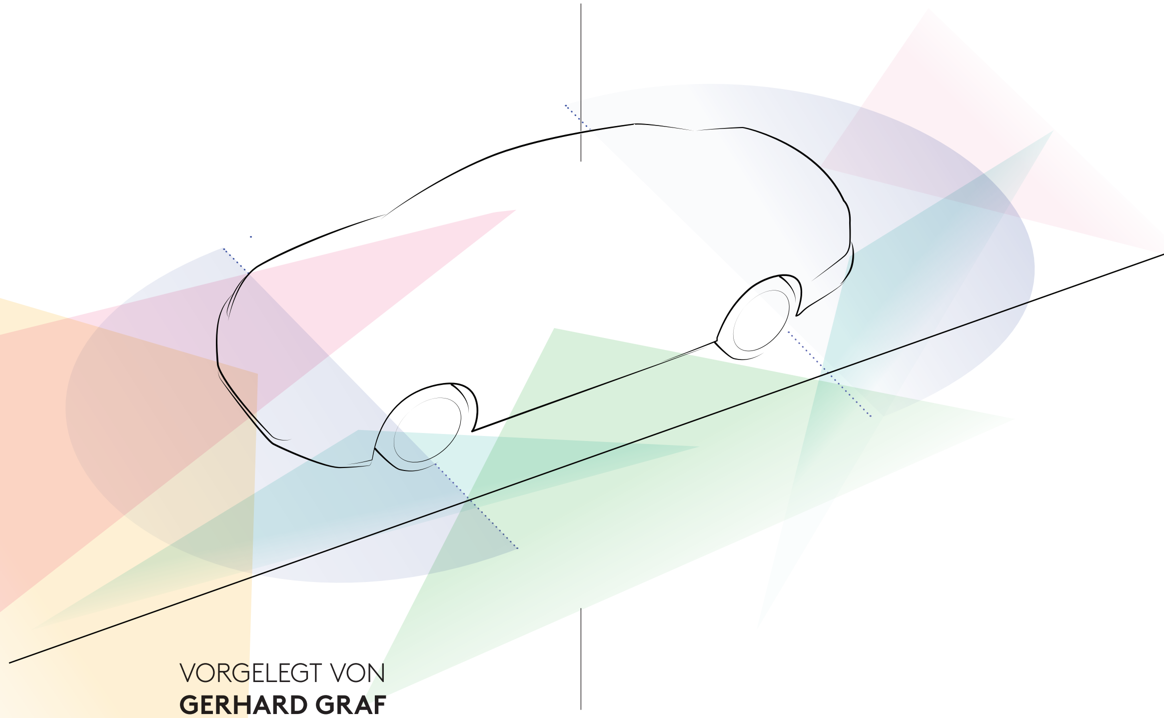
FAR OUT
INTERFACE AND INTERACTION CONCEPTS FOR TELEOPERATED AUTONOMOUS VEHICLES

DISSERTATION
AN DER FAKULTÄT FÜR MATHEMATIK,
INFORMATIK UND STATISTIK
DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN



VORGELEGT VON
GERHARD GRAF
M.SC. HUMAN–COMPUTER INTERACTION

MÜNCHEN, 4. APRIL 2024

FAR OUT
INTERFACE AND INTERACTION CONCEPTS FOR TELEOPERATED AUTONOMOUS VEHICLES

DISSERTATION
AN DER FAKULTÄT FÜR MATHEMATIK,
INFORMATIK UND STATISTIK
DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

VORGELEGT VON
GERHARD GRAF
M.SC. HUMAN–COMPUTER INTERACTION

MÜNCHEN, 4. APRIL 2024

ERSTGUTACHTER: PROF. DR. SVEN MAYER
ZWEITGUTACHTER: PROF. DR. RONALD SCHROETER

TAG DER MÜNDLICHEN PRÜFUNG:
6 DECEMBER 2024

Zusammenfassung

Mit den Fortschritten der Fahrzeugautomatisierungstechnologie und den ersten autonomen Ride-Hailing-Flotten, die den Testbetrieb aufgenommen haben, werden praktische Überlegungen und Anforderungen an das Anbieten von Diensten, die auf der Technologie der Autonomen Fahrzeuge basieren, deutlich. Situationen, in denen die Infrastruktur ausfällt, das Erkennen und Interpretieren von Verkehrskontrollen, z. B. durch Handzeichen der Polizei, oder Situationen, in denen das Autonome Fahrzeug mit Hindernissen zurechtkommen muss, sind einige Beispiele, die den Bedarf an menschlicher Unterstützung erhöhen können. Zu diesem Zweck wurde die Teleoperation als Lösung vorgeschlagen. Zwangsläufig muss sich die Teleoperation einigen Herausforderungen stellen, wie z. B. einer hohen Latenzzeit und einem geringen Situationsbewusstsein aus der Ferne. Um diese Probleme zu lösen, präsentieren wir in dieser Dissertation die Evaluierung einer Reihe von Schnittstellen- und Interaktionskonzepten, die darauf abzielen, das Situationsbewusstsein des Teleoperators zu verbessern. Insbesondere diskutieren wir die Ergebnisse verschiedener Studien zur Verbesserung der räumlichen Wahrnehmung, des Zustandsbewusstseins und der Entscheidungsprozesse des Teleoperators.

Der Prozess der Gestaltung eines sicheren und ergonomischen Teleoperator-Arbeitsplatzes begann mit einer explorativen Phase, in der wir die Anforderungen der Bediener aus erster Hand sammelten. Insbesondere im Hinblick auf die

Forschungsfrage: Welche Informationen werden benötigt, um das Situationsbewusstsein bei der Fernsteuerung eines Autonomen Fahrzeuge zu verbessern. Hiermit wurden die relevantesten Anforderungen in funktionale Designkonzepte übersetzt. Zunächst wurden zwei Human-Machine-Interface-Konzepte für die Übernahmeaufforderung entworfen und evaluiert. Beide sollten die Übernahme der Kontrolle über das Fahrzeug erleichtern. Weiterhin entwickelten wir eine Software, die es dem Teleoperator ermöglichte, schnell zu erkennen, was in der entfernten Situation die Übernahmeanforderung auslösen könnte oder sogar ausgelöst hatte. Um einen schnellen Überblick über die entfernte Umgebung zu erhalten, entwickelten wir ein Interaktionskonzept für die Kamerasteuerung. Basierend auf der Kopfposition und dem Blickbereich des Bedieners konnte die Software vorhersagen, welche Kamera angezeigt werden muss. Schließlich haben wir ein fortschrittliches Teleoperator-Assistenzsystem namens Predictive Corridor entwickelt. Der Predictive Corridor bietet betriebliche Entscheidungsunterstützung, indem er dem Teleoperator die voraussichtliche Position des Autonomen Fahrzeuges und den Bereich anzeigt, in dem das Autonome Fahrzeug bei unerwarteten Netzausfällen weiterfahren wird.

Abstract

As vehicle automation technology progresses and the first autonomous ride-hailing fleets have started test operations, practical considerations and requirements of offering services based on Autonomous Vehicle technology become apparent. Situations of infrastructure failure, recognition, and interpretation of traffic controls, e.g., by police hand signals, or situations when the Autonomous Vehicle has to deal with obstacles, are a few examples that can raise the need for human assistance. Toward this end, teleoperation has been proposed as a solution. Inevitably, teleoperation must face several challenges, such as high latency and low remote Situation Awareness. To solve these issues, this dissertation evaluates a series of interface and interaction concepts that seek to enhance the teleoperator's Situation Awareness. In particular, we discuss the results of different studies designed to improve the teleoperator's spatial perception, state awareness, and teleoperator's decision-making processes.

The process of designing a safe and ergonomic teleoperator workplace started with an explorative phase, in which we collected first-hand operator requirements. In particular, concerning the research question: Which information is needed to enhance Situation Awareness when remote controlling an Autonomous Vehicle. Herewith, the most relevant requirements have been translated into functional design concepts. We began by designing and evaluating two Human-Machine Interface concepts for the Takeover Request. Both

were intended to facilitate the establishment of control over the vehicle. We further developed software that enabled the teleoperator to quickly acknowledge what in the remote situation may cause, or even what had caused the Takeover Request. We developed an interaction concept for camera control to overview the remote environment quickly. Based on the operator's head position and gaze region of interest, the software could predict which camera has to be displayed. Lastly, we designed an Advanced Teleoperator Assistance System called Predictive Corridor. The Predictive Corridor offers operational decision support by showing the teleoperator the Autonomous Vehicle predicted position and the area in which the Autonomous Vehicle will continue to travel in cases of unexpected network losses.

Acknowledgment

To begin with, I want to express my sincere thank to Prof. Dr. Heinrich Hussmann. Prof. Hussmann was essential to my academic journey, and his guidance and support will not be forgotten. Losing Prof. Hussmann was a difficult and deeply saddening experience. Next, I would love to thank Prof. Dr. Sven Mayer, Prof. Dr. Ronald Schroeter, and Dr. Henri Palleis for their time and constructive comments. Likewise, I want to thank all my colleagues, Ph.D. fellows, and of course, all the students who made this thesis possible. Thanks to Noyan Tillman Sahin, Hao Xu, Zhenxin Zhang, Zhidong Zhang, and Aleksa Ristic. Lastly, I would like to express my most obsequious thanks to my family, who have supported my efforts seamlessly throughout all the phases of my journey. To you, a heartfelt hug.

Table of Contents

1 Introduction **13**

1.1 Research Approach 18

1.2 Contribution and Collaboration Statement 19

1.3 Thesis Outline 22

BACKGROUND AND RELATED WORK **25**

2 Teleoperation **27**

2.1 Conventional Teleoperation System Models 29

2.2 Principal factors that affect remote operations 32

2.2.1 Video Image 32

2.2.2 Field of View 34

2.2.3 Camera Viewpoint 35

2.2.4 Orientation 39

2.2.5 Depth Perception 40

2.2.6 Time Delay 41

2.3 Teleoperation Human–Machine Interfaces Concepts 42

2.3.1 Interfaces for Teleoperation Inputs 42

2.3.2 Interfaces for Teleoperation Feedback 53

2.4 Discussion 56

2.5 Summary 61

3 **Sensing Operator's Situation Awareness** **63**
3.1 Individual Situation Awareness 63
3.2 Team Situation Awareness 67
3.3 System Situation Awareness 69
3.4 Measuring Situation Awareness 71
3.5 Discussion 73
3.6 Summary 75

4 **Sensing Operator's Cognitive Workload** **77**
4.1 Cognitive Workload 77
4.1.1 Measuring Cognitive Workload 78
4.2 Methodology 79
4.2.1 Systematic Literature Review 80
4.3 Findings of the Systematic Literature Review 81
4.3.1 Thermal Imaging Application 81
4.3.2 Thermal Imaging Data Capture 84
4.3.3 Thermal Imaging Data Processing 86
4.4 Results of the Meta-Analysis 88
4.5 Discussion 91
4.6 Summary 93

TOWARD DESIGING AN EFFECTIVE TELEOPERATOR WORKPLACE **94**

5 **Subject Matter Expert Analysis** **97**
5.1 Related Work 98
5.2 Requirement Assessment 101
5.2.1 Procedure and Tasks 102
5.2.2 Participants 102
5.2.3 Apparatus and Materials 103
5.2.4 Results 103
5.3 Requirements Clustering and Rating 108
5.3.1 Procedure and Tasks 108

5.3.2 Participants 110
5.3.3 Apparatus and Materials 111
5.3.4 Results 111
5.4 Discussion 113
5.5 Summary 114

ON-BOARDING: REMOTE TAKEOVER REQUEST 116

6 Remote Takeover 119
6.1 Related Work 119
6.2 Our Approach 121
6.2.1 Primary Display Design 123
6.2.2 Secondary Display Design 124
6.3 User Study 125
6.3.1 Participants 125
6.3.2 Apparatus and Materials 126
6.3.3 Study Design 126
6.3.4 Procedure 128
6.3.5 Results 129
6.3.6 Discussion 132
6.4 Summary 134

7 Detecting and Visualizing Anomalies 135
7.1 Related Work 135
7.2 Our Approach 138
7.3 User Study 140
7.3.1 Implementation Set-Up 140
7.3.2 Participants 142
7.3.3 Apparatus and Materials 142
7.3.4 Study Design 143
7.3.5 Procedure 144
7.3.6 Result 144
7.4 Discussion 146
7.5 Summary 147

EN-ROUTE: TELEOPERATOR ASSISTANCE SYSTEMS 148

8 The Predictive Corridor 151
8.1 Related Work 152
8.2 Our Approach 154
8.3 User Study 156
8.3.1 Participants 156
8.3.2 Apparatus and Materials 157
8.3.3 Study Design 158
8.3.4 Procedure 160
8.3.5 Results 161
8.4 Discussion 167
8.5 Summary 168

9 Dynamic Camera Control 171
9.1 Related Work 171
9.2 Our Approach 175
9.3 User Study 177
9.3.1 Implementation Set-Up 177
9.3.2 Participants 178
9.3.3 Apparatus and Materials 179
9.3.4 Study Design 180
9.3.5 Procedure 182
9.3.6 Results 183
9.4 Discussion 188
9.5 Summary 191

REFLECTIONS 192

10 Conclusion and Outlook 195
10.1 Summary 195
10.1.1 On-Boarding: Takeover Request 198
10.1.2 En-Route: Teleoperator Assistance Systems 199
10.2 Contributions 202

10.3 Limitations 203
10.4 Future Work 204
10.5 Closing Remarks 207

Bibliography **209**

Glossary **240**

List of Figures **247**

List of Tables **261**

1

Introduction

The Autonomous Vehicle (AV) holds the promise of improving our live road safety and offering new mobility services. Several new technology companies, traditional automakers, and original equipment suppliers carry out extensive technological research to bring their products to the market [4]. Hereby, AVs will transport passengers and materials without the driver's presence inside the vehicle [132]. The On-Road Automated Driving (ORAD) committee provides a classification of six levels of driving automation, ranging from no automation to full automation [37]. The guide, formally known as *"Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems"*, is used by engineers, legislators, and journalists to cluster AVs in their level of automation. According to the ORAD committee, the critical distinction lies between levels two and three. In level two, the human driver performs part of the dynamic driving task, i.e., execution of steering and acceleration/deceleration. In contrast, in level three, the automated driving system performs the entire dynamic driving task, i.e., able to monitor the driving environment.

Current vehicles already offer many partially autonomous features, such as lane change warning, adaptive cruise control, blind-spot detection, side collision warning, or emergency braking. However, to develop high or full AVs, many

Level	**Steering, acceleration, and deceleration**	**Monitoring of driving environment**	**Fallback performance of dynamic driving task**
Zero	Human	Human	Human
One	Human	Human	Human
Two	**System**	Human	Human
Three	**System**	**System**	Human
Four	**System**	**System**	**System**
Five	**System**	**System**	**System**

Table 1.1: The table provides an overview of the six levels of driving automation [37].

technology companies are developing methods to array the numerous sensors with software to allow the vehicle to perceive the world in 360 degrees (Figure 1.1). Here, sensors such as Light Detecting and Ranging (LiDAR), Long- Mid- and Short-Range Radars as well as cameras, constantly scan the environment to detect any dynamic and static objects around the vehicle, e.g., pedestrians, cyclists, vehicles, traffic lights, obstructions, and other road features. The sensors work together seamlessly to create a detailed 3D image of the world, and to continuously translate dynamic driving tasks into commands, i.e., the operational commands such as steering, braking, accelerating, and monitoring the vehicle on the roadway, and tactical ones such as responding to events, determining when to change lanes, turn, or use signals. For each dynamic artifact on the road, the software predicts the future movements and behaviors based, among others, on the current speed and trajectory of the vehicle. Once the information is received, the software will calculate the exact trajectory, speed, lane, and steering maneuvers needed to progress along its route safely. However, despite the maturity of the industry and vehicle technology, it is still impracticable to automate all complex driving situations arbitrarily. In this regard, teleoperation, which denotes the operation of a vehicle at a distance, was proposed as a solution. Situations of infrastructure failure, recognition, and interpretation of traffic controls, e.g., police hand signals, or situations when the AV has to deal with obstacles, are a few examples that can raise the need for human assistance (Figure 1.1).

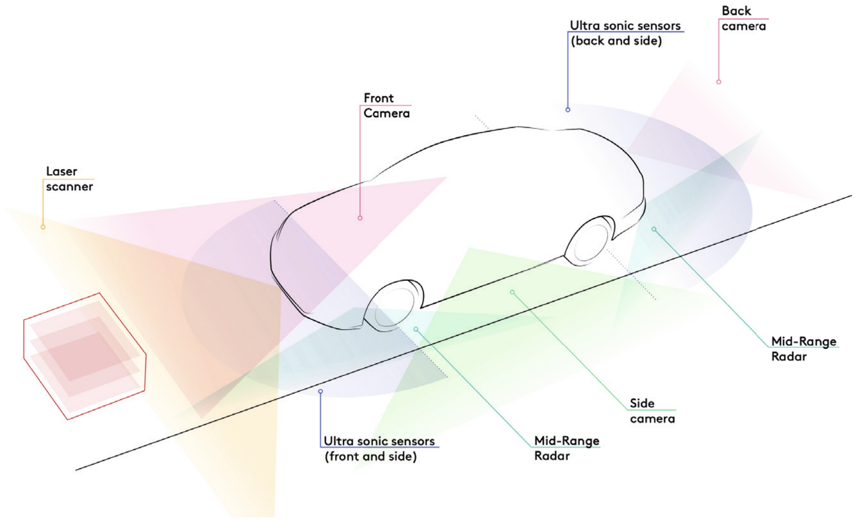


Figure 1.1: This illustration shows the environmental sensor of a On-Road Automated Driving (ORAD) Level 3 vehicle. Moreover, we see a possible teleoperation scenario, i.e., an unexpected non-moving obstacle lying on the vehicle path. In such a scenario, according to traffic regulations, the Autonomous Vehicle (AV) is not allowed to drive over the solid marking lane. However, the AV may wirelessly send a Takeover Request to the teleoperator to solve the issue.

Inevitably, teleoperation must face several challenges, such as limited field of view, management of viewpoints cameras, orientation in a foreign environment, low depth perception, or the challenge of time latency.

Over the course of research in recent years, solutions and systems have been proposed in the past years of research years to support the operator in the analysis and execution of maneuvers, for example, by implementing forces that can counterbalance the operator's movements on the controllers, adopting multimodal feedback, or advanced visualization to predict the remote vehicle movements. Multimodal feedback has also been employed in a variety of different teleoperation systems. These included visual and haptic feedback, visual and audio feedback, or even a combination. Multimodal feedback has been proven to provide better immersion, orientation, and communication of alerts.

Chapter	Research Problem [169]	Short description	Research Question
Chapter 2	Conceptual	Systematic literature review on principal factors that affect remote operations and teleoperation interfaces concepts.	What are the potential benefits and challenges of using teleoperation to improve the safety and functionality of autonomous vehicles?
			What are the principal factors that affect the performance of teleoperation, and how can they be addressed to improve the efficacy of teleoperation?
			What are the state-of-the-art human-machine interface concepts for vehicle teleoperation, and how do they affect the performance and success of teleoperation tasks?
Chapter 3	Conceptual	Systematic literature review on individual and distributed Situation Awareness.	What are the methods for improving Situational Awareness in remote operation contexts, and how do these methods affect performance and task success?
Chapter 4	Conceptual, Empirical	Systematic literature review and meta-analysis on thermal imaging in driving simulators.	What is the effectiveness of using infrared thermal imaging to measure cognitive load in driving situations, and how does it compare to traditional physiological sensors?

Chapter 5	Constructive, Empirical	Subject Matter Expert analysis to support the discovery of users’ requirements and the development of novel interfaces.	What are the necessary user requirements for vehicle teleoperation-based interfaces that support high levels of Situation Awareness, and how should they be presented to the operators?
Chapter 6	Constructive, Empirical	An empirical study on novel interfaces for remote Takeover Requests.	What is the effect of implementing Takeover Request interfaces in improving the performance of remote vehicle control, and how does it affect the Situation Awareness and cognitive load in teleoperated vehicle operations?
Chapter 7	Constructive, Empirical	An empirical study on a novel interface for vehicles’ disengagement.	How can the concept of introspection be applied to Autonomous Vehicles, and what is its effect on improving Situation Awareness and reducing Takeover Times in teleoperated vehicle operations?
Chapter 8	Constructive, Empirical	An empirical study on a novel interface for vehicle prediction position to mitigate low depth perception and time latency.	How can the accuracy of current Predictive Displays be improved, and what is its effect on driving performance and cognitive load in teleoperated vehicle operations?
Chapter 9	Constructive, Empirical	An empirical study on novel interfaces for camera viewpoint control.	What is the effect of implementing "hands-free" camera control mechanisms on cognitive workload and Situation Awareness in teleoperated vehicle operations?

