design process for concept development and evaluation of teleoperation in automated driving*. PhD thesis, Technische Hochschule Ingolstadt.

[19] Schuß, M., Rollwagen, A., and Riener, A. (2022). Understanding operator influence in automated urban shuttle buses and recommendations for future development. *Multimodal Technologies and Interaction*, 6(12):109.

[20] Triantafyllidis, E., Mcgreavy, C., Gu, J., and Li, Z. (2020). Study of multimodal interfaces and the improvements on teleoperation. *IEEE Access*, 8:78213–78227.

[21] Zhang, T. (2020). Toward automated vehicle teleoperation: Vision, opportunities, and challenges. *IEEE Internet of Things Journal*, 7(12):11347–11354.

[22] Zhao, L., Nybacka, M., Drugge, L., Rothhämel, M., Habibovic, A., and Hvitfeldt, H. (2024). The influence of motion-cueing, sound and vibration feedback on driving behavior and experience – a virtual teleoperation experiment. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–13.

[23] Zhu, D. and Gedeon, T. (2011). Exploring camera viewpoint control models for a multi-tasking setting in teleoperation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 53–62. ACM.

[24] Zou, X., O'Hern, S., Ens, B., Coxon, S., Mater, P., Chow, R., Neylan, M., and Vu, H. L. (2021). On-road virtual reality autonomous vehicle (vrav) simulator: An empirical study on user experience. *Transportation Research Part C: Emerging Technologies*, 126:103090.