Table 1.2: The table provides a structured summary of the research questions addressed [169].

Moreover, augmented reality techniques are demonstrated to support operator spatial awareness, including orientation and depth perception. However, despite the examples of existing research, Situation Awareness remains a challenge when developing interface and interaction concepts for teleoperated AV.

In this context, the main objectives of our work are to improve teleoperators' spatial perception, state awareness, and teleoperators' decision-making processes during the Takeover Request and vehicle conduction. We will present a series of experiments that will examine the teleoperator's Situation Awareness by employing advanced technologies and user interface design concepts. The aim is to increase and improve the usability of the teleoperator information systems while complying with safety requirements. Hence the motivation for this work is to create a safe and ergonomic teleoperation workplace. These are requirements that are indispensable for secure interaction with AVs. Thereby, the following overarching research questions will be examined:

What information should be presented to the teleoperator? How should it be presented and what effect will the information have on the teleoperator?

These questions will be evaluated based on the theoretical principles from the literature and examined in several experiments using a qualitative and quantitative methodological approach. Eventually, Table 1.2 outlines the research questions explored in this dissertation, following the framework of Oulasvirta and Hornbæk [169].

1.1 Research Approach

During the course of this thesis, we will present user-interface and interaction concepts to improve teleoperators' spatial perception, state awareness, and decision-making processes. To address our objective, we will set the empirical and conceptual research upon concepts from cognitive science, precisely upon the concept of Situation Awareness and cognitive workload. This theoretical framework will help us set a foundation and a common denominator among the works we evaluate.

To find the answers to our questions, we started with a literature research. The literature research helped us find qualitative requirements, and models for tasks/functions shared between operators and vehicles. Which ultimately helped us present a solution space for AV teleoperation systems. In this early phase, first-hand experts' requirements were included to complement the literature research findings. Thus, we conducted several workshops dedicated to a user-centered design approach.

A further methodology we included in this thesis was an iterative design process that included early and rapid evaluations that built on each other. This was efficiently realizable in the context of my doctoral thesis. In particular, the implementation and evaluation of display and operating concepts for AV teleoperator applications were based on previously identified requirements to derive ultimately desirable Human-Machine Interface (HMI) concepts. This included prototyping from low to high applications, evaluating designs with qualitative and quantitative research methods, and lastly, analyzing the data gathered to endeavor generalizations and conclusions.

1.2 Contribution and Collaboration Statement

This thesis was supported by Bayerische Motoren Werke (BMW). However, the analyses, considerations, and conclusions presented in this document are entirely independent and have not been influenced, sponsored, or endorsed by the organization. Additionally, during the course of this research, I had the privilege of collaborating with students and colleagues whose insights and contributions have been instrumental in shaping this work (cf. Table 1.3). Beyond Chapter 2 and Chapter 3, which were developed entirely by me and cover, respectively: (a) a systematic literature review on the principal factors influencing remote operations and teleoperation interface concepts, and (b) a systematic literature review on individual and distributed Situation Awareness, this thesis further explores key empirical and collaborative research efforts.

In particular, Chapter 4 focuses on a systematic literature review and meta-analysis examining the application of thermal imaging technology in driving simulators.

Personal Contribution	Collaborators’ Contributions	References
I managed and supervised the entire project, providing detailed feedback and strategic direction during the literature review process. Additionally, I was solely responsible for selecting the papers to be included and conducting the meta-analysis.	Noyan Tillman Sahin supported this work by systematically reviewing the literature and contributing to the overall structure of the study.	[89]
I led both of the user studies that informed the empirical Subject Matter Expert analysis and was directly responsible for the comprehensive evaluation of the findings.	Heinrich Hußmann provided constructive feedback and guidance during the process of preparing and submitting the research paper.	[85]
I managed and supervised the project during the conceptualization and design process of the graphical user interface, as well as during the execution of the user study. I was also responsible for analyzing and evaluating the results of the study.	Zhidong Zhang played a significant role by developing the graphical user interface and assisting with the execution of the user study.	[82, 90]
I was responsible for the entire process, which included designing the graphical user interface, executing the user study, and analyzing and interpreting the results.	Zhenxin Zhang developed the functional prototype for the interface, while the concept of predictive disengagement was introduced and elaborated by Christopher Kuhn.	[115]
I was fully responsible for designing the graphical user interface, conducting the user study, and carrying out a thorough evaluation of the results.	Hao Xu supported the project by developing the prototype and conducting the initial open-loop evaluation of the system..	[83, 84, 87]
I undertook the responsibility of designing the graphical user interface, organizing and conducting the user study, and performing the evaluation of the results.	Zhenxin Zhang contributed to this research by developing the prototype necessary for the study.	[88]

Table 1.3: The table provides a structured summary of the main contributions, highlighting the collaborator, as well as the research papers and the patents derived from it.

My contributions to this chapter were extensive, as I managed and supervised the entire project. I provided detailed feedback and strategic direction throughout the literature review process, ensuring a rigorous and structured approach. Additionally, I was solely responsible for selecting the papers to be included in the analysis and conducting the meta-analysis itself [89]. This work was supported by Noyan Tillman Sahin, who contributed by systematically reviewing the literature and helping to structure the overall study.

Chapter 5 delves into an analysis conducted with Subject Matter Experts to uncover user requirements and support the design and development of innovative user interfaces. I played a central role in this research, leading both user studies that informed the empirical analysis and taking full responsibility for the comprehensive evaluation of the findings. Heinrich Hußmann contributed significantly by providing constructive feedback and guidance during the preparation and submission of the research paper [82, 85].

Chapter 6 describes an empirical study investigating innovative user interfaces designed to manage remote Takeover Requests in the context of autonomous vehicles. My role in this project involved managing and supervising the entire process, guiding the conceptualization and design of the graphical user interface, and overseeing the execution of the user study [82, 90]. Furthermore, I conducted a thorough analysis and evaluation of the study results. This research was supported by Zhidong Zhang, who developed the graphical user interface and assisted with conducting the user study.

Chapter 7, I present an empirical study focused on the design and evaluation of a novel user interface for vehicle disengagement in autonomous systems. My responsibilities included designing the graphical user interface, organizing and executing the user study, and analyzing and interpreting the results to draw meaningful conclusions. Collaborative contributions came from Zhenxin Zhang, who developed the functional prototype, and Christopher Kuhn, who introduced and elaborated on the concept of predictive disengagement [113].

Chapter 8 explores an empirical study investigating a novel user interface designed to predict vehicle positions. This interface aims to address challenges related to low depth perception and time latency in teleoperation scenarios. My role in this project encompassed designing the graphical user interface, conducting the user study, and carrying out a comprehensive evaluation of the results. Hao Xu supported the research by developing the prototype and conducting the initial open-loop evaluation, which provided critical insights for further development [83, 84, 87].

Finally, Chapter 9 focuses on an empirical study examining novel user interfaces for controlling camera viewpoints in teleoperation contexts. My contributions included designing the graphical user interface, organizing and conducting the user study, and evaluating the results in detail [88]. The research was supported by Zhenxin Zhang, who developed the prototype required to implement and test the interface.

Other peer-reviewed papers, not included in this thesis but nonetheless instrumental in shaping its development, are the follows: Graf et al. [86], Schitz et al. [194], Schitz et al. [195], and Schitz et al. [196].

1.3 Thesis Outline

Having introduced in Chapter 1 the objectives of the dissertation and problem position, in Chapter 2, we delve into the literature supporting our work. This includes existing teleoperation HMIs and the principal factors affecting remote operation. A visual overview of the interplay of the content of my research can be assessed in Figure 1.2.

Chapter 3 and Chapter 4 report the theoretical framework that helped us set a foundation and a common denominator among the works evaluated later. Foremost, to address our objective, we set the empirical and conceptual research upon concepts from cognitive science, particularly the concept of Situation Awareness and cognitive workload. In Chapter 4, we were also interested in whether the novel technology for cognitive sensing such as the thermal imaging could provide valuable support for the work. Thus we conducted a systematic literature review and meta-analysis to get a clear picture of the effectiveness of this technology.

The process of designing a safe and ergonomic teleoperator workplace started with an explorative phase, in which we collected first-hand user requirements. Inspired by the requirement analysis of related research fields,

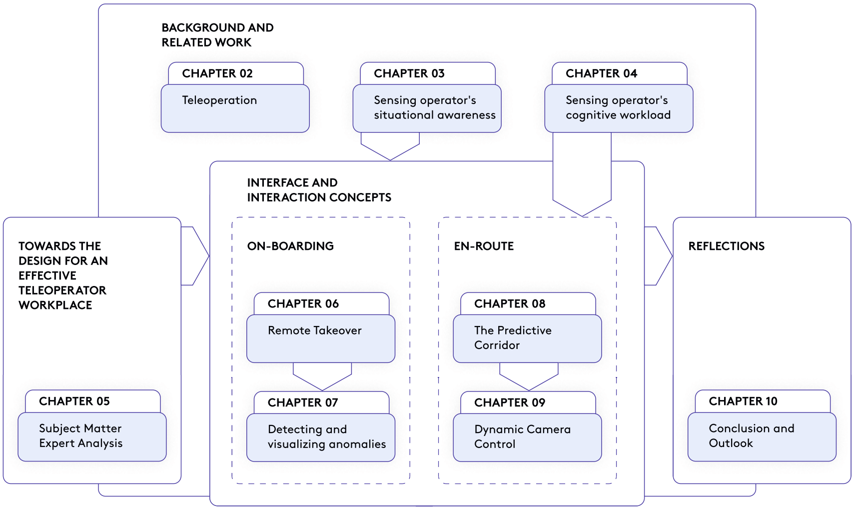


Figure 1.2: The diagram visually shows an overview of the content interplay in my research.

in Chapter 5, we present a comprehensive Situation Awareness requirements framework and analysis for AV teleoperation-based interfaces. The chapter aimed to create a foundation from which we could develop novel design concepts for teleoperated AV.

In contrast to Chapter 6 and Chapter 7, in which we focused on developing HMI that could intuitively explain the Takeover Request, in Chapter 8 and Chapter 9, we addressed, respectively, the challenge of time latency, low depth perception, and the hands-busy problem. Therefore, while Chapter 6 and Chapter 7 refer to HMIs that aim to support the operator in acquiring information before the driving maneuver, in Chapter 8 and Chapter 9, we describe user interactions and interfaces focused on supporting the teleoperator during the driving maneuver. This separation of before and during the driving maneuver inspired us to employ terms like "on-boarding" and "en-route." These clearly refer to the terminology used in the automotive context, in which the customer, after having booked a vehicle via a mobile phone application, is guided to their destination.

When it comes to the methodology employed, the decision to introduce thermal imaging for cognitive load assessment in Chapter 8 and Chapter 9 was driven by a deliberate shift in approach. Following the exploration of subjective methods in chapters Chapter 6 and Chapter 7, specifically the NASA Task Load Index (NASA-TLX) questionnaire, it became apparent that an objective approach was necessary to complement and enhance the assessment framework. Recognizing the inherent subjectivity in tools like NASA-TLX, the introduction of thermal imaging was a strategic move to bring a more objective dimension to the evaluation of cognitive load. This sequential progression allows for a comprehensive examination, starting with established methods and gradually incorporating innovative approaches to offer a well-rounded perspective on cognitive load assessment.

Finally, in Chapter 10, we summarize the work, critically reflect on its limitations, and disclose possible pathways for future research. This agenda spans the insights gained in the presented thesis and can thus help other researchers as guidance and inspiration.

Background and Related Work

2

Teleoperation

As vehicle automation technology progresses and the first autonomous ride-hailing fleets have started test operations, practical considerations and requirements of offering services based on Autonomous Vehicle (AV) technology become apparent. Thus, despite the rapid progress in machine learning, ubiquitous technologies, and communication infrastructure, in some situations, AVs will continue to require human situational assessment [132].

On public roads, it might happen that the traffic regulations do not correspond with the actual traffic situation, e.g., when the traffic requires the regulation of a person. This might occur when road constructions are not fully finished or do not match the standard, e.g., when the lane markings are partially nonexistent, or signposts are missing. Road construction sites may also generate narrow spots on the carriageway and require more attention to avoid collision with other vehicles. Alternatively, some vehicles (e.g., buses, trucks) may need more space at the intersection or other points; in such situations, the AV must act and reset its position to allow the other vehicle to turn. Public shows, events, or sports competitions held in a public place might involve a crowd of people. The AV has to drive slowly through the crowd and attract the attention of vulnerable road users. On other occasions, an obstacle might

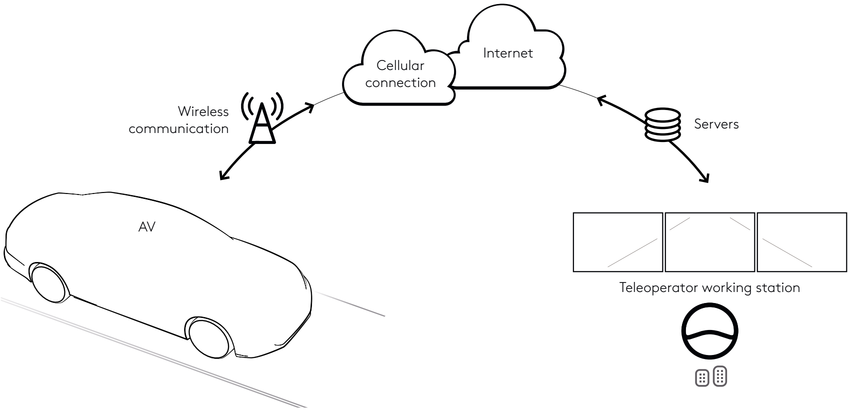


Figure 2.1: The data transmission scheme for teleoperated AVs. The AV transfers radio-based environmental data to the teleoperator, who can remotely assess and manage the vehicle. Operators interpret the data and transfer operational and strategic commands to the vehicle.

block the roadway, and the AV has to drive over the solid marking lane. Other inner-city scenarios might include dealing with emergency vehicles and traffic jams, i. e., entering/exiting to/from the main road. Therefore, despite the AV full functionality, an additional human assessment may improve the quality of the autonomous driving experience.

In this context, vehicle teleoperation will enable *"the capability to sense and act in a remote location"* [68] (p. 13). Having received environmental data from the vehicle, the operator interprets the data and transfers operational and strategic commands to the vehicle (Figure 2.1). Hence, according to Winfield's teleoperation definition, the vehicle can be seen as an extension of the operator's senses, a robot device remotely conducted/supervised by a human being [242].

To control the remote vehicle, the most common and used operating concepts are based on the direct manipulation of the actuators, whereby recently indirect operating concepts have also been proposed in case of dilated time delays and low data transmission capabilities[1] [106]. Indirect operating models such as sequential trajectory designation [75], sequential maneuver selection

---

[1] We refer to Designated Driver, Einride, Ottopia, Phantom, Starsky Robotics, Udelv, and Voysys for direct control. However, companies that implement indirect control are Drive.ai, General Motors, Nissan, Scotty Lab, Waymo, and Zoox (May 2019).

[67], and sequential target coordinates [106], not only might be employed to overcome the needs of high-bandwidth data transmission but also to bypass the operator's cognitive overload. In the next section (Section 2.1), we will describe in greater detail current teleoperation concepts by adopting Sheridan's conventional system model [203].

2.1 Conventional Teleoperation System Models

Sheridan [203] distinguished five degrees of automation to classify teleoperation models, grouped into three primary control paradigms: direct control, collaborative control, and automatic controls (Figure 2.2).

Historically, direct controls are the most widespread method to conduct a vehicle remotely. By using control inputs as a steering wheel, pedal, or joysticks, the operator during the operation is responsible for observing and perceiving through displays, the remote environment and deciding on an appropriate strategy, i.e., in the case of direct driving, the execution of steering and acceleration/deceleration [32, 75]. The telepresence of the operator at this stage is mandatory as only the teleoperator can close the control loop and stabilize the vehicle through the driving maneuver [203].

Although the vehicle, and to some extent the interface, might assist the teleoperator during the task, the operator is still in charge of perceiving the remote environment, making decisions, and executing the maneuvers. Currently, vehicles including Unmanned Aerial Vehicle (UAV), Unmanned Underwater Vehicle (UUV), and Unmanned Ground Vehicle (UGV) continue to rely on direct command, and additionally, several Advanced Teleoperator Assistance System (ATAS) have been proposed to support the operator's decision making. An example of a direct control ATAS for AVs is the Free Corridor (FC) [32]. The FC is a safety concept for direct teleoperation driving in which a path of full braking, based on parameters such as current speed, friction coefficient, and the radius of curvature, is shown to the teleoperator, who is responsible for keeping the path free from obstacles (Figure 8.1a). If there is no connection to the teleoperator, the vehicle will stop at the end of the calculated corridor.

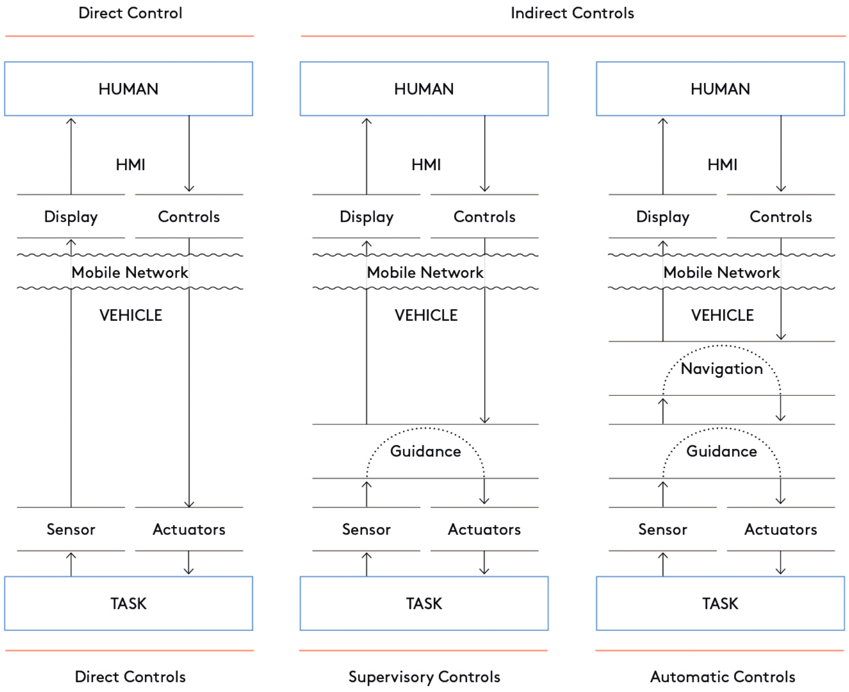


Figure 2.2: An adaptation of the diagram of automation level by Sheridan [203]. Many different authors have adopted Sheridan's [203] model, e.g., [32, 68, 75], though with some adaptation, i.e., instead of having five layers of automation, there were only three, and in all, computer aiding mediation is present. It can be observed that the more abstract the task is, the higher degree of automation is required, i.e., onboard vehicle intelligence.

Another ATAS example is the Predictive Display (PD) [35]. As the name implies, the PD predicts the remote vehicle movements to forecast its position and compensate for the time latency (Figure 8.1b).

Compared to manual controls, supervisory controls already entail some essential autonomous functions, such as path control and obstacle avoidance [106], i.e., a local guidance loop is closed autonomously by the vehicle. Sheridan [202] coined the term *supervisory control*, referring to the analogy between the employer and their subordinates. In this sense, the teleoperator delivers the strategic objective while the execution is left to the vehicle. Although the operator at this stage is relieved from directly guiding the vehicle, as Sheridan [202] noted, the human has yet to continuously find an appropriate navigation strategy for the vehicle, e.g., by designating a sequential trajectory or sequential target coordinates. ATASs examples of indirect controls methods have been developed by Kay [106] and Fong [68], and most recently by Gnatzig [75]. The waypoint-based control developed by Kay [106], for instance, enables the operator to specify a sequence of target coordinates that the vehicle must pass through autonomously. The target coordinates can be based on images [106] or maps [68]. Alternatively, in trajectory-based control [75], the operator designs trajectory segments using the conventional steering wheel-pedal combination. Thus, in supervisory controls, we still see the active presence of the human operator.

Automatic control entails a high level of automation. As Fong [68] pointed out, the automatic control expression is to some extent misleading since such systems will never be fully automatic. There will always be situations where the human operator needs to decide for the vehicle, e.g., when the traffic rule has to be infringed. In automatic control, the control strategy may be delivered before execution or interleaved with the execution [68]. The system is responsible for the vehicle navigation and guidance, while the operator must supervise and provide high-level commands, e.g., when a pull-over is required. Thus, the abstract nature of the task is the essence of the difference between supervisory and automatic control. An ATAS example that can be implemented on such an automation level might be the Conduct-by-Wire concept [105] or pieDrive [70], initially developed for in-vehicle usage. In both concepts, the currently available maneuvers are presented to the driver, who has to decide which maneuver should be passed to the vehicle, e.g., pull-over. The vehicle then has to translate the command into driving functions. Thus, the operator has more responsibility for deciding *what* task should be performed, whereas the vehicle decides *how* to perform the maneuver.

Therefore, conventional teleoperation classification can be distinguished between direct and indirect controls (i.e., supervisory and automatic controls). The stress between direct and indirect controls depends not only on the vehicle on-board intelligence but also on how much responsibility we would like to share with the vehicle. The more abstract commands we give, the more automation/responsibility is delegated to the vehicle. On the one hand, direct controls depend upon the operator's decisions and motor movements [68], which makes the loop more prone to encounter errors [203], and upon the reliance on high-bandwidth data transmission [68]. However, the teleoperator is a valuable aid to the system due to the abundance of driving experiences and the ability to anticipate scenarios [3, 217]. On the other hand, indirect control requires more high-level commands, i.e., the maneuver is initiated by the operator and the vehicle responsible for the execution. Hence, indirect driving releases the operator from the stabilization task while reducing the cognitive effort and error logs [105, 203]. However, according to Kauer et al. [105], motor movement and strategic decision-making skills might be compromised by relieving the operator from controlling the vehicle.

In this regard, it can be observed that vehicle teleoperation is a multidimensional domain that shows the tension between automation limitation and teleoperation requirements. On the one hand, the difficulty in automating all complex driving situations arbitrarily. On the other hand, the need to master high communication delays, managing Situation Awareness, and AV capabilities.

2.2 Principal factors that affect remote operations

From Section 2.2.1 to Section 2.2.6, we will delve into the key factors influencing direct and indirect remote operations (cf. Figure 2.2) and examples of existing approaches. We will discuss the factors that degrade the quality of the video images, low Frame Rate (FR), pixel density, and color depth. Also, we will examine the challenges of teleoperating with a limited Field of View (FoV), managing camera viewpoints, orientation in a foreign environment, the difficulty of depth perception judgment, and the challenge of time latency (cf. Table 2.1).

2.2.1 Video Image

A good perception of the remote environment is essential for a successful teleoperation execution. Researchers have stated that the video channel is the most fundamental component of the complete control loop and that the video image quality will determine the nature of the operator’s performance [227]. Typical factors that can degrade video quality are low FR, pixel density, and color depth [31].

In an empirical study conducted by Darken et al. [43], participants were asked to report the remote camera movement on a sheet of paper at different FR, 1.43 Hz, 6.71 Hz, 17.9 Hz, and 21.37 Hz. The analysis unveiled that participants had trouble maintaining spatial orientation at the lowest FR, i.e., more prone to reporting wrong camera positions and misclassified objects in the remote environment. Many similar studies have evaluated operational performance with various imaging frequencies. In a series of experiments, Van Erp and Padmos [227] recommend a FR of 10 Hz since participants recorded significant lateral control variations at lower frequencies. A further approach to assess FR effect on operative performance was proposed by Massimino and Sheridan [139]. The authors have shown that at 3 Hz, a placement task was doubtful to be executed successfully. By increasing the frequency to 6 Hz, the probability of success increased; however, performances were significantly affected.

Similarly, Chen et al. [29], in a simulated study with UGV and UAV, found that at 5 Hz imaging frequencies, the teleoperator’s performances were diminished. A comprehensive review of low FR effects on human performance is provided by Chen and Thropp [31]. The authors examined more than 50 publications and presented a theoretical framework that compiles the findings into four core dimensions: psychomotor performance, perceptual performance, behavioral effects, and subjective perception. Chen and Thropp [31] inferred that a 15 Hz image frequency might be a suitable threshold for psychomotor and perceptual-based tasks, and 10 Hz might be a tolerable frequency for operative performances.

Other facts that can impair teleoperation quality are pixel density (i.e., resolution quality) and color depth [31]. The low resolution of the video image might exclude essential cues in building an appropriate mental model of the environment [29], and the distortion associated with the picture reconstruction has been suggested to increase the operator’s cognitive load. Nevertheless, in early studies, Sheridan [203] demonstrated that trained subjects could also operate with the degraded signal transmission in resolution scale and color depth.

Other studies have shown that color depth could impact target identification, particularly when matched with low FR [31, 66]. For instance, Fisher et al. [66] reported that subjects who handled the task with color images were 24% faster on target identification time than those given grayscale images.

2.2.2 Field of View

Often teleoperators rely only upon a partial representation of the remote environment, demanding manipulating the remote camera point of view to gain Situation Awareness (SA). This phenomenon is known as the *keyhole effect* in the literature. The *keyhole effect* requires operators additional effort to study the surroundings, as only a tiny representation of the environment is available [246]. Settling themself properly in a remote environment, and remembering their path, has been proven to be complicated even for expert operators. Participants in a study were asked to keep their orientation by following the remote vehicle position on paper and naming objects by placing them in the correct position. The authors observed an "extreme" difficulty and lower performance on spatial comprehension and object classification [43]. Limited Field of View (FoV) has also been shown to reduce the operator's spatial comprehension during the World Trade Center rescue responses [26]. The authors reported that a robot employed in the task force stumbled on its way, and operators could not diagnose the reason for its freezing. In their analysis, Casper and Murphy [26] remarked that the limited FoV and the egocentric perspective played a significant role in detecting the operation's problem.

According to Van Erp and Padmos [227], with a reduced FoV, operators might have trouble judging the vehicle speed, have a degraded perception, and mislocations of obstructions. The authors underlined that peripheral vision is relevant for lane-keeping when driving on curved roads and when planning the turning maneuver, as the driver must be prepared 1-2 seconds before. Enlarging the FoV does not necessarily improve the teleoperator's placement capabilities in the remote scene. In an article on the effects of camera orientation and spatial judgment, Thomas and Wickens [221] showed that participants tended to assess only the information presented on the central screen area and ended the observation for other relevant information in the periphery soon. In particular, when comparing the exocentric frame of reference (with a limited FoV) and the egocentric frame of reference (with a large FoV), participants, in the prior condition, felt a sense of over-salience, causing the state of (display-induced) cognitive tunneling [221]. Moreover, by increasing the FoV, the perceived vehicle velocity rises, and consequently, the operator reduces the speed [208]. Notwithstanding, a broader FoV can be employed before the driving task, when for example, the teleoperator must assess the remote situation, thus planning tactics and strategic responses [199].

2.2.3 Camera Viewpoint

Camera placement in UGV and UAV teleoperation contexts is oftentimes on the vehicle body. Depending on the type of vehicle under control, teleoperators can be provided with ego- or exocentric viewpoints. In particular, cameras mounted on UGV usually present an egocentric view, whereas an exocentric view is more common to be seen in UAVs. Dede [46], in an article, stated that the egocentric viewpoints enable an *"actional immersion and motivation through embodied"* (p. 66). On the contrary, exocentric viewpoints raise *"abstract, symbolic insights gained from distancing oneself from the context"* (p. 66). However, like those provided by the exocentric viewpoint, unnatural perspectives may degrade operative performance due to wrong position estimation [151]. Luck et al. [130] proved that ego perspective produced a faster and safer subject performance in a rescue experiment scenario. However, participants stated the wish to have a more holistic vision of the situation and, therefore, an exocentric viewpoint.

Information integration of a diverse viewpoint, e.g., exocentric and egocentric, had presented the potential for human performance improvements.

Additional information (e.g., sensor data) can also improve teleoperation quality, including better SA [251] and less cognitive load [53]. Hughes and Lewis [97] suggested a dual screens strategy, on the one screen showing sensor-driven data, and on the other screen, the live video streaming of the remote environment. Furthermore, the authors noted that an independent camera had improved the search performances. In contrast, other studies suggest that HMIs with multiple perspectives may fatigue the operator [166, 222].

Factor		Description	Effect on teleoperation quality	Reference
Video Image	Frame rate	The number of frames per second in a video or display.	A low frame rate (< 10 Hz) can make it difficult for operators to maintain spatial orientation and accurately perform tasks in a remote environment.	[29, 31, 43, 139, 227]
	Pixel density	The number of pixels per unit of area in a display.	Low pixel density can exclude essential cues for building an appropriate mental model of the environment and can increase the operator's cognitive load.	[29, 31]
	Color depth	The number of bits is used to represent a single pixel's color in a display.	Color depth can impact target identification, especially along a low frame rate.	[31, 66]
Field of view	Keyhole effect	The limited field of view in a video or display.	By relying only on a partial representation of the remote environment, the teleoperator must manipulate the remote camera point of view to gain SA.	[246]
	Spatial comprehension	The ability to understand the spatial relationships between objects in an environment.	A limited FoV reduces the operator's ability to comprehend the spatial layout of the environment.	[26, 43]
	Object classification	The ability to identify and classify objects in an environment.	A limited FoV reduces the operator's ability to identify and differentiate objects within their surroundings.	[43, 227]
	Perceived vehicle speed	The speed at which the operator perceives a vehicle.	Operators with a limited FoV may need help to accurately judge a vehicle's speed. Nevertheless, an enlarged FoV might cause the operator to perceive that the vehicle is moving faster, leading to a reduced speed.	[208, 227]

Camera Viewpoint	Egocentric viewpoints	A view from the perspective of the operator.	An egocentric viewpoint can provide a sense of immersion and motivation by allowing the user to experience the environment as if they were physically present. However, it might make it difficult to perceive and navigate correctly and accurately the robot in the surrounding environment.	[26, 46, 130]
	Exocentric viewpoints	A view from an external perspective.	Using an exocentric point of view can allow for the gain of abstract, symbolic insights by removing oneself from the immediate context. However, unnatural perspectives may degrade operative performance due to wrong position estimation.	[46, 151]
	Multiple viewpoint	Using multiple perspectives in a video or display.	Using two screens simultaneously can be effective, with one displaying sensor data and the other showing a live video stream of the remote environment. However, with multiple viewpoints, the operator must constantly integrate the information from multiple perspectives leading to fatigue.	[53, 97, 166, 222, 251]
Orientation	Ego-referenced maps	Maps that are based on the operator’s perspective.	Ego-referenced is suitable for vehicle conduction. However, when asked to recall the route using ego-referenced maps, the operators’ performance significantly deteriorates.	[24, 25]
	World-referenced maps	Maps that are based on an external perspective.	Holding a north-up map (world-referenced) is recommended when planning the journey. However, world-referenced maps are less effective during the navigation task.	[24, 25]
	Fixed viewpoints	A static perspective in a video or display.	Maintaining spatial orientation in unfamiliar environments might be difficult with fixed pitch and roll angle information. A solution might be prived by an independent camera.	[97, 171]

Factor		Description	Effect on teleoperation quality	Reference
Depth Perception	Underestimation of spatial information.	The underestimation of the distance or size of objects in an environment	Standard displays can alter depth perception and diminish performance, as operators must translate 3D information on a 2D surface.	[104, 118, 141, 221, 244]
	Lower viewpoint	A lower viewpoint in a video or display.	Lowering the teleoperator's viewpoint leads to more navigational errors, and depth perception is impaired in unstructured environments.	[26, 141]
	Stereoscopic displays	Displays that use two slightly offset images to create a 3D effect.	Stereoscopic displays provide depth information. However, they might increase operators' motion sickness and perceived stress levels. Furthermore, stereoscopic displays were only valuable for novel users.	[48]
	Integrated Sensor Data	The integration of sensor data, such as depth information, into a video or display.	Fused sensor displays may lead to potential data misperception and cognitive tunneling, such as unexpected objects.	[216]
Time delay	Move-and-wait Phenomenon	The repeated stop-and-go movements of the operator in a teleoperation task.	Move-and-Wait Phenomenon occurs when the time delay is more than one second, causing the operator to wait for feedback responses instead of sending continuous commands. The presence of a time delay in the control loop can reduce the efficiency of the human operator, i.e., it increases task completion time, error rate, and movement time.	[119, 177, 197, 202, 204]

Table 2.1: The table outlines the literature concerning the principal factors that affect remote operations. In particular, a summary of the findings of video image quality, Field of View (FoV), camera viewpoint, orientation, depth perception, and the time delay is provided.

In particular, they require scanning the images, perceiving the relevant elements, and understanding the information acquired. This process must be carried out for each perspective shown to the operator, which ultimately had to integrate them in a meaningful fashion, careful not to be biased prone [221]. To overcome this issue, Olmos et al. [166] suggested organizing the diverse visual feedback with a technique used in cinematography, such as visual momentum, i.e., visual continuity [245], and sound design.

2.2.4 Orientation

Operators should be provided with intuitive orientation solutions to operate in a foreign environment, and must have a global and local understanding of the environment. The former will provide all necessary information about the vehicle location, and the latter will support the operator during the driving maneuvers. If these references are missing, the operator may feel disoriented; thus, finding his position will ultimately be more confusing [29].

Ego-referenced maps, i.e., maps with a rotating point of view, have been systematically contributing to a more reliable global awareness [9, 25, 42, 128, 237, 239]. In a series of studies, Casner [24, 25] evaluated operator navigational awareness by employing, on the one hand, a north-up map and, on the other hand, a track-up map and asked to fly over a list of points of interest in an unfamiliar territory. The result inferred that pilots who handled the task using ego-referenced-based maps navigated more precisely than the others. However, when asked to recall the route, the pilots experienced a significant deterioration in performance. In contrast, the group that had previously performed the task using north-up-based maps performed better. The passive role assumed by the user in being driven to the destination contributed to a worse performance, at least when asked to recall the path. On the contrary, participants who were asked to play a more active role in the navigation planning, remembered the route better. The authors concluded that holding a north-up map (world-referenced) is recommended when planning the journey. However, when conducting the vehicle, a track-up map is suggested (ego-referenced).

Despite being equipped with maps and compass features, teleoperators can continue to have issues maintaining spatial orientation [141]. Not only orientation in unfamiliar environment is complex, but understanding the orientation of the own vehicle might also be challenging. Pastore [171] demonstrated that subjects had trouble extrapolating pitch and roll angle information when the vehicle was equipped with fixed cameras. Showing the vehicle position with other artifacts solves the problem, as operators have references that they can rely upon [29].

2.2.5 Depth Perception

Inspecting the remote environment via 2D display devices, such as screens, has been shown to alter the operator's depth perception [221], resulting in diminished operative performance [104]. The underestimation of spatial information occurs when operators translate 3D information on a 2D surface [118, 244]. Although humans tend to underestimate distances in natural conditions, this phenomenon is more acute when mediated by display devices [118, 244]. McGovern [141] demonstrated that subjects who were asked to control a UGV repeatedly underestimated the distance from and to the obstacles. By lowering the perspective and thus the teleoperator's viewpoint, operators made more navigational errors than when maneuvering a UAV, which presented an exocentric viewpoint of the remote scene. Depth perception is even more impaired if there is a loss of references, or salient size cues, e.g., in environments where objects are disorganized. During the 9/11 rescue responses at the World Trade Center, teleoperators had trouble assessing remote depth cues, due to the unstructured environment [26].

A solution to this problem is the use of stereoscopic displays. In a study on the effect of stereoscopic displays on teleoperator performance, Scribner and Gombash [199] reported statistical significance favoring stereoscopic displays, as they contribute to more accurate perception over monocular displays. However, stereoscopic displays increased motion sickness and perceived stress level. Furthermore, Draper et al. [48] demonstrated that stereoscopic displays were useful for novel users only. Other suggestions to support the operator's depth judgment include HMIs overlay and integrating sensor data onto video streaming. Marble et al. [134], in a qualitative assessment of the subject's preferences, observed participant willingness to overlay depth information onto the video feedback. Nielsen [161] found that integrated video data compensated for the insufficiency of information, though it slowed down the operative performance [107]. Furthermore, fused sensor displays might carry the potential of data misperception (e.g., unexpected objects) and cognitive tunneling [216].

2.2.6 Time Delay

Time delay, or simply latency, is a subject that is widely studied in teleoperation. Latency refers to the time delay that exists from the remote vehicle to the teleoperator's workstation. In this regard, the control commands sent by the operators to the vehicle are received with several milliseconds delay. Viceversa, vehicle sensor data, e.g., video images, LiDAR, and any other sensors data, are delivered at the operator workstation with some latency. Typically remote operation of UGVs or UAVs requires wireless communication contrary to many UUVs that remain connected by cable with the operator's workstation. Therefore, most of the acquired latency is produced by the communication infrastructure [197], which at the moment cannot wholly support ultra-low latency systems.

Literature findings have shown that the *move-and-wait* phenomena occur when the time delay is more than one second. Thus instead of sending continuous commands to the machine, operators prefer to wait for feedback responses [119]. It has been shown that the presence of time delay in the control loop has two effects: latency can make the closed-loop unstable [202], and secondly, the efficiency of the human operator is reduced [177]. Early studies by Sheridan and Ferrell [204] demonstrated that task completion time increased with increasing time delay. According to Sheridan and Ferrell [204], latency in the control loop profoundly impacted the teleoperator performance. Since these initial findings, researchers have continued to investigate time-delay effects on operators. Scott and Colin [197] discovered that by increasing the time delay of 225 ms in the control loop, the participant's error rate increased by 214% and movement time by 63.9%. Lane et al. [119] tested subjects under a more significant time delay, up to 3 seconds, and confirmed by Scott and Colin [197] findings. They moreover recorded a very similar increment of completion time of 213% when testing the subject with and without time delay.

Although it took longer to complete the task, the authors noted no statistically significant change of performance in conditions tested below 1 second. Lane et al. [119] noted that fluid video streaming with significant fixed latency is less harmful than shorter and variable ones. A strategy that was later employed in many other studies [44, 87, 124, 131].

Along with Sheridan [203] and Draper et al. [49] suggested tailoring HMIs that mediate for the time latency, for example, by employing advanced technology such as the Predictive Display, a safety requirement when relying on communication infrastructure. Ricks et al. [181] showed significant improvements of 17% in navigation and 1/5 of collisions by adopting HMIs as the Predictive Display.

2.3 Teleoperation Human–Machine Interfaces Concepts

In teleoperation, the interface is the means through which the operator provides inputs and receives feedback from the system being controlled. Several types of interfaces can be used for teleoperation, including manual interfaces, such as joysticks and buttons, haptic interfaces that provide a sense of touch and force feedback, and virtual reality interfaces that create a fully immersive environment for the operator. The type of interface used will depend on the specific application and the operator's needs.

In the following two sections, we review the state-of-the-art of direct and indirect teleoperation HMI concepts. We divided the literature concerning interfaces for teleoperation *inputs* and interfaces for teleoperation *feedback*. Thus, in Section 2.3.1, we include controls-based, touch-based, gestures-based, speech-based, gaze-based, and physiological-based inputs. In Section 2.3.2, we include visual devices, auditory devices, and haptic devices.

2.3.1 Interfaces for Teleoperation Inputs

Teleoperation input interfaces allow an operator to remotely control a vehicle or robot, usually through a computer or control device. These interfaces may include joysticks, touch screens, keypads, or other inputs that allow the operator to send real-time commands to the vehicle or robot. In the following section, a comprehensive review is provided (cf. Table 2.2).

2.3.1.1 Control-based input

Typically, teleoperation workstations rely upon a master/slave configuration[1]. In teleoperation workstations, primary devices try to simulate or reproduce in scale the remote vehicle control devices [206]. To this end, teleoperator inputs are mapped accordingly to the replica. Control devices such as pedals and steering wheels are widely spread in UGV and have been employed simultaneously to maneuver the robots longitudinally and laterally. Managing the vehicle with controls with which the operator is already familiar, as in the case of a steering wheel, had facilitated the task accomplishment, as the mental models already existed. Thereby the interaction with the environment is less challenging and more intuitive. This might be the case of the test vehicle teleoperated with steering wheels and pedals in Shen et al. [201] or the experiment by Kim and Ryu [109] that designed a shared vehicle teleoperation system to control multiple vehicles.

Devices such as joysticks or other similar controllers have also been extensively used in standard teleoperation workstations to communicate ongoing changes to the remote vehicle or actuators (Figure 2.3), e.g., to adjust the car position at the desired speed [153, 163, 233]. Thus, proportionally to their displacement, joysticks enable positioning control and rating control simultaneously [162]. Furthermore, dual joysticks provide separate input commands when, for example, the remote vehicle requires control camera movement [162].

Joysticks and controllers up to six Degree of Freedom have been used for controlling small robotic vehicles [225], and are employed in domains including art, medical training, and virtual environments [116, 138], and studied for recreational reasons [137, 193].

---

[1] We will refer to master/slave configuration as primary/replica.

Figure 2.3: In this image, a *Phantom Auto* remote operator controls an autonomous mobile delivery robot via a dual joystick controller that asked for support. An image from the 2020 *Phantom Auto* Press Kit.

2.3.1.2 Touch-Based Input

Most recently, the availability and affordability of touch-sensitive displays, e.g., mobile tablets, have begun to take hold in telerobotics. For instance, besides other listed techniques, Lockwood et al. [126] presented the interaction with touch-sensitive devices as the primary interaction technique for controlling and monitoring remote AVs (Figure 2.4). The interaction with the displays has been described as intuitive that does not require calibration. In a series of experiments, Lee and Zhai [121] could demonstrate that soft buttons enhanced with synthetic feedback (e.g., vibrato-tactile or audio) might provide better or at least similar performance compared to the standard hard buttons. Hoffman et al. [95] also suggested additional input interactions for controlling robots in search and rescue operations, including linear finger swipe, curved finger swipe, multi-finger swipe, multi-finger gesture, tap, and long press. Okishiba et al. [165] have designed input methods for controlling excavators for urban construction sites. The interface enabled the teleoperator to directly control and visualize the remote robot arm. The implemented HMI allowed to control the bucket velocity using virtual joysticks and the bucket position by dragging. Compared with the formal International Organization for Standardization (ISO)

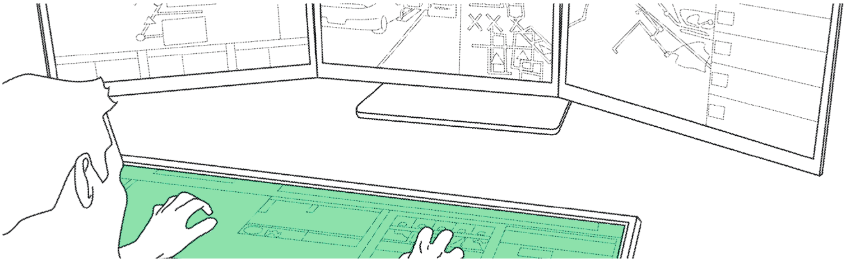


Figure 2.4: The illustration shows a teleoperator workstation envisioned by Lockwood et al. [126]. In particular, the teleoperator can assess and conduct the remote vehicle using touch-sensitive input devices.

teleoperation interface, the HMI proposed by Okishiba et al. [165] was superior to the conventional interface in task execution time, accuracy, and subjective usability evaluation.

The authors concluded that touch-sensitive displays could support an extended range of autonomy, from low to high, enabling simple commands and more complex movements. However, despite the intuitiveness of direct touch, challenges such as the interpretation of 2D inputs to 3D space mapping [234] or the design of adequate gesture-based touch controls remain present [129].

2.3.1.3 Gesture-Based Input

The recognition and classification of human gestures are extensively studied in HCI, including in robotics. For example Cohen et al. [36] researched human gesturing for vehicle controls and created a vocabulary of 24 "terms" for remotely controlling vehicles. In particular, the gestures were designed to control the camera motor arm. With a more elaborated approach, Frigola et al. [72] have set up a laboratory environment that could translate human gestures such as up, down, stop, turn, approach, and go in real-time. Furthermore, Ardizzone et al. [8] provided a software architecture for automatic recognition of human arms poses, with the intent to train robots to follow an arbitrary path. Other similar prototypes (Figure 2.5) have used visual gesturing to control small robots [69] or perform simple interactions with the virtual world [38]. Although being a valuable method, gesture-based interactions are often imprecise, ambiguous, or

(a) **To the right [69]** (b) **To the left [69]** (c) **Stop [69]**

Figure 2.5: This array of images shows the gesture-based driving mode proposed by Fong et al. [69]. To start, the operator raises his or her left hand to activate the robot and uses the right hand to specify the direction, e.g., to the right (a), to the left (b), or stop (c).

irregularly performed by the user. This method has been shown to frustrate the user as the interaction is complicated and exhausting [69]. Multimodal interfaces, i.e., speech as an input method, could solve this problem by discriminating input ambiguity.

2.3.1.4 Speech-Based Input

Another interaction methods assessed for vehicle teleoperation are voice commands. Voice commands are often combined with other control input devices [30]. For instance, to bypass the hand-busy problem, Peñín et al. [174] developed a telerobotic system for a live power line enhanced by voice commands. As the teleoperator's hands were both on the primary control devices, communication with the interface was done through voice input and speech synthesis. Thus, the interaction with the interface was achieved through voice instructions, which were valuable to control the camera point of view. Although the empirical studies indicated that the time to perform teleoperation tasks was similar to the time spent by experienced users doing the same tasks manually, the results obtained are promising as teleoperation for power line management could lower the risk of injuries. Similarly, researchers at the National Aeronautics and Space Administration (NASA) and the Defense Advanced Research Projects Agency (DARPA) had employed voice commands to control camera views, zooming, autonomy levels, and other functions when teleoperator's hands were busy controlling the remote robot actuators [81]. The authors noted that the operator's fatigue increased rapidly without voice.

Voice commands have also been used in flight/navigation control tasks. Draper et al. [51] showed that voice commands were significantly better than manual entries in time completion, accuracy, navigation measures, and pilot ratings, as HMIs for UAV workstations ordinarily featured multiple menu items and pages. Thus, the authors have improved operational performance by replacing the numerous sequential button presses with voice commands. Martín-Barrio et al. [136] confirmed this hypothesis, showing that voice commands are efficient and could reduce subjective cognitive workload. Medicherla and Sekmen [142], furthermore, have shown that voice control mechanisms for Human-Robot Interaction were reliable predictors of efficiency, with most of the participants, 75%, preferring voice-control over manual control.

This has also been documented for other robot teleoperation settings, where 94% of the subjects preferred voice command against conventional control devices [136]. Input through language seems a natural interaction method for humans to communicate information and is widely welcomed. The recent advances in natural language processing speech interaction, as shown, could be a reliable solution for communicating and expressing human statements.

2.3.1.5 Gaze-Based Input

In most common HMIs for vehicle teleoperation, control devices such as joysticks, keyboards, and computer mouses have been widely researched in the field of teleoperation. However, other streams of literature explored gazed-based interaction techniques for vehicle maneuvering. For instance, Yu et al. [254] approached gaze methods as a direct interaction technique for teleoperating small robots. The authors developed a non-invasive gaze tracking system based on 2D mapping estimation. The first studies showed reliable results for conducting vehicles to the desired target regions. Yet, as shown by Carreto et al. [23], gaze-based control interaction extends task completion time compared to traditional input devices like keyboards and mice.

Besides vehicle controlling, the operator often must control multiple devices simultaneously. This might be the case when the teleoperator must steer the vehicle while assessing the remote cameras. To solve this issue, authors have proposed gaze-based techniques that leverage natural patterns of interactions for camera control [226]. Gaze interaction methods have been shown to improve operational performances, as long as the teleoperator has to deal with multiple control devices and tasks, e.g., when maneuvering the remote AV and controlling the cameras [94, 257, 258]. Section 9.1 covers different gaze-based HMI concepts that adopt former eye-tracking mechanisms for vehicle teleoperation in greater detail.

As presented, gaze-based HMIs are intuitive interaction systems that do not require prior knowledge. However, eye-tracking mechanisms create the challenges of the Midas[1] touch issue [100]. That is, when eye gaze is used as an interactive media interfacing with computer systems, the eye primary function, to look and perceive visual information, should be distinguished from those deliberate interactions with the computer systems [228]. Thus, focusing on a display region could be interpreted as a selection mechanism, while in the states of the saccade, for instance, could be employed for screening or fatigue detection.

2.3.1.6 Psychological-Based Input

When it comes to vehicle teleoperation, Brain Computer Interface (BCI) has been used to investigate the feasibility of brainwaves to control small robots. BCIs technology turned out to be helpful, especially in contexts where the user cannot control the vehicle with their extremities, e.g., those with physical impairments or severe disabilities. Amai et al. [6] developed a brainwave-controlled mobile robot system that showed promising results in a proof-of-concept demonstration. In their study, participants were asked to wear a commercial Electroencephalography (EEG) device that used primary physiological input to control the vehicle speed and direction (Figure 2.6). In particular, brainwave signals

---

[1] Historically, King Midas is remembered in Greek mythology to turn everything he touched into gold. The problem arose when he touched his food and drink, which hardened into gold. As a result, unfortunately he could not reverse the action he had taken.

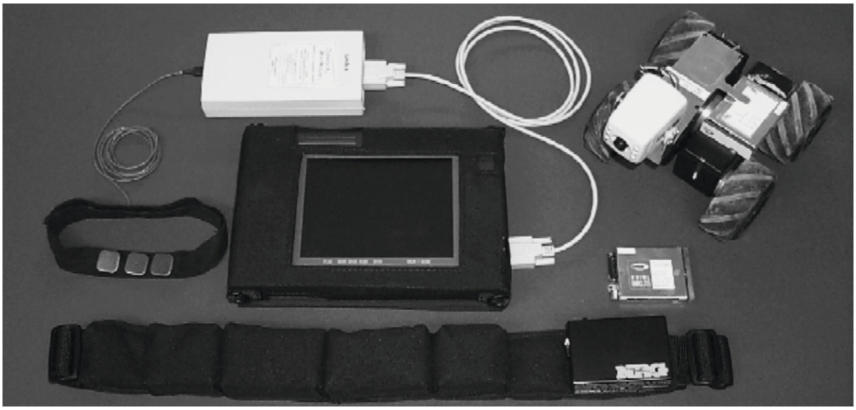


Figure 2.6: In this image, the equipment in Amai et al. [6]. Clockwise from left, the MindMouse headband and its hardware attached to a portable computer, the vehicle robot, radio data hardware, lastly, a battery belt.

and small facial muscular movements such as beta wave amplitude, jaw clench, and the movement of looking left/right were coded by a wearable computer that transmitted operational input wirelessly to the vehicle. The first results demonstrated that the beta wave to control the vehicle works well. However, the operator's mental arousal increased, inducing the vehicle to move faster as the beta wave increased. The less successful implementation was the turning maneuver, which had caused a significant delay in the control loop.

A comparison of two different BCI headsets was recently made by Vourvopoulos and Liarokapis [232]. The authors compared raw brainwave signals of the Neurosky Mindset[1] to the Emotiv Epoc headset[2]. The prior was based on the user's attention levels, decoded by the robot to accelerate or decelerate. The latter used a headset equipped with 14 sensors to control the vehicle. The authors noted that the usage of both devices was intuitive. However, the instability in the communication channel raised the question of the feasibility of such systems. Not to mention the training time factor.

---

[1] https://neurosky.com

[2] https://emotiv.com

Modality		Description	Reference
Input			
	Control-based input	Uses pedals, steering wheels, joysticks, and other similar controllers to control the remote vehicle. These devices are often familiar to the operator and facilitate task accomplishment. Complex control mapping on the devices might need clarification during the operation.	[109, 153, 162, 163, 201, 206, 225, 233]
	Touch-based input	Uses touch-sensitive displays, such as mobile tablets, to control and monitor remote vehicles. These interfaces are intuitive and do not require prior knowledge. However, it can be affected by the precision and dexterity of the operator's fingers.	[95, 121, 129, 165, 234]
	Gesture-based input	Uses human gestures to control robots. This method has been studied to translate gestures such as "up," "down," "stop," "turn," "approach," and "go" in real-time. However, it can be imprecise, ambiguous, and irregular, leading to user frustration.	[8, 36, 38, 69, 72]
	Speech-based input	Uses voice commands to control robots. This method can be combined with other control input devices, allowing the teleoperator to control the interface through voice instructions. However, background noise and accent and pronunciation variations can be affected.	[30, 51, 81, 136, 142, 174]
	Gaze-based input	Uses eye-tracking mechanisms to control the remote vehicle. These interfaces are intuitive and do not require prior knowledge. However, it can be affected by the "Midas touch" issue, where the eye's primary function of looking and perceiving visual information is distinguished from deliberate interactions with the computer system.	[23, 94, 100, 226, 228, 254, 257, 258]

	Psychological-based input	Uses brainwave signals to control the remote vehicle. These interfaces are helpful in contexts where the operator cannot control the vehicle with their extremities. Though they are limited by the instability of the communication channel and the training time required.	[6, 15]
Feedback			
Visual devices	**Monoscopic displays**	Monoscopic displays, such as LCD or LED monitors, convey visual feedback. They are widely available and affordable. However, they lack depth perception.	[117, 180]
	Stereoscopic displays	Stereoscopic displays provide 3D images by presenting slightly different perspectives to each eye. They improved perception and reduced cognitive demand. Nevertheless, the benefits are temporary, and the usefulness is limited in specific situations.	[52, 180, 223]
	Projections and caves	Projections and caves are larger and more immersive displays that surround the user and provide a wide FoV. They increase the sense of immersion and ability to show peripheral information. However, larger displays are limited in availability and potentially cause motion sickness.	[99, 125, 140, 167, 186]
	Head-mounted displays	HMDs are small displays worn on the user's head, providing a personal view of the remote environment. HMDs, increase the immersion and ability to build a mental model of the remote situation. However, there is potential for motion sickness and reduced focus on peripheral information.	[32, 96, 211]

Modality		Description	Reference
Auditory Devices	**Monophonic audio feedback**	Monophonic audio feedback provides information to the operator by carrying information classes that improve spatial awareness. Monophonic, yet, are limited in providing directional cues.	[154, 205]
	Stereophonic and spatial sound systems	Stereophonic systems could provide two or more sound channels and directionally map sound sources. They increase the sense of immersion and ability to provide cues on the size, nature, condition, and sounds of objects in the remote environment. Stereophonic systems require specialized equipment and may be more challenging to implement.	[18, 51, 146]
Haptic Devices	**Mobile haptic displays**	Mobile haptic displays are mechanical devices that transfer cutaneous sense data to the user, providing a sense of touch. Haptic feedbacks are fast, distinguishable, intuitive for users, and can provide warning messages and information about vehicle orientation and velocity. However, they may require specialized equipment and may need to be more widely available or affordable.	[20, 34, 111, 146, 184]
	Stationary haptic displays	Stationary haptic displays are larger and more complex haptic displays. They provide high-precision feedback over the user's entire body by reproducing the remote vehicle's acceleration, braking, and centripetal forces. However, they are limited in availability and affordability.	[133]

Table 2.2: The table resumes the review of the current teleoperation HMI concepts. We divided the literature findings in interfaces for teleoperation *inputs* and interfaces for teleoperation *feedback*. In particular, we include controls-based, touch-based, gesture-based, speech-based, gaze-based, and physiological-based input modalities. For feedback modalities, we include visual devices, auditory devices, and haptic/tactile devices.

Maneuvering vehicles using BCI technology is still at the beginning, as the control accuracy is insufficient to move vehicles firmly. Instead, BCI technology can be employed to predict the user's state, as the psychological and physiological data, as shown, can predict operational performance [15].

2.3.2 Interfaces for Teleoperation Feedback

In teleoperation, visual, auditory, and haptic interfaces deliver feedback to the operator. Visual interfaces give the operator a visual representation of the device or system being controlled, while auditory interfaces supply auditory feedback, such as sounds or speech. Lastly, haptic interfaces offer the operator force feedback, for example, when touching a display surface. In the following section, we review the current interfaces for teleoperation feedback (cf. Table 2.2).

2.3.2.1 Visual devices

One of the most applied modalities to receive feedback in teleoperation settings is via the visual channel. Visual displays are essential for teleoperators as most relevant information can be conveyed visually in an intuitive manner [35].

Currently, most teleoperation workstations rely on monoscopic displays, such as Liquid Crystal Display (LCD) or Light Emitting Diode (LED) monitors. In this regard, monoscopic monitors are often used stationary linked to a computer, and used to convey visual feedback [117]. A rich body of literature studied the effect of more immersive displays, such as stereoscopic displays, for telemanipulation. In a manual tracking task, Richard et al. [180] demonstrated that performances on a stereoscopic display were significantly better. Richard et al. [180] showed that task completion time was significantly lower in favor of the stereoscopic display, proving that stereoscopic images were less cognitively demanding. Besides, Thropp and Chen [223] underlined a correlation between stereo- and monoscopic images and FR, noting that better performances are registered when using stereoscopic images with high FRs. Despite these positive findings, Drascic [52] underlined that the gain we might have with stereoscopic images would be temporary. Moreover, the authors proposed longer-lasting benefits in situations where stereoscopic depth cues are crucial.

Figure 2.7: In this image, the professional drifter Vaughn Gittin Jr. drives an AV remotely using a steering wheel and pedals as input devices and an Head Mounted Display as output devices. An image from the 2019 *Samsung* Press Kit.

Other immersive displays, such as projections and caves, have been widely reported in teleoperation [125, 140, 186]. Voysys[1], for instance, promote a field of view of 210 degrees suggesting that the peripheral area can be compressed by hardware that can sense the user's head direction [167]. Therefore, areas that are not right in the operator's field of view can be shown in low quality. A teleoperator workstation with projections has also been presented by Iagnemma [99].

Alternatively, Sportillo et al. [211] used Head Mounted Display (HMD)s (Figure 2.7) to operate a semi-autonomous vehicle in a driving simulation, and Hosseini [96] demonstrated that HMDs helped the driver to build a quick mental model of the remote situation, bringing a better immersion. However, as shown, HMDs, compared to other devices, had induced motion sickness and led users to focus only on what is presented in their central area at the cost of other relevant information shown on the periphery [32]. Also, by increasing the display size, the perceived vehicle velocity rose, and the operator reduced the speed [32].

---

[1] https://voysys.se

2.3.2.2 Auditory Devices

Auditory displays can provide additional feedback when orientation is needed to alert the operator of imminent danger or support communication with in-vehicle passengers. Auditory cues are a valuable contribution to visual ones as they can improve spatial awareness and carry information classes when the visual channel is loaded [205]. In a study on audio feedback systems for simulated space robot teleoperation, Nagai et al. [154] demonstrated that auditory cues could reduce the operator's workload and provide practical support for decision-making.

To this end, monophonic audio feedback can yield valuable signals for the operator, or operators, as multiple users in parallel can receive the feedback. Stereophonic and spatial sound systems give a greater level of immersion since sound can be directionally mapped. Spatial sound systems have the strength to provide cues on the size, nature, condition, and sounds associated with individual objects of the remote environment [146]. A system that was implemented for teleoperation is described by Bourdot et al. [18], in which the authors investigate spatial sound systems in virtual reality settings. In particular, in their experiment, Bourdot et al. [18] used auditory feedback to enhance participant's limited FoV. The authors used audio streams that afforded a perceptual link to obtain information from the environment outside the FoV, enabling the inference of ongoing events outside the display. Moreover, Draper et al. [51] have proved that spatial audio displays have increased subject's situation SA in UAV operation. Similarly, we can imagine that spatial sound systems could be employed in AV teleoperation settings, as the teleoperator might not be conscious of the environmental changes happening laterally and behind.

2.3.2.3 Haptic Devices

Haptic displays are mechanical devices designed to transfer cutaneous sense data to the user, thus using as feedback systems modality. Haptic displays differ in their kinematic structure, work-space, and output force [146] and can be found in various forms and scales, from mobile to stationary. However, in both cases, the force feedback execution must be fast to simulate the human reception, e.g., a vibration on the steering wheel to simulate a car braking behavior or a force on the pedal that counterbalances an excessive speed. Haptic displays have been used to provide a warning message [34], e.g., for abrupt braking behavior. Also, to deliver information concerning the vehicle orientation, direction [34] and vehicle velocity [184].

Haptic feedback has been shown to be fast recognizable and intuitive from a user perspective [111]. This was also demonstrated in simulated UAV missions by Calhoun et al. [20]. Although the empirical study did not show statistical evidence favoring tactile feedback, participants preferred them over auditory cues, considering tactile feedback more salient and faster in attracting their attention [20].

More complex stationary haptic displays have been discussed in Mallwitz et al. [133]. Stationary haptic displays have the advantage of providing high precision on the whole teleoperator's body, e.g., an exoskeleton [133], or in the case of UGV, 6-axis simulators can reproduce the acceleration, braking, and centripetal forces of the remote vehicle. However, despite the technological advancement, economic and scalability factors make stationary haptic displays not widely used.

2.4 Discussion

Although machines will acquire more and more responsibility in executing specific tasks, vehicles will always demand expert control and supervision. In this context, human operators remain crucial for executing low-level commands in direct teleoperation workstations and defining high-level goals in indirect teleoperation workstations (cf. Figure 2.2). The choice of suitable HMIs for vehicle teleoperation will depend on the specific needs and constraints of teleoperation tasks, as well as the preferences and abilities of teleoperators.

Various solutions and systems have been proposed to aid operators in analyzing and executing maneuvers during direct teleoperation. Examples include implementing counterbalancing forces on controllers, adopting multimodal feedback, and employing advanced visualization to predict remote vehicle movements. Multimodal feedback has also been employed in a variety of different indirect teleoperation systems. These included visual and haptic feedback, visual and audio feedback, or even a combination of all of them. Multimodal feedback has been proven to provide better immersion, orientation, and communication of alerts. Moreover, augmented reality techniques are demonstrated to support operator spatial awareness, including orientation and depth perception.

In light of the literature reviewed in this chapter, we have derived a set of requirements, from both direct and indirect teleoperation, that have been considered during the course of this dissertation. (Table 2.3). In particular, the requirements embrace software and hardware solutions. They include: a) fluid video streaming with fixed latency, b) the FR at a minimum of 15 Hz, c) enlarged FoV, d) visual continuity, e) egocentric exocentric viewpoints, f) independent camera, g) world- and ego-referenced maps and h) predictive displays.

A fluid video streaming is critical for maintaining the viewer's attention to the video, as buffering or interruptions can hinder the viewer and disrupt the viewing experience. In case of connection unreliability, i.e., when the cellular connection can not provide a constant speed, the video streaming should be at fixed latency to consistently transmit video package data, which can help to ensure a smooth and uninterrupted viewing experience. In relation to this, the FR is an important factor in the quality and smoothness of a video (min. 15 Hz). A higher FR means that more frames are displayed per second, which can result in a more fluid and natural viewing experience. However, a higher FR can also require more processing power and bandwidth, which may not be suitable in all cases.

When it comes to assessing and operating in a remote scene, an enlarged FoV can provide the viewer with more visual information and a sense of immersion in the remote environment. A broader FoV can be achieved through wide-angle lenses or by digitally increasing the FoV. With regard to this, visual continuity is essential for maintaining the viewer's attention and interest in the video. A video with poor visual continuity may disorient or confuse the viewer. Thus, multiple vehicle cameras', in the best of cases, need to be stitched to create a seamless view of the remote environment.

Requirement	Description	Effect on teleoperation quality	Reference
Fluid video streaming with fixed latency	Video streaming has a consistent flow without delays or disruptions.	Video streaming with a consistent and slightly higher latency is less disruptive than fluctuating and shorter latency.	[119]
Frame rate at a minimum of 15 Hz	The video has a high frame rate, allowing for smooth and precise motion.	A minimum of 15 Hz image frequency is a suitable threshold for psychomotor and perceptual-based tasks.	[31]
Enlarged Field of View	The field of view of the video is expanded, providing a more comprehensive view of the environment.	Peripheral vision is relevant for lane-keeping when driving on curved roads and when planning the turning maneuver, as the driver must be prepared 1-2 seconds before. Moreover, a broader FoV can be employed before the driving task when for example, the teleoperator must assess the remote situation, thus planning tactics and strategic responses. However, to avoid over-salience, HMIs must be designed to consider a suitable threshold between limited and enlarged FoV.	[199, 227]
Visual continuity	The video maintains a consistent visual appearance, with no abrupt changes or disruptions.	Organize the diverse visual feedback using a cinematography technique such as visual continuity, i.e., sticking the camera viewpoint when possible.	[245]

Egocentric and exocentric viewpoints	The egocentric viewpoint displays the environment from the perspective of the person or objects it is focused. The exocentric viewpoint displays the environment from above, not tied to any specific person or object.	An egocentric viewpoint can provide a sense of immersion and motivation by allowing the user to experience the environment as if they were physically present. In contrast, an exocentric point of view can allow for the gain of abstract, symbolic insights by removing oneself from the immediate context.	[46, 130]
Independent camera	The video is captured by a standalone camera, not attached to any person or object.	Having an independent camera has been shown to improve search performance.	[97]
World- and ego-referenced maps	The video includes maps showing the camera's location concerning the surrounding environment and the person or object it is focused.	Holding a north-up map (world-referenced) is recommended when planning the journey. However, when conducting the vehicle, a track-up map is suggested (ego-referenced).	[24, 25]
Predictive Displays	The video includes a graphical user interface that anticipates and predicts future events, allowing for better decision-making and navigation.	To support the operator's depth judgment, researchers' suggestions include HMI overlays and integrating sensor data onto video streaming. Moreover, Predictive Displays can also mitigate the effects of time delay and improve performance.	[49, 134, 161, 181, 203]

Table 2.3: The table summarizes the qualitative requirements from the literature research on vehicle teleoperation embracing hardware and software solutions.

Along, egocentric and exocentric viewpoints can create different viewing experiences for the operator. Egocentric viewpoints can provide a sense of immersion and allow the viewer to feel as if they are experiencing the video first-hand. Exocentric viewpoints, on the other hand, can provide a more objective and detached perspective of the video. In this case, an independent camera that can be moved and operated independently of the viewer can increase the operators' understanding. Thus, separate cameras can be used to capture video from a variety of different perspectives and viewpoints, creating a more dynamic and immersive viewing experience.

Concerning knowledge and navigation of the surroundings, world- and ego-referenced maps should be provided. World-referenced maps are based on an external frame of reference and can help provide a larger-scale view of the environment. On the other hand, ego-referenced maps are based on the viewer's position and orientation and are valuable for navigating and providing a more detailed view of the environment that is specific to the viewer's location.

Eventuality, a Predictive Display can show the expected future state of the vehicle based on the time delay from and to the vehicle. The Predictive Display can provide operators with additional information and improve their situational awareness, allowing them to make more informed decisions.

The choice of suitable HMIs for vehicle teleoperation will ultimately depend on the specific needs and constraints of the teleoperation tasks, as well as the preferences and abilities of the teleoperator. Moreover, as Casper and Murphy [26] underlined, Situation Awareness remains a challenge when developing HMIs for remote vehicles. In relation to this, the main issue we will address in the following chapters deals with remote Situation Awareness. Thus, by detaching the operator from the actual environment, it is necessary to ask what information the teleoperator needs to directly conduct the remote vehicle safely. In particular, the overarching research questions we postulated were:

What information should be presented to the teleoperator during direct teleoperation? How should it be presented and what effect will the information have on the teleoperator?

2.5 Summary

Teleoperation is the remote control of a device or system from a distance, typically by employing technology such as robotics or computer-controlled systems. In this regard, teleoperation allows the manipulation of objects or processes without the direct physical presence of the operator in the remote environment. Prior research has distinguished three degrees of automation to classify teleoperation models: direct control, collaborative control, and automatic controls (Figure 2.2). Direct controls are the most widespread method for conducting a vehicle remotely, using control inputs such as steering wheels, pedals, and joysticks. By directly operating the vehicle, the operator is responsible for observing and perceiving the remote environment through displays and deciding on an appropriate strategy. Collaborative control involves a shared decision-making process between the operator and the vehicle, with both contributing to the control of the vehicle. Consequently, collaborative controls involve some autonomous functions, such as path control and obstacle avoidance, whereas the human operator must continuously find a navigation strategy for the vehicle. Automatic control entails a high level of automation, with the system responsible for vehicle navigation and guidance while the operator provides tactical and strategic commands.

However, despite the progress of automation, challenges still exist in designing interfaces for vehicle teleoperation. For instance, most factors that affect remote operations include the video channel quality, such as the frame rate, pixel density, and color depth, as well as the operator's cognitive load and Situation Awareness. Other factors include the limited Field of View, keyhole effect, communication delay, and the limitations of the remote system. These factors can impact the operator's ability to control and manipulate the remote system effectively, leading to reduced performance and potential safety risks. To solve these issues, engineers have employed visual and auditory devices in teleoperation settings to provide feedback to the operator. Visual displays, such as monoscopic and stereoscopic displays, are commonly employed to convey relevant information intuitively. Immersive displays, such as Head Mounted Display and projection systems, have also been used to enhance the operator's experience. On the other hand, auditory displays can provide additional cues to support the operator's decision-making and improve their spatial awareness. Such displays can be monophonic or stereophonic and provide directional information to help the operator understand the environment they are interacting.

Teleoperation workstations we reviewed commonly employ control-based input devices such as steering wheels and pedals to simulate the control of a remote vehicle. Joysticks and controllers with multiple degrees of freedom have also been used for controlling small robotic vehicles. More recently, we saw that the use of touch-sensitive displays has increased in teleoperation, allowing for more intuitive and direct control of remote systems. However, challenges still exist in interpreting 2D inputs in 3D space and designing adequate gesture-based touch controls. Eventually, despite being extensively studied in the field of human-computer interaction, gesture-based input still poses concerns, such as accurately detecting and interpreting gestures and designing intuitive gestures for specific tasks.

In conclusion, designing effective and intuitive interfaces for teleoperation remains challenging. However, from the literature we reviewed, we are confident that teleoperation technology will continue to develop and will likely have significant impacts in various fields.

3

Sensing Operator's Situation Awareness

In the previous chapter, we described the state-of-the-art of teleoperation. We illustrated current remote operation concepts and gave an overview of the significant challenges and requirements. We have seen that despite the numerous solutions that have been provided to compensate for the limitations of operating in remote, maintaining Situation Awareness (SA) remains a significant challenge. Towards this end, the acquisition and maintenance of SA is a critical motivation for developing cognitive ergonomic HMIs. In the following chapter, we will introduce established models and methods to assess SA before employing concrete methods and presenting empirical results in Chapter 6, Chapter 7, and Chapter 8.

3.1 Individual Situation Awareness

Based on individual characteristics and decision-making processes, the conduction of a vehicle is a complicated task. The task success will depend on the knowledge acquired about the remote traffic situation, which, according to Endsley [61], can be called Situation Awareness (SA). Originated in aviation and later spread to other domains, SA has been shown to be critical in many fields, from assessing pilot performance errors [189], to modeling ground vehicle crash avoidance systems [149]. Studies in the related field of research indicate that 60% of aircraft incidents were caused due to temporary loss of SA [59]. Thus, despite the expertise acquired during their carrier, from technical and theoretical training programs, operators continued to experience SA losses [59].

Earlier SA theories were created with the common purpose to model human information acquisition while performing a task [12, 58, 191, 207, 220]. Among these theories, the most polarizing and the most used have been developed by Endsley [58] and by Smith and Hancock [207]. Although both have been developed to explain the SA of a single operator, they differ in terms of the psychological approach. On the one hand, SA is seen as the product [58]. On the other hand, SA is seen as a process [207]. Specifically, Endsley's [58] three-level model takes an information processing approach, while Smith and Hancock [207] use a perceptual cycle model approach. According to Stanton et al. [213], these two perspectives dominate the SA literature.

Endsley [58] defines SA as *"the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"* (p. 97). In this frame of reference, SA is formed from three layers, perception (1° SA Level), comprehension (2° SA Level), and projection (3° SA Level). The first SA level defines the perception of the relevant elements and status information in the environment. The second SA level regards how the operator understands the situation, i.e., comprehends the current situation. This level refers to the perceived elements of the first level and combined them in a meaningful way. After perceiving and understanding the current situation, the third level to build SA involves projecting the information in their near future status (Figure 3.1).

These levels described SA as a product that leads to decision-making and performance of actions. According to Endsley's model, the acquisition and maintenance of SA are influenced by individual factors (e.g., abilities, experience, training) and system factors (e.g., system capability, interface design, stress/workload, complexity, automation).

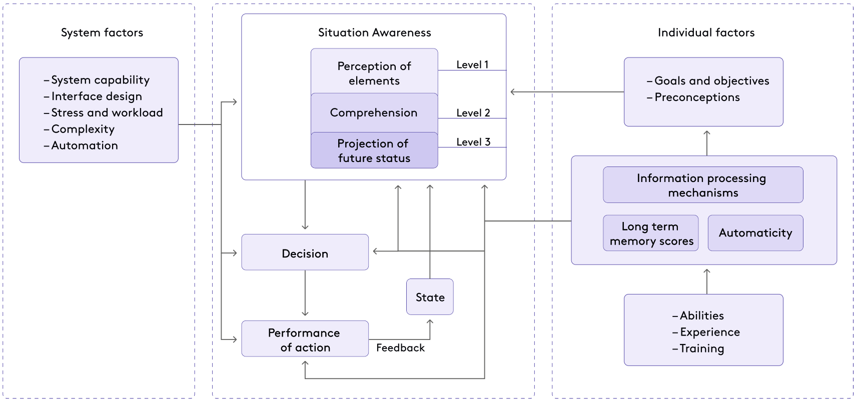


Figure 3.1: The Situation Awareness (SA) model in dynamic decision-making processes proposed by Endsley [63]. Here the SA, formed by three layers, is mediated by systemic and operator factors.

An essential aspect of vehicle operation is that the operator's SA is mediated by systemic factors. Thus, the task success will depend not only on the quality of the SA obtained but also on the interface and interaction component quality. An incomplete system contributes to operational errors and lesser immersion [152].

A well-known counter-theory to Endsley's [58] three-level model has been proposed by Smith and Hancock [207] that developed an SA theory based upon Neisser [158] perceptual cycle model, describing SA as a *"generative process of knowledge creation and informed action taking"* (p. 138). The model describes the cycle of perception and action, i.e., the human interaction with the world (Figure 3.2). In particular, human interactions are driven by pre-acquired mental models. The results of these interactions update, by modifying or confirming, our mental models. Thus the result of the interaction not only modifies the original patterns but also directs further interactions/explorations. The interaction with the world, which is defined as exploration in this theory, continues in an infinite cyclical nature.

The most substantial difference we see in Smith and Hancock's model compared to the prior described is suggested by the SA position. In fact, in the perceptual cycle model, SA resides neither in the world nor in the person but in

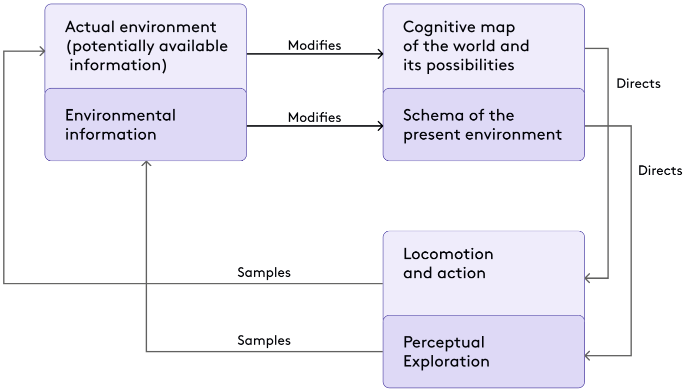


Figure 3.2: The Situation Awareness (SA) model by Smith and Hancock [207] is based upon Niesser's [158] perceptual cycle model. The model describes the cycle of perception and action, i.e., the human interaction with the world and thus the formation of SA.

the person's interaction with the world. The establishment and maintenance of SA pivots around internally held mental models that contain information regarding given situations. These mental models facilitate the anticipation of future events by directing their course of action. The operators will eventually supervise the development of action to ensure that it conforms to their expectations. Any unexpected event will elicit further interactions, or research, altering the operator's existing model.

Yet, Smith and Hancock's model describes a more dynamic aspect of SA, and thus according to Stanton et al. [212], more holistic. Endsley's three-level model, on the other hand, offered a more structured description of SA that has made it possible to measure SA intuitively, enabling the extraction of SA requirements at each level [188]. Its usefulness demonstrates the popularity of Endsley's model in informing the design and evaluation of systems. However, both models do not include the explanation of SA within operational teams and in operational systems.

3.2 Team Situation Awareness

As complexity increased and tasks have started to be shared among teams, prior SA models [58, 191, 220] have been updated to include inter- and intra-team SA [187]. Salas et al. [187] have defined SA held by a team as *"the shared understanding of a situation among team members at one point in time"* (p. 131). The prerogative of Salas et al.'s model is firstly the acquisition of SA by the individual and, secondly, the distribution of this knowledge among the team members to achieve the common purpose. The main feature of this model is the coordination of the awareness that the individual members have gained (Figure 3.3). Therefore, a proper team SA is more complicated to gain than individual SA as there is more information to decode and convey, not to mention the coordination of team members within highly complex systems [187].

Other definitions have been proposed to explain intra-team SA. For example, Endsley [62] defined intra-team SA by relating the individual SA, defining intra-team SA as the overlay of different SAs generated by different requirements [60]. Endsley [62] suggests that, during task performance, the SA of the individual may overlap with that of other members, as according to the three-level model, operators may have in common the perception, understanding, and projection of specific elements, but which have different purposes based on their task and role within the team, and thus the generation of different SAs with common elements. Endsley and Robertson [64] inferred that a strong team performance challenges single operators to hold a great SA upon their task and on the understanding of the distributed SA. This means that each member is required to share only the highest/abstract SA level, not to overload other operators. This may include merely the meaning and consequences for the team's goals. In this context, the main factors related to the task's performance quality are the shared objectives.

Wellens [238] further argues that a consistent overlap between the SA of the team members is essential for the smooth performance of the task. However, the overlap must continue to allow members to acquire SA individually. In these

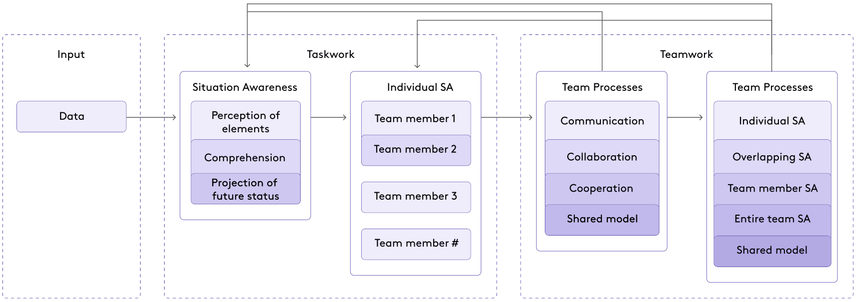


Figure 3.3: The model that explains team Situation Awareness (SA). We see team members must possess the SA related to their roles and goals while also holding the SA related to other team members.

terms, Wellens [238] defines team SA as: *"the sharing of a common perspective between two or more individuals regarding current environmental events, their meaning and projected future status"* (p. 272).

Bolstad and Endsley [16] suggested four core factors involved in shared SA, requirements, devices, mechanisms, and processes. *Shared SA requirements* refer to the already processed SA of the individual operator to be shared, including information about the status of tasks and other members. *Shared SA devices* are those devices available to the team for sharing information, including verbal communication, shared displays, or a shared virtual environment. By *shared SA mechanisms*, the authors mean the mental models that individual operators have and share. According to the authors, this is the core component as shared mental models function to support interpreting information in the same way within the team. Shared mental models are considered to have the role of facilitators for communication and prediction of other behaviors [175]. Lastly, *shared SA processes* refer to the processes for exchanging information among team members, including coordination and prioritization of tasks and the scheduling establishment.

We can conclude that team members must possess the SA related to their roles and goals while also holding the SA related to other team members. In particular, this overall SA includes the awareness that the operator has of the activities, roles, and responsibilities of other team members and the team as a whole, including goals and performance. The SA is then distributed within the team through communication, coordination, and collaboration to inform and modify other SA members. Thus, the SA in teams is threefold: SA of the individual team member, SA of other team members, and SA of the team as a whole.

3.3 System Situation Awareness

The increasing interest in systems thinking in HCI [22, 55, 235, 241] has also touched SA literature, introduced as Distributed Situation Awareness (DSA). DSA refers to the notion that systems can have and share SA. DSA, however, have a more historical basis, in particular, is based on the theory of distributed cognition approach proposed by Hutchins [98]. Hutchins suggests that in distributed cognition, people and the artifacts of a system combine to form a *"joint cognitive system."* The cognition is achieved through coordination among the elements of the system [11].

When it comes to DSA, cognition is seen as the awareness that connects the elements of the system. SA is, therefore, a system's component, which unlike the inter- and intra-team SA, is not generated by the operator's understanding but by the interaction between the elements. Thus, intrinsic of the system, like in Niesser's 1976 model. Moreover, Artman [10] stated that the interaction *"emerges in a context where artefacts and information technology partly structure the possibility of sharing and distributing information"* (p 1113).

Differently, Stanton et al. [214] suggested that the SA is the product of the cooperation between the elements. In this regard, despite being in the same situation, the system's elements held different SA. However, their SA can overlay, compatible, complementary, and SA's insufficiencies can be compensated. Stanton et al. [214] define DSA as *"knowledge activated for a specific task, at a specific time within a system"* (p.1291). The contribution of DSA systems explains and complements the literature on SA by considering SA not only individually or within a team but also collaborative systems.

Model	Author	Definition	Theoretical background
Individual	Endsley [58]	*"the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."* (p. 97)	Endsley's three-level model
	Smith and Hancock [207]	*"generative process of knowledge creation and informed action taking"* (p. 138)	Perceptual cycle model
Team	Salas et al. [187]	*"the shared understanding of a situation among team members at one point in time"* (p. 131)	Endsley's three-level model, Team work theory
	Wellens [238]	*"the sharing of a common perspective between two or more individuals regarding current environmental events, their meaning and projected future status."* (p. 272)	Endsley's three-level model, Distributed decision making model
System	Artman [10]	[awareness] *"emerges in a context where artefacts and information technology partly structure the possibility of sharing and distributing information"* (p. 1113)	Distributed cognition theory
	Stanton et al. [214]	*"knowledge activated for a specific task, at a specific time within a system"* (p. 1291)	Perceptual cycle, Distributed cognition theory, Distributed SA theory

Table 3.1: The table provides an overview of the reviewed Situational Awareness (SA) models. In particular, we report the SA frameworks developed to explain individual and distributed decision-making processes.

3.4 Measuring Situation Awareness

Various tools for designers and engineers are available for measuring different aspects of SA. However, the most widely used techniques include freeze probe, real-time probe, and post-trial self-rating [214] (Table 3.2). Historically, prior SA assessment methods have been developed to measure the operator's SA, e.g., via the Situation Awareness Global Assessment Technique (SAGAT) or Situation Awareness Rating Technique (SART) questionnaires. With SA knowledge and team awareness development, it is likewise feasible to be assessed, e.g., via the Coordinated Awareness of Teams (CAST) technique [80]. However, it is an open question on how to measure System SA.

Freeze probe techniques allow the assessment of SA through questionnaires. In this context, the freezing of the task is necessary to allow the user to answer one or more questions based on the prior acquired knowledge. During the assessment, the user cannot see the situation and is therefore asked to recall and elaborate on past events related to the situation. For example, during the SAGAT [63] the user is allowed to answer a series of questions relating to the three levels of SA described by Endsley [63], namely perception (1st SA Level), comprehension (2nd SA Level), and projection (3rd SA Level). Once users have completed the questionnaires, they will resume the interrupted task until a new query appears. Having assessed the user's SA, the answers are then compared with the simulation ground truth data or the expert's acceptance criteria. Although the freeze probe technique is a valuable method to assess operator SA during the task and solve the problem of collecting post-trial data, it can be intrusive as it interrupts the user in solving the task. A real-time probe might provide a solution to this problem, e.g., via the Situation Present Assessment Method (SPAM) [56], facilitating the participant to answer the queries without freezing and nonetheless enabling the assessment of SA in real environment. Nevertheless, it is still intrusive and biased as the whole amount of information is available to the operator.

Other questionnaires have been developed to be employed post-trial, without interrupting the user during the task. However, they lack the sensitivity of the freezing and real-time probe technique. Usually, post-trial questioners

Method	Description	Tool
Freeze probe	SA is assessed through questionnaires, yet, the task is frozen to allow the user to answer. During the freeze, the user cannot assess the situation and must recall and elaborate on the memories based on past events. The participant answers are then compared with ground truth data.	Situation Awareness Global Assessment Technique (SAGAT) [63]
Real-time probe	SA is assessed through questionnaires without freezing the task, allowing the user to answer in real time. Nevertheless, it is still intrusive and biased as the whole amount of information is available to the operator.	Situation-Present Assessment Method (SPAM) [56]
Post-trial self-rating	SA is assessed through self-rated questionnaires after the task has been completed. However, they need more sensitivity to the freezing and real-time probe technique.	Situation Awareness Rating Technique (SART) [220]

Table 3.2: The table summarizes the different methods for measuring Situation Awareness.

are self-rated techniques that measure the operator's SA in various domains. For example, having administrated the Situation Awareness Rating Technique (SART) query [220], inference on operators arousal, spare mental capacity, concentration, and division of attention is available. Also, data about the quality of the information, instability, complexity, variability, and familiarity with the situation is retrievable.

While the literature on individual SA is extensive, there has been limited emphasis on developing measures for team SA. The Coordinated Awareness of Situation by Teams (CAST) technique, proposed by Gorman et al. [80], addresses the collective SA levels within a team and among its members. Another approach involves using the SAGAT query, administered to all team members. However, the SAGAT questionnaire shares criticisms with those mentioned earlier, particularly its challenge in real-world collaborative tasks due to impractical task freezing.

To address this issue, some studies evaluate team SA by correlating it with individual member performances, such as counting errors. Nonetheless, this method has faced criticism, as good individual performance does not necessarily indicate good team SA [188].

3.5 Discussion

Situation Awareness is a complex concept that refers to an individual's perception of the elements in their environment, their comprehension of the current situation, and their projection of future events. It is a critical factor in the pursuit an adequate performance of complex tasks. Nevertheless, the concept of SA has been the subject of criticism and debate in the academic literature.

One major criticism of SA is the subjectivity of the matter, i.e., SA is based on an individual's perception of their environment. Different operators may have different interpretations of the same situation, making it challenging to assess SA accurately. The subjective nature of SA could drive difficulties in comparing individuals' SA or generalizing the findings from SA questionnaires. A further criticism of SA is the dependencies on the situation in which it is being evaluated. Reviewers questioned the transfer of the findings from one situation to another or the development of effective training programs across different contexts. Measuring SA could likewise be problematic, as it is a complex and multifaceted concept involving cognitive and emotional processes. Various tools are available for assessing SA, such as questionnaires and performance measures. However, each has limitations and may not provide a comprehensive assessment of an individual's SA. Moreover, SA might be influenced by factors outside an individual's control. This might be the case with the system's design or information availability, making it questionable to attribute any SA deficits solely to an individual's knowledge. Eventually, SA is one of many factors influencing performance in complex systems. Other factors, such as knowledge, skills, and motivation, may play a role in an individual's performance, and it could be challenging to attribute performance outcomes exclusively to SA.

Nevertheless, despite these criticisms, SA remained widely assessed and promoted as an essential concept in human performance to improve the effectiveness of complex systems. There are several benefits to assessing SA. These include: a) identifying areas of strength and weakness, b) improving performance, c) enhancing safety, d) enhancing efficiency, and e) facilitating decision-making.

Assessing SA can help researchers identify an individual's strengths and weaknesses regarding their ability to perceive, comprehend and anticipate environmental events. This information could tailor training and development programs to address specific areas of need. For example, if an individual scores inadequately on items related to projection, they may benefit from training on scenario-based planning or decision-making skills. By identifying areas of SA deficit, researchers can implement strategies to improve SA and, in turn, improve performance. This could include providing training, improving the system's design, or increasing the availability of information. For instance, if individuals struggle with SA due to inadequate information presentation, providing more data or explicit content may improve their SA and performance. Consequently, reduced risk of errors and accidents might be obtained. Proper SA is alike necessary for the efficient performance of tasks. By identifying and addressing deficits in SA, researchers could improve efficiency and reduce the time and resources required to complete tasks. This might be the case when an individual progresses with difficulty during task completion for over-salience. A more significant amount of information may require additional time for the operator to complete the tasks or cause more fallacies, leading to inefficiencies. Improving SA by delivering the right information might increase efficiency and reduce the time and resources required to complete tasks. Thus, leading to better decision-making and improving the decisions' quality. Ultimately, assessing SA can provide valuable insights into an individual's abilities, improve performance, enhance safety, increase efficiency, and facilitate better decision-making. Various tools are available for assessing SA, such as questionnaires, performance measures, debriefing, simulations, and training. A combination of these tools may be needed to comprehensively understand the individual's SA and identify specific areas for improvement. Certainly, further research is needed to understand the limitations of SA better and to develop more accurate and comprehensive methods.

3.6 Summary

In this chapter, the acquisition and maintenance of SA has been discussed as a critical motivation for developing cognitive ergonomic interfaces. In this context, SA refers to the ability to understand the current state of an environment, predict future events, and effectively respond to changing circumstances. SA is a critical concept in aviation, military operations, and emergency response, where the ability to make fast and accurate decisions could lead to severe consequences.

Therefore, SA requires constant attention, environmental monitoring, and irrelevant or ambiguous information screening. It also requires the ability to quickly and accurately analyze and interpret information and to make decisions based on this analysis. A high level of SA might be achieved via the use of HMIs that helps to gather and analyze the remote environment. An adequate SA is essential in many high-stakes environments, as it can help individuals and teams respond effectively to changing circumstances and make informed decisions.

4

Sensing Operator's Cognitive Workload

In Chapter 3, we presented the theoretical framework of Situation Awareness (SA). Specifically, we introduced established models and methods of individual and distributed SA. In the following chapter, after briefly defining cognitive workload in automotive literature, a novel method for cognitive sensing, namely thermal imaging, is presented and evaluated. In particular, we attempt to describe the state-of-the-art in using infrared thermal imaging in driving simulators by conducting a systematic literature review and meta-analysis. This eventually led us to gain a clear picture of the effectiveness of this technology, which was later used as an evaluative method in Chapters 8 and 9.

4.1 Cognitive Workload

Cognitive load has always been a topic of importance in HCI [253]. It is determined by external and internal factors, i.e.: (a) extraneous processing, triggered by cognitive processes that do not support the subject's objective; (b) intrinsic processing, in which the subject is engaged in comprehending the task; and lastly (c) germane processing in which the subject is engaged in cognitive processing as a mental organization of the information received [45]. Similarly, in automotive research, the cognitive workload is used to describe the amount of mental resources an operator can spend on performing a particular task. It is widely used to study the causes and effects of distracted driving or measure the consequences of affective states on the driving task [101, 127, 253]. Young et al. [253] highlight that several often conflicting definitions of cognitive workload exist. However, they describe cognitive workload as a multidimensional measurement of mental capacity, with attention being a limited resource that diminishes task accuracy and efficiency when overloaded [253]. For example, they highlight the standard technique of inducing cognitive overload in driver studies by employing secondary tasks that force the subject to use more working memory. Understanding cognitive workload is pivotal in both HCI and automotive research, as it not only reflects the mental resources invested in tasks but also impacts task accuracy and efficiency under cognitive overload.

4.1.1 Measuring Cognitive Workload

Varied ways of measuring cognitive workload exists, such as self-assessment questionnaires like the NASA Task Load Index (NASA-TLX) [92], Subjective Workload Assessment Technique [179], and Rating Scale Mental Effort [259], measuring task error rate and efficiency, or by measuring physiological parameters such as heart rate or brain activity [127, 253]. As there often exists a dissociation between subjective workload and task efficiency (i.e., subjects report higher cognitive workload while task efficiency stays the same) as Vidulich and Wickens [229] show, physiological measurements are captured in cognitive workload studies to provide an objective truth [253]. Measuring these physiological parameters is usually achieved by attaching sensors to the skin [127], which may constrict movement, impeding on primary task performance such as the driving task in an automotive study, proving impractical for real-time and continuous tracking of subjects [236].

An emerging and exciting response to this deficiency is using infrared thermal imaging cameras to capture physiological data of test subjects [127]. Like a camera used in photography, thermal infrared cameras use a lens to project electromagnetic radiation from a scene onto a two-dimensional sensor plane. This plane is equipped with photo-sensors capable of sensing radiation in the visible light spectrum. Infrared thermal cameras utilize sensors sensitive to thermal radiation between 3 $\mu\mathrm{m}$ to 5 $\mu\mathrm{m}$ for Mid-wave Infrared cameras or between 8 $\mu\mathrm{m}$ to 14 $\mu\mathrm{m}$ for the Long-wave Infrared cameras. Thus, covering a thermal radiation band from 3 $\mu\mathrm{m}$ to 14 $\mu\mathrm{m}$ [74]. This produces an RGB video image, where measured temperature values are encoded as color values. Because the human skin emits thermal radiation in this spectrum, its emissivity is close to perfect black-body radiation [91]. Thus, thermal imaging is ideal for capturing body temperature [74] and determining spatial differences between body regions. Prior research shows that a correlation exists between self-assessed levels of cognitive effort and temperature readings captured using a thermal infrared camera [168]. Since then, the usage of infrared thermal imaging has been increasing in automotive research [127].

In this chapter, we performed an analysis of current literature to show the development of infrared thermal imaging technology in driving simulator studies. We limited our literature to automotive research to ensure a standard experiment setup that is well suited for comparison and also because automotive research can benefit much from an unobtrusive, real-time measurement of cognitive load. In this regard, we determined and formalized three areas of focus on which the literature was analyzed: the application of thermal imaging, the method of data capture, and how the captured image data was processed. Then, we conducted a meta-analysis that establishes the effectiveness of using thermal cameras for detecting cognitive distraction in a driving context. Eventually, we discuss our findings and suggest improvements in experimental design, highlighting opportunities and challenges for future research.

4.2 Methodology

Our goal in this chapter is to assess the effectiveness of utilizing infrared thermal imaging to measure the cognitive workload of study subjects in HCI research, and identify areas of improvement. As we intended to perform statistical anal-

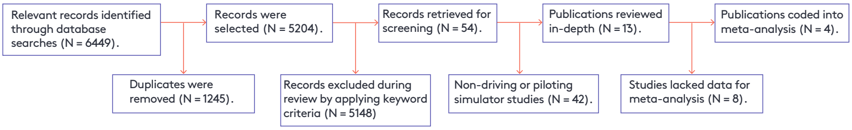


Figure 4.1: A summary illustration of the literature we collected following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [148]. Of 5204 unique records, 54 were retrieved for screening, 13 studies were analyzed in-depth, and four coded for the meta-analysis using the Population, Intervention, Comparison group, Outcomes, Time frame, Setting (PICOTS) framework [183]. Unfortunately, eight studies lacked the necessary statistics needed for the meta-analysis.

ysis, we focused our literature collection on the field of automotive research, which enabled us to compare similar experimental setups (i.e., participants in driving or piloting simulators), significantly reducing external variables.

Our approach is threefold: first, we conducted a systematic literature review of available literature and identified common points of comparison. Secondly, we conducted a statistical meta-analysis on the collected literature on the effectiveness of measuring cognitive workload using infrared thermal imaging. Finally, we discuss our findings.

4.2.1 Systematic Literature Review

We performed the systematic literature review using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [148], for which a summative illustration can be found in Figure 4.1. The search was conducted in January 2020 through a series of electronic search engines: ACM DL[1], IEEE Xplore[2], SAGE[3], Springer[4], as well as Elsevier[5] (i.e., ScienceDirect), using the keywords: facial temperature, facial thermography, thermography, infrared camera, thermal camera, and thermal imaging. After duplicates were removed a total of 5204 articles was found. Two additional records were further identified from other sources and added to the data set. These terms

---

[1] https://dl.acm.org
[2] https://ieeexplore.ieee.org
[3] https://journals.sagepub.com
[4] https://link.springer.com
[5] https://sciencedirect.com

were primarily thought to not exclude early approaches or proposals by simply searching very specifically. Upon having downloaded from each database a CSV file, we parsed the titles and keywords applying the following terms: cognitive strain, cognitive load, cognitive workload (or work-load), mental strain, mental load, mental workload (or work-load). We used the keyword variants of driving (e.g., driv*, driv* behavior) to further contain the research. Toward this end, 54 articles were retrieved for screening, and 13 papers were selected for in-depth review. These papers were selected based on the criteria for using a driving or piloting simulator and utilizing an infrared thermal imaging camera.

Eventually, we conducted a meta-analysis of the collected records to test the overall effectiveness of using thermal imaging to measure cognitive load in driving simulators. Unfortunately, eight studies lacked the primary data for the meta-analysis, i.e., did not report the necessary statistics needed to calculate the Standardized Mean-Difference (SMD) and their Confidence Intervals (CI) [122]. Therefore, for the meta-analysis, four studies of 12 have been coded using the PICOTS framework (Population, Intervention, Comparison group, Outcomes, Time frame, Setting) [183].

4.3 Findings of the Systematic Literature Review

We obtained thirteen records from our literature collection method, which we analyzed in detail. After primary analysis of the literature, we formalized three main focus areas for analyzing and comparing papers: (a) Thermal Imaging Application, (b) Data Capture, and (c) Image Data Processing. These areas were common in all the papers and were selected and worded to be immediately apparent, readily comparable between different studies, and, taken together, they describe an experiment in its entirety from data collection to final results.

4.3.1 Thermal Imaging Application

Several trends can be determined in thermal imaging in the literature we reviewed. Six research papers attempt to demonstrate the validity of using infrared thermal imaging as a tool to provide physiological measurements of a subject in a driving simulator [5, 7, 103, 150, 240, 255]. Two papers use these measurements to validate an experimental setup or automotive design [102, 108]. Two papers demonstrate the usage of infrared thermal imaging to determine moments of driver distraction or the high cognitive workload of a driver in real-time [170, 172]. One paper utilizes infrared thermal imaging to determine head position and orientation in the car [247]. Finally, one additional paper proposes a theoretical method but not implementation details [147].

4.3.1.1 Thermal Imaging as a Physiological Measurement Tool

In six articles we reviewed the attempt to develop methods of using infrared thermal imaging as a physiological measurement tool. In two experiments, Or and Duffy [168] have shown that forehead temperature, in arousal states, tends to remain stable while the nose temperature decreases. In the first experiment, participants were asked to drive in a simulated environment and were exposed to a secondary task (mental arithmetics). The facial temperature was measured pre- and post-stimulus, and the Modified Cooper-Harper questionnaire was assessed. First, the authors found a correlation between nose and forehead temperature with the Modified Cooper-Harper questionnaire. Secondly, they observed that nose temperature decreased when the participant was occupied with the secondary task, while forehead temperature remained constant. The second experiment was conducted in real and simulated driving conditions. Similar results were found for the simulation, i.e., decreased nose temperature and constant forehead temperature. However, these results were not confirmed under the actual driving condition.

Anzengruber and Riener [7] show that their method is adequate to infer "mental conditions of the driver" but point out that the difference in states is very low (0.25 °C) and that other factors such as food intake, fatigue, and body position might have a more significant impact on body temperature, which limits the approach in real-world scenarios. Kajiwara [103] shows that while the combination of physiological features captured is a valid measure of driver’s mental workload, noise and sunlight significantly inhibit the potential of this system to be implemented in real-world conditions. Murai et al. [150] demonstrate that the difference between nasal and forehead temperatures is quicker to react to moments of the cognitive load than measurements of high- low-frequency heart rate analyses. Zhang et al. [255] demonstrate that facial temperatures change in driving situations that require sudden and quick driver input (such as an unobserved vehicle swerving into the driver's lane). Finally, Wesley et al. [240] show that the mean temperature in the supra-orbital signal is significantly higher when participants in a driving simulator have to fulfill distracting tasks, showing that infrared thermal imaging can be used to determine the driver's distraction.

In contrast, Altschaffel et al. [5] find that temperature variations observed in their data sets are not statistically significant to determine stress or cognitive load. Temperature differences occurred sporadically or with a significant delay between the increased workload demand and skin temperature reaction, showing that the method utilized was inadequate as a workload measure. Since the reason is not immediately evident from the data provided, there could remain several unidentified factors that influence the efficiency of thermal imaging as a physiological measurement tool.

4.3.1.2 Thermal Imaging to Validate Experimental Setups or Designs

Both Jeong et al. [102] and Kim et al. [108] use infrared thermal imaging as a measurement to validate experimental setups. Jeong et al. [102] utilize variations in participants' nasal temperature to argue that a vehicle control method utilizing joystick control produces less driver workload and "stress" than using a steering wheel to control a vehicle in identical driving conditions. Moreover, Kim et al. [108] use their participants' skin temperature measurements to validate a proposed simulated highway design over another. Both of the stated papers find that their measurements are in-line with additional physiological measurements.

4.3.1.3 Thermal Imaging as a Real-time Measure of Driver Workload

In two research papers, the authors show that a real-time measurement of driver workload and driver distractions is feasible using an infrared thermal imaging camera [170, 172]. For instance, Pavlidis et al. [172] use the measure of perinasal perspiration to activate a biofeedback sensor that alerts the driver of distracted

Study	Camera Model	Sensitivity	Refresh rate
Jeong et al. [102]	FLIR A20	.1°C	30 hz
Kim et al. [108]	FLIR A20	.1°C	30 hz
Or and Duffy [168]	Mikron MicroScan 7200V	.08°C	n/a
Pavlidis et al. [172]	FLIR Tau 640	.05°C	n/a
Anzengruber and Riener [7]	FLIR SC655	.05°C	n/a
Wesley et al. [240]	ThermoVision SC6000 MWIR	.01°C	n/a
Kajiwara [103]	Vionics Air 32	n/a	10 hz
Zhang et al. [255]	Optris PI640	.075°C	n/a
Murai et al. [150]	n/a	.1°C	n/a
Wu et al. [247]	n/a	n/a	n/a
Altschaffel et al. [5]	n/a	n/a	n/a
Panagopoulos et al. [170]	Previously collected dataset	n/a	1 Hz
Mitas and Ryguła [147]	Theoretic setup	n/a	n/a

Table 4.1: The table shows an overview of the thermal cameras used in the reviewed literature. Sensitivity is specified as the temperature difference where a change is indistinguishable from noise, the Noise Equivalent Temperature Difference (NETD).

driving using a self-starting Cumulative Sums algorithm that activates and deactivates a warning light when perspiration passes a threshold, notifying a driver of his distracted mental state. Further, Panagopoulos and Pavlidis [170] show that perinasal perspiration combined with other physiological measurements can accurately predict distracted and aggressive driving moments in a data set from a distracted driving experiment.

4.3.1.4 Thermal Imaging to Determine the Driver's Head Position

The only paper Wu et al. [247] that does not directly use infrared thermal imaging as a physiological measurement demonstrates that tracking the driver's head position in a cab using thermal infrared imaging is feasible in an unobtrusive manner. Thermal imaging was only used to identify facial landmarks to calculate head tilt and position. They also demonstrate minor pitch and yaw changes, ranging from one to two degrees, in the driver's head positions.

4.3.2 Thermal Imaging Data Capture

As using an infrared thermal imaging camera was one of the criteria for inclusion in the in-depth review, all of the papers analyzed use thermal imaging to capture

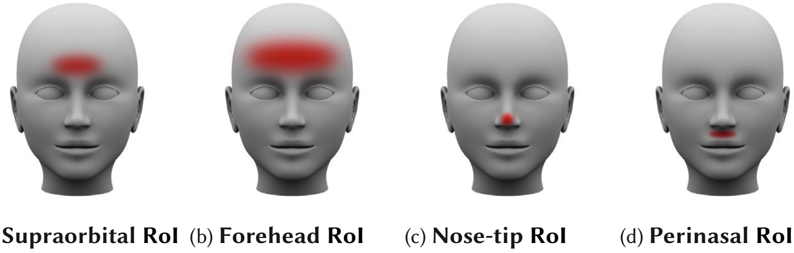

(a) **Supraorbital RoI** (b) **Forehead RoI** (c) **Nose-tip RoI** (d) **Perinasal RoI**

Figure 4.2: An overview of the Regions of Interest (RoI) we identified during the literature analysis. From left to right, (a) supraorbital, (b) forehead, (c) nose-tip, and (d) perinasal regions.

subject data. Most of the research papers examined set their studies in simulators, except for Wu et al. [247], who performed a study in a moving vehicle, and for Or and Duffy [168], who conducted multiple studies in both, natural and simulated environments. Other studies used a pre-captured dataset of drivers [170] or described a theoretic experiment setup [147]. Specifications such as camera models, image sensitivity, and refresh rate are provided in Table 4.1.

4.3.2.1 Regions of Interest

Measuring the cognitive effort of a driver is achieved by measuring a change in face skin temperature [74]. This is in line with the seminal paper by Or and Duffy [168], which shows that forehead temperature does not change significantly between start and end of an experiment setup, but that the nasal temperature does, and that it decreases with the amount of cognitive effort used by the driver. No age effects were observed, however; Or and Duffy [168] observe that the cognitive effort is significantly lower when piloting an actual vehicle compared to driving in a simulator.

Measuring nose tip and forehead temperature has been repeatedly utilized in the literature [102, 103, 150, 255]. In these studies, the nose tip and forehead region have been determined manually [102, 103] or through spectacles of precise dimensions [150]. Alternatively, a Histograms of Oriented Gradients (HOG) detector was trained on a set of manually labeled snapshots to detect the regions automatically [255].

Additional Regions of Interest (RoI) have been considered in other simulated studies. Wesley et al. [240] trace the area above the eye affected by the supraorbital nerve. The perinasal region, as defined by Shastri et al. [200], which describes the area between the upper lip and the nose, is utilized by two studies [170, 172].

Further research, however, did not limit their data collection to a specific area of the face; instead, they recorded temperature values of the entire face [5, 7, 247]. In contrast, Mitas and Ryguła [147] suggest employing infrared thermal imaging to measure Percentage Eye Openness Tracking, limiting their regions of interest to the eyes only. Figure 4.2 illustrates the identified RoIs.

4.3.2.2 The Capture of Further Physiological Data

In addition to capturing infrared thermal imaging data, papers obtained in our analysis record additional physiological data to confirm their findings or demonstrate the validity of their method. Jeong et al. [102] captured the Electrocardiogram (ECG), Electromyogram (EMG), Electrooculogram (EOG), and Galvanic Skin Response (GSR) of the study participants to demonstrate the correlation between other physiological measurements and data obtained through infrared thermal imaging.

Zhang et al. [255] collected ECG and GSR to correlate their findings. Kim et al. [108] collected ECG, EMG, EOG, and GSR data in the same manner as Jeong et al. [102]. Murai et al. [150] measured the heart rate of their participant to verify their method of using thermal imaging. Panagopoulos and Pavlidis [170] combined the thermal image readings of their data set with heart rate, breathing rate, and GSR readings to measure cognitive workload.

4.3.3 Thermal Imaging Data Processing

As with digital video cameras, thermal cameras are subject to noise and other image artifacts [74]. Valuable temperature data, in this regard, can be obtained from image processing tools. Processing the image data and extracting temperature values is required to reproduce an experiment and verify the data obtained accurately. This subsection will detail the methodologies utilized to extract data in the reviewed literature.

Study	N	Age		Gender	
		Mean	SD	Male	Female
Panagopoulos and Pavlidis [170] [a]	59	n/a	n/a	26	33
Pavlidis et al. [172] [b]	47	n/a	n/a	n/a	n/a
Or and Duffy [168], Experiment I [c]	16	24.2	4.9	16	17
	17	48.8	7.1	–	–
Or and Duffy [168], Experiment II [c]	6	25.7	5.2	9	4
	7	46.7	6.4	–	–
Kim et al. [108]	33	n/a	n/a	30	3
Zhang et al. [255]	18	27.5	4.5	14	4
Jeong et al. [102]	13	25.2	2.0	13	0
Wesley et al. [240]	11	27.5	n/a	4	7
Anzengruber and Riener [7]	9	n/a	n/a	9	0
Kajiwara [103]	4	22.0	.71	4	0
Wu et al. [247]	1	n/a	n/a	1	0
Murai et al. [150]	1	n/a	n/a	1	0
Altschaffel et al. [5]	n/a	n/a	n/a	n/a	n/a
Mitas and Ryguła [147]	n/a	n/a	n/a	n/a	n/a

Table 4.2: A detailed overview of the participants recruited in the reviewed literature. Note: some research papers have separated the participants into groups of age: *a)* two groups of ages: 18-27 and 60+, *b)* two groups of ages: 18-27 and 55+, gender-balanced, *c)* two groups of ages: 18-35 and 36-64.

Kajiwara [103] extracts monochrome square pixel values from the nose tip and forehead regions, averages the pixel density per frame per square, and subtracts the density of the forehead area from the nose area, obtaining the difference in temperature signals between the forehead and nose over time. This value is combined with other physiological features such as skin potential change and skin conductance over time to measure the driver's mental workload. Zhang et al. [255] apply the same method of difference in temperature signals between forehead and nose-tip, identified by their HOG detector. Wesley et al. [240] produce a mean of all collected temperature values in the tracked region of interest per measurement, obtaining a one-dimensional signal from the two-dimensional thermal data. Noise in the thermal data was reduced using a Fast Fourier Transformation-based method.

Anzengruber and Riener [7] utilize an estimation-maximization algorithm to compute a Gaussian Mixture Model to identify regions of the face with the highest temperature. By calculating a mean, these values are then normalized by subtracting the starting temperature of the regions identified, thus determining the change in temperature of a facial region during a testing configuration.

Pavlidis et al. [172], and Panagopoulos and Pavlidis [170] determine the amount of perspiration in the perinasal area by performing a contour-based black top-hat transformation which generates an image of areas of perspiration on a black background, thus enabling the calculation of the amount of perspiration as the energy in the image generated by the colored spots. They utilize the previously described method of Shastri et al. [200].

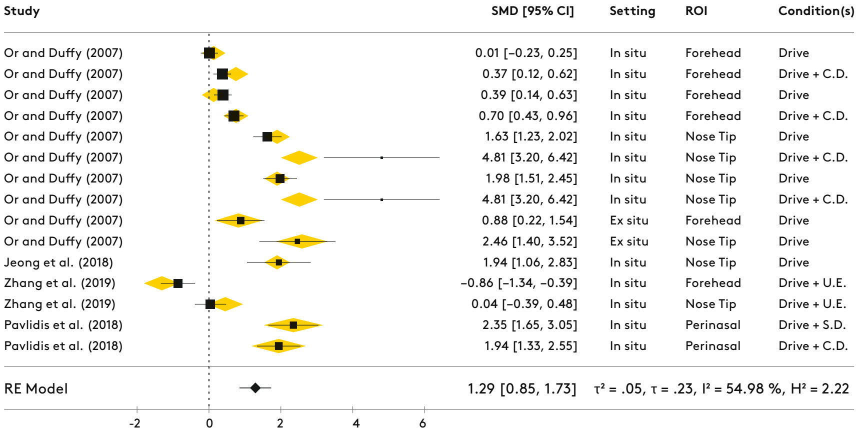
Lastly, Wu et al. [247] utilize the Supervised Descent Method by Xiong and De la Torre [250] to produce a point-cloud of facial features and calculate 3D coordinates for each facial feature through vector calculation, thus determining the head position and orientation in three-dimensional space.

4.4 Results of the Meta-Analysis

To determine the validity of the thermal camera as a tool for assessing cognitive load, we conducted a meta-analysis. We were interested in whether the null hypothesis of no effect could be rejected. From the 13 studies that we analyzed in-depth, four were suitable to be coded using the Population, Intervention, Comparison group, Outcomes, Time frame, Setting (PICOTS) framework [183]. However, despite the limited amount of records available, we extracted 15 data points containing Standardized Mean-Difference (SMD) and Confidence Intervals (CI). Using this data, we ran a Random-Effects analysis using the R METAFOR package[1] [230], carried out moderator and Meta-Regression analyses, and created Forest and Funnel plots (Figure 4.3 and 4.4). We used a Random-Effects model since it does not assume that the estimated effects arise from a single homogeneous population; instead, that the true effect sizes differ between each study.

---

[1] https://metafor-project.org/doku.php



Study	SMD [95% CI]	Setting	ROI	Condition(s)
Or and Duffy (2007)	0.01 [−0.23, 0.25]	In situ	Forehead	Drive
Or and Duffy (2007)	0.37 [0.12, 0.62]	In situ	Forehead	Drive + C.D.
Or and Duffy (2007)	0.39 [0.14, 0.63]	In situ	Forehead	Drive
Or and Duffy (2007)	0.70 [0.43, 0.96]	In situ	Forehead	Drive + C.D.
Or and Duffy (2007)	1.63 [1.23, 2.02]	In situ	Nose Tip	Drive
Or and Duffy (2007)	4.81 [3.20, 6.42]	In situ	Nose Tip	Drive + C.D.
Or and Duffy (2007)	1.98 [1.51, 2.45]	In situ	Nose Tip	Drive
Or and Duffy (2007)	4.81 [3.20, 6.42]	In situ	Nose Tip	Drive + C.D.
Or and Duffy (2007)	0.88 [0.22, 1.54]	Ex situ	Forehead	Drive
Or and Duffy (2007)	2.46 [1.40, 3.52]	Ex situ	Nose Tip	Drive
Jeong et al. (2018)	1.94 [1.06, 2.83]	In situ	Nose Tip	Drive
Zhang et al. (2019)	−0.86 [−1.34, −0.39]	In situ	Forehead	Drive + U.E.
Zhang et al. (2019)	0.04 [−0.39, 0.48]	In situ	Nose Tip	Drive + U.E.
Pavlidis et al. (2018)	2.35 [1.65, 3.05]	In situ	Perinasal	Drive + S.D.
Pavlidis et al. (2018)	1.94 [1.33, 2.55]	In situ	Perinasal	Drive + C.D.
RE Model	1.29 [0.85, 1.73]	$\tau^2 = .05$, $\tau = .23$, $I^2 = 54.98$ %, $H^2 = 2.22$		

Figure 4.3: The Forest Plot shows the Standardized Mean-Difference (SMD) and the Confidence Intervals (CI) of the studies taken into consideration. The last line shows the size, with its CI, of the Random Effects model. The Forest Plot also includes yellow "diamonds" for each study. These are the predicted effect sizes computed from the meta-regression model. The column *Setting* indicates where the study was conducted (in-situ or ex-situ), the *RoIs* where the thermophysiological data was extracted (forehead, nose tip, and perinasal). In column *Conditions*, C.D. stands for Cognitive Distraction, U.E. for Unexpected Event, and S.D. for Sensorimotor Distraction. To understand the output analysis, the SMD value is interpreted similar to the effect size (*d*) defined as: *d* (.01) = very small, *d* (.2) = small, *d* (.5) = medium, *d* (.8) = large, *d* (1.2) = very large [192]. Moreover, the CI indicates the precision with which we can calculate the SMD, facilitating the comparison between the studies.

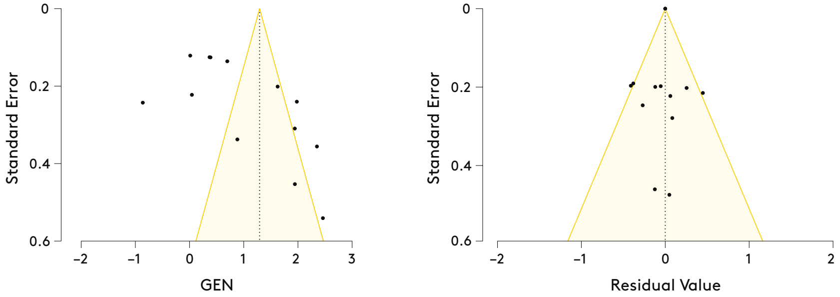
The Omnibus test of the model coefficients shows to be significant (Q = 32.96, DF = 1, $p < .001$). Also, the test of Residual Heterogeneity shows to be significant (Q = 248.46 DF = 14, $p < .001$). The significance of both tests indicates a rejection of the null hypothesis of no effect and the hypothesis of homogeneity. The prior implies that the model (without explanatory variables such as study setting, the RoIs, and conditions) is significant. The latter implies that thermophysiological results differed across studies (heterogeneity of the effect sizes). To explain the excess of variance among the studies, we carried out a meta-regression that includes factors as RoIs where the thermophysiological data was extracted (e.g., forehead, nose tip, and perinasal region), the study setting, i.e., if the study was carried out in-situ (laboratory via driving simulator) or ex-situ (real environment). Moreover, we included study design factors as

					95% C. I.	
	Estimate	**SE**	**z**	**p**	**Lower**	**Upper**
Intercept	0.822	0.343	2.400	0.016	0.151	1.493
Setting (In situ)	-0.690	0.363	-1.899	0.058	-1.403	0.022
ROIs (Nose Tip)	1.764	0.212	8.310	$< .001$	1.348	2.180
ROIs (Perinasal)	1.191	0.426	2.798	0.005	0.357	2.025
Condition(s) (Drive + C.D.)	0.618	0.238	2.598	0.009	0.152	1.085
Condition(s) (Drive + U.E.)	-1.443	0.269	-5.354	$< .001$	-1.971	-0.915
Condition(s) (Drive + S.D.)	1.026	0.621	1.654	0.098	-0.190	2.243

Table 4.3: The table shows the estimated coefficients (via Wald test) for the Random-Effects analysis, along with their significance. Here, next to the intercept, one coefficient for Setting (in situ), two for ROIs (Nose Tip and Perinasal), and three for Conditions: Drive + Cognitive Distraction (C.D.), Unexpected Event (U.E.), and Sensorimotor Distraction (S.D.). Note: SE stands for Standard Error.

a type of distraction added to the condition (e.g., cognitive or sensorimotor distraction). Despite the specification cited, the Omnibus test of the model coefficients shows to be significant (Q = 121,26, DF = 6, p < .001), indicating only two coefficients to be non-significant: *Drive + Sensorimotor Distraction* and *In-situ setting* (Table 4.3). The test of Residual Heterogeneity also shows to be significant (Q = 34,42 DF = 8, p < .001). The test indicates that although the moderators explain the differences between the studies, there is an unexplained excess of variance. Furthermore, we conducted an analysis to assess publication biases. The Rank correlation test for funnel plot asymmetry (Kendall's τ = .42, p = .03) shows minor signs of asymmetry (Figure 4.4a), indicating a bias of effect sizes included in our meta-analysis [17]. Moreover, the Funnel Plot (visually) confirms an excess of heterogeneity due to the few data points (> 5%) that lie outside the confidence triangle.

Despite the variance of the observed effect size, overall (as can be seen in the Forest Plot Figure 4.3), the meta-analysis suggests a significant positive effect in detecting cognitive load using thermal cameras, SMD = 1.29, p < .01, 95% CI [0.85, 1.72]. This is a promising result; however, the low number of studies pulled into the meta-analysis (due to inadequate design or diagnostic accuracy of the other nine studies) leads us to conclude that the results obtained must be viewed carefully and require a systematic investigation.



(a) **Funnel Plot without explanatory** (b) **Funnel Plot with explanatory**

Figure 4.4: (a) The Funnel Plot of the Random-Effects analysis without explanatory variables, and (b) the Funnel Plot with explanatory variables (i.e., study setting, RoIs, and conditions). In (a), the data points are expected to lay in a confidence triangle 95% of the time. However, this is not the case; an excessive amount of points are outside the interval due to heterogeneity. In (b), the Funnel Plot of the meta-regression model is shown. Herewith, we see that the plot is symmetrical in the vertical axis around the meta-analytic compounded effect size estimate. Suggesting that the moderators (i.e., study setting, RoIs, and conditions) do explain the differences between the studies; however, there is some excess of variance (> 5%).

4.5 Discussion

This chapter provides a first comprehensive understanding of the usage of thermal imaging in driving simulators. Despite the methodological limitations found during the inclusion of the studies, our outcomes show that infrared thermal imaging has a positive effect in detecting cognitive distraction in control settings. However, some moderate inconsistencies between studies ($I^2 = 54.98\%$) suggest that the results differ by more than would be expected, reducing the confidence that we might have of the results. These differences might have been generated by several factors, primarily due to inadequate study design, e.g., how the collected data was gathered or which variables were selected [190].

Overall, the thermal camera appears to be a valuable addition to the automotive researcher's toolbox to assess the driver's physiological effects. When additional physiological parameters were observed, we noticed they always corroborate the data collected through infrared thermal imaging, making infrared thermal imaging a valuable alternative to a classical physiological sensor technology, as no physical sensors have to be placed on the participant's body. This may increase the participant's sense of freedom, as sensors measuring like ECG or GSR can be constricting. Thermal cameras also have the advantage of reduced experiment set-up time, as experiments need to be set-up only once by placing the camera instead of placing sensors for every participant. Moreover, it may render unnecessary the technical know-how needed to place these sensors on the subject. Furthermore, thermography is already being utilized to provide real-time information of cognitive workload to the driver.

Given a thorough and founded base knowledge on how participants face temperatures react to driving situations, we feel that infrared thermal imaging is a good alternative to classical physiological sensing methods. However, despite these first positive results, the limited finding requires further investigation and replication studies.

Given the high excess of variance that we observed in our statistical meta-analysis, some unknown effects on study efficiency remain. Here we identify a need for additional detail in study design in future work. We suggest developing a clear and detailed framework on how infrared thermal imaging should be done, and we propose a methodology based on the focus areas we formalized in this chapter. Regions of Interest must be defined, as well as the method of determining these regions. Data extraction from the color image must be described, and noise reduction or further data processing explained with methodology and parameters clearly stated.

Future researchers should also consider the participant variance and group size in their studies. Furthermore, we suggest including more gender diversity when recruiting subjects. The study participants' gender is skewed towards males (Male N = 127, females N = 68) and none from others genders. Additional data on participants' traits, such as affinity towards anxiety or personality profiles, should also factor into the analysis. Indeed, prior research has shown that stress can affect sympathetic responses, thereby influencing the results [210], as do different personality types on specific driving behaviors [176]. The addition of these elements in the data collection process could help identify additional driver effects or effects on the thermal image and provide a more accurate thermography evaluation as a physiological measurement tool.

Despite limiting our analysis to the field of automotive research, our results provide insights to the entire field of cognitive workload research using thermal imaging. Identifying further effects on measurement efficiency is only possible when detailed descriptions of data capture and data processing methodologies are provided. We provide an approach of comparison between research in thermal imaging to improve and make readily accessible standardized and robust methodologies in future studies.

4.6 Summary

The chapter presents a method for measuring cognitive workload in automotive research using infrared thermal imaging cameras. It discusses the limitations of traditional methods for measuring cognitive workload and highlights the potential of thermal imaging in providing a real-time, continuous, and non-intrusive measurement of physiological parameters. The chapter also presents a systematic literature review and meta-analysis of thermal imaging in driving simulators, showing its effectiveness in measuring cognitive workload and its potential for future research in the field of HCI.

Toward Desiging an Effective Teleoperator Workplace

5

Subject Matter Expert Analysis

We complement the theoretical approach described in Section 2.4, i.e., with the collection of requirements inferred from the literature review, with an explorative phase in which we gather first-hand user requirements. Inspired by the requirement analysis of related research fields, in this chapter, we present a comprehensive SA requirements framework and analysis of AV teleoperation-based interfaces.

This chapter provides a detailed exploration of the following publication:

Gaetano Graf and Heinrich Hussmann. “User Requirements for Remote Teleoperation-Based Interfaces.” In: AutomotiveUI ’20 (2020), pp. 85–88. DOI: `10.1145/3409251.3411730`

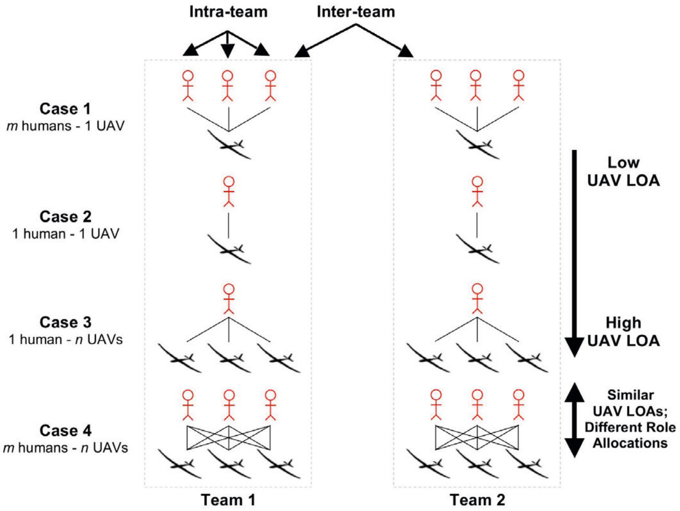
We begin with an overview of prior requirement analysis, and then we present the results of two studies. For these, we employed two methodologies, in-depth interviews and traditional statistical analysis. The in-depth interview enabled us to collect qualitative user requirements, and the statistical analysis helped us answer the hypothesis if there was a statistical preference over the requirements collected. This chapter sought to adapt and extend prior research of user requirements for unmanned vehicles. Also, to support the development of HMI by helping the operator achieve and maintain high SA levels.

5.1 Related Work

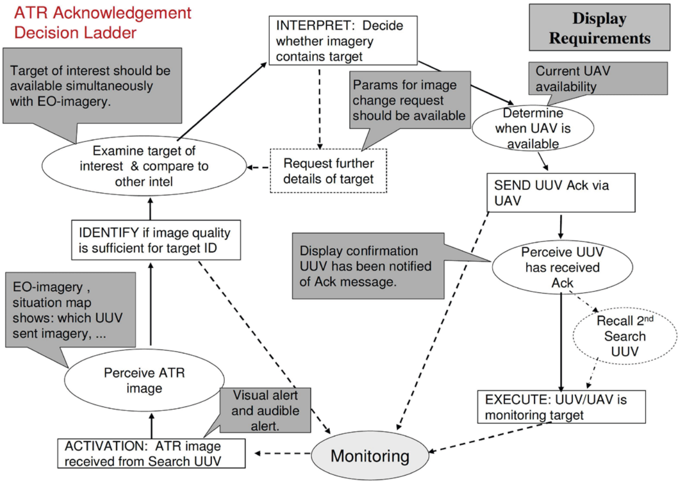
In recent years, many requirement studies in HCI were conducted to target and support the operator's SA, particularly, research in Unmanned Aerial Vehicle [27, 54], Unmanned Underwater Vehicle [157, 198], as well as in Unmanned Ground Vehicle [59, 182].

For example, Drury and Scott [54] have proposed HMI assets for Unmanned Aerial Vehicle (UAV) for single and multiple human operators. The peculiarity of their analysis lies in the fact that the authors model the requirements collected with the levels of automation of the remote vehicles. As levels of automation of the vehicle increase, the level of information required by an operator decreases. In the most basic case, i.e., when one operator controls one vehicle, the framework is constructed upon three blocks. The first block describes Human-to-UAV awareness. The second represents UAV-to-Human awareness, and the third block the general awareness.

First, the Human-to-UAV awareness unfolds the understanding that an operator has of the vehicle. This level includes the spatial relationships (e.g., geographical coordinates and speed obstacles or targets), the vehicle capabilities (e.g., sensors or communication status), the health of the vehicle (e.g., charge or fuel level), other non-health statuses (e.g., sensors in use), weather condition (current and predicted), and least, the certainty of the components (e.g., the probability of an anticipated event). The second block refers to UAV-to-Human awareness, i.e., what the vehicle needs to know from the operator. The level includes the operator's commands (e.g., where to go) and the operator's given constraints (e.g., preprogrammed fail-safe). The third and last block describes the general awareness, i.e., the overall mission awareness.



(a) **Human-UAV Control Strategy**



(b) **Decision Ladder**

Figure 5.1: (a) The Human-UAV control strategy developed by Drury and Scott [54]. Here, the authors describe the requirements with the Levels of Automation of the remote vehicles. (b) The decision ladder. That is, to generate and collect user requirements, Nehme et al. [157] proposed a decision ladder that intends to replicate the decision-making processes of the human operator.

This level includes task purpose (e.g., monitoring), customers and other stakeholders (e.g., who requested the task and who has the interest), task progress, time constraints, and related missions (e.g., relevant information of other vehicles).

Upon having defined the base case (one human, one vehicle), Drury and Scott [54] expand the framework assuming that multiple operators simultaneously control various vehicles. Here, the authors updated the framework with two additional blocks. These are Human-to-Human awareness and Vehicle-to-Vehicle awareness. The prior refers to the understanding that one operator has of another operator or another team, for instance, when locations, intentions, and additional information must be shared. Similarly, the latter, the Vehicle-to-Vehicle block, refers to the knowledge that one vehicle has over the others, e.g., about the commands, tasks, or planes it has.

Additional requirements for UAV and UUV have been provided by Nehme et al. [157]. To gather user requirements, the authors adopt a different approach in their research. Their framework is based on the assumption that the design of novel HMI concepts does not involve domain experts, a common problem when collecting user requirements [39, 198]. To solve this issue, the authors propose a decision ladder that should compensate for the lack of understanding that experts have over the subject. The decision ladder is built upon three stages that replicate the human operator's decision-making processes. The first stage involves the generation of a scenario, the second stage the generation of the related event flow diagram, and the third the generation of SA requirements. The result is the decision ladder from which information and display requirements are obtained.

Differently, Riley and Endsley [182] present a collection of user requirements for UGV obtained from field observations elicited during robot-assisted rescue training sessions. In particular, the requirements are the result of three kinds of observations: direct observation, post-task interview sessions, and review of the rescue videos. Although the requirements obtained are based on a specific task, i.e., depending on the interviewee's assignment, the observations reported in this research reveal common features that can eventually be generalized and applied in similar contexts, as for AV teleoperation. For instance, in their paper, the authors highlight difficulties in localizing the vehicle, challenges in understanding the remote situation, difficulties in interpreting the information, and difficulties for individuals/teams to share SA.

Although prior works offered an extensive literature review of how to mitigate low remote SA and human-performance issues [31, 216], little has been done to collect first-hand operator's requirements for AV teleoperation-based interfaces. Informed by prior requirement analysis of UAV [27, 54], UUV [157, 198], as well as of UGV [59, 182], in this chapter our research aims to address the following overarching research question:

What are the essential user requirements for vehicle teleoperation-based interfaces that support high levels of Situation Awareness, and how should these requirements be effectively presented to operators?

Building on the foundations laid by Nehme et al. [157], who employed a decision ladder approach to gather user requirements, and Riley and Endsley [182], who derived requirements from field observations during UGV robot-assisted rescue training sessions, our research adopts a comprehensive approach.

5.2 Requirement Assessment

In alignment with the methodology employed in the studies discussed above, i.e., focusing on the design of a SA-oriented teleoperator workspace, we adopted the Goal-Directed Task Analysis (GDTA) method. As defined by Endsley [59], GDTA serves as a cognitive task analysis tool, particularly valuable in the early design phase. Our implementation of GDTA concentrated on understanding the tasks that teleoperators undertake to achieve specific goals.

To gather insights into the cognitive processes and considerations involved, we opted for semi-structured interviews, providing a degree of flexibility during the discussions. This approach allowed us to adapt the interview format based on participants' responses, reordering questions and exploring intriguing ideas that emerged. This nuanced exploration proved instrumental in comprehending participants' attitudes, preferences, concerns, practices, and ergonomic needs related to the safe control of remote AVs.

Figure 5.2: A screenshot from the driver perspective: the image shows the starting point from which the participants have begun the user study.

Thus, the interview was conducted in a driving simulator study lab that extended our investigation beyond the simple interview setting.. This holistic approach, informed by the principles of GDTA, enabled us to delve deeper into the practical aspects of achieving SA and ergonomic control in the context of remote AV operations.

5.2.1 Procedure and Tasks

In the first study, participants were asked to imagine a situation in which an AV required the operator assistance and they needed to drive the vehicle to a safe position. Before starting the driving task, when the screen was off, we asked the subjects what they expected to see and what they would expect to happen. Then, having turned the display on, the participant received "live" video images of the remote vehicle (Figure 5.2). The video images depicted the remote situation from the driver's perspective showing three overlapping camera streams covering a horizontal field of view of 180 degrees. Upon turning on the screen, we asked the subject to describe what they saw and which functions they would like to have to conduct the AV safely. Then the driving task began, asking the participant to follow the turn-by-turn navigation instructions. Having completed the assignment, we conducted a post-task interview, which encouraged the subject to elaborate on their thoughts and concerns about their experience.

5.2.2 Participants

The interviews were conducted with IT professionals working in the automotive industry. We recruited 18 participants (4 female) ranging from 22 to 35 years

(M = 28, SD = 3.66). On average, participants held a driving license for about ten years (SD = 3.67), and more than half of the participants (N = 10) already had some experience with remote controlling, e.g., of drones. Lastly, none of the participants had experience in remote controlling AVs.

5.2.3 Apparatus and Materials

Our experimental setup consisted of one Display of 48.9-inch Ultra Wide (3840 × 1080 Pixel). We installed a steering wheel and pedals (Logitech G29) to control the remote vehicle. The steering wheel provided force feedback and a 900-degree rotation. Finally, one computer enabled the communication of input and output signals between the simulator and the automotive control elements (i.e., steering wheel and pedals).

5.2.4 Results

We analyzed the data and removed redundant statements, obtaining a total of 80 requirements (Figure 5.3, and Figure 5.4). Each requirement was sorted into a category, yielding a general theme that describes the category from the participant's perspective. The requirements provide information not only about the vehicle but also about the surrounding environment. Covering all three macro levels of SA: 1) perception of elements in the environment, 2) comprehension of the current situation, and 3) future status projection.

Vehicle position: The vehicle position describes how and where the AV is located. Thus, necessary primary information includes: 360° remote view, vehicle heading, location, steering wheel orientation angle, and wheels orientation. In this regard, the vehicle speed is also taken into account.

Vehicle status: Other important information on the AV status may include the motor/battery state, i.e., vehicle charge/fuel level, motor/battery temperature, and motor oil level. We also entail information about other vehicle damage, e.g., tire, light, and similar. Moreover, last vehicle inspection and overall vehicle damage, e.g., on the car body.

Vehicle issues: The teleoperator may be concerned about important information regarding the past, present, and near-future vehicle issues, including location and actions, unexpected technical issues, unexpected traffic situations, unforeseen customer issues, maintenance routines. This information may allow the teleoperator to build a quick understanding of the situation.

Vehicle characteristics: With vehicle characteristics, we refer to the actual vehicle size, i.e., length, width, and height, including the AV weight. Also, besides the vehicle specific weight, information on the vehicle cargo state, i.e., load/unload and type of cargo, may be essential for the operator to adjust their driving behavior to the situation.

Vehicle operations: To anticipate and support the teleoperator in understanding the future situation, the projected destination of the vehicle (in the short and long term), as well as the projected vehicle stop position, is shown. Also, potential longitudinal and lateral collision warnings may be forward to the teleoperator. Additionally, the system may suggest possible locations where to place the vehicle, i.e., to pull-over the AV safely. Alternatively, the system may recommend turning maneuvers and similar control actions. All this information should not neglect the communication delay from and to the AV.

Task objectives: The teleoperator might need to know how long the operation is taking, its impact on the customer plane, and any time constraints. For instance, when the teleoperator has to work under time limits, it might help to know the projected time to task completion, following tasks, and priority. The total number of assignments that must be completed should also be displayed. Last, we include the projected probability to complete the tasks or the projected need to reject the tasks.

On-board sensor: Herewith, we include functionality and state (i.e., on/off) of on-board sensors such as Light Detecting and Ranging (LiDAR), long- mid- and short-range radars, as well as cameras. Cameras may hold further information as orientation and zoom level. Besides the sensors functionality and status, additional detailed information such as the number of sensors and the type of sensors detecting the obstacle should be available. Perhaps, it might be helpful for expert teleoperators to know the results of the past detection.

Communication state: Information about the quality and condition of the communication connection from and to the AV must be considered. That is bandwidth from the vehicle to operator and vice versa, the available bandwidth, bandwidth requirements, current signal strength, signal future trends, a list of mobile providers, and the current mobile provider in use.

Objects, obstacles: The teleoperator, to understand the spatial relationship between the AV and the obstructions around, might need to know the distance and the obstacles location. Other details may include the object characteristics such as material (i.e., solid or liquid) and size (i.e., length, width, depth, height).

Environmental information: Environmental information entails essential details on the surrounded environment. It may include information about the speed limit ruling the area and other traffic signaling and infrastructure information, e.g., traffic lights. Further information may include the location of other fleet-related AVs assigned to the area.

Weather conditions: Current and predicted weather conditions are shown to the teleoperator. Poor weather conditions such as heavy precipitation, winds, and icing will affect operative performance. The teleoperator must consider both current and predicted weather information to adapt their driving behavior.

Terrain features: Further spatial information about the terrain features might enhance the teleoperator's understanding of how and where they should maneuver the vehicle. We include features such as landmarks/barriers, clutter, debris, and uneven terrain in this category.

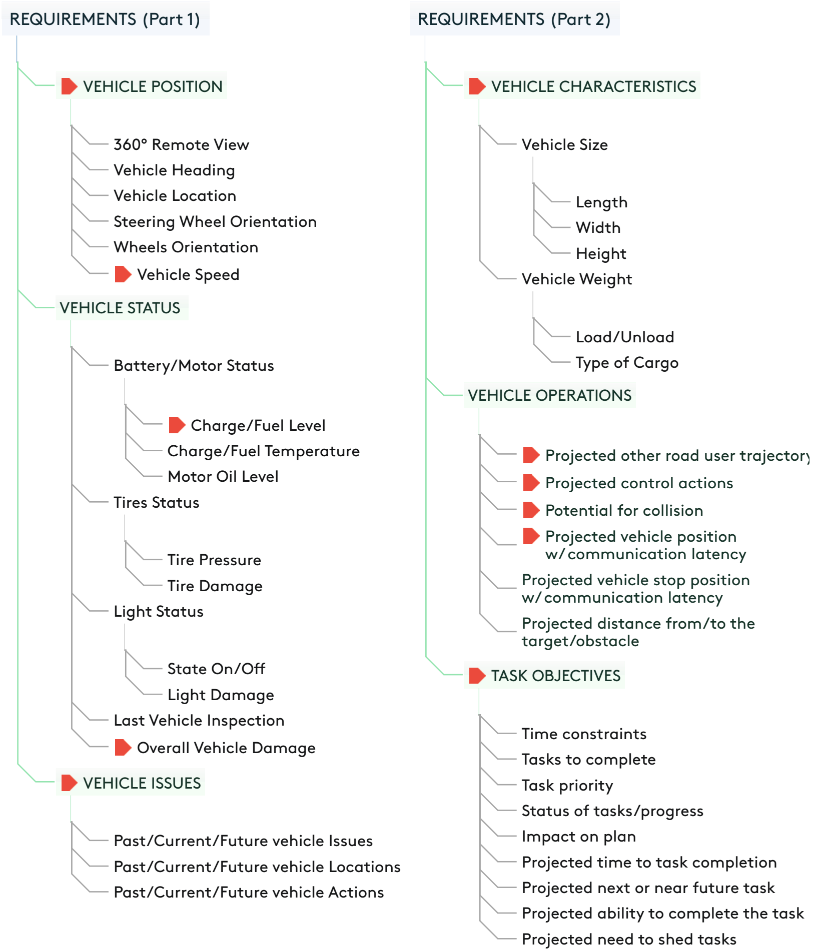


Figure 5.3: A collection of 80 users' requirements was collected in the first interview. Flagged in red are the requirements that the participants in the second interview have chosen.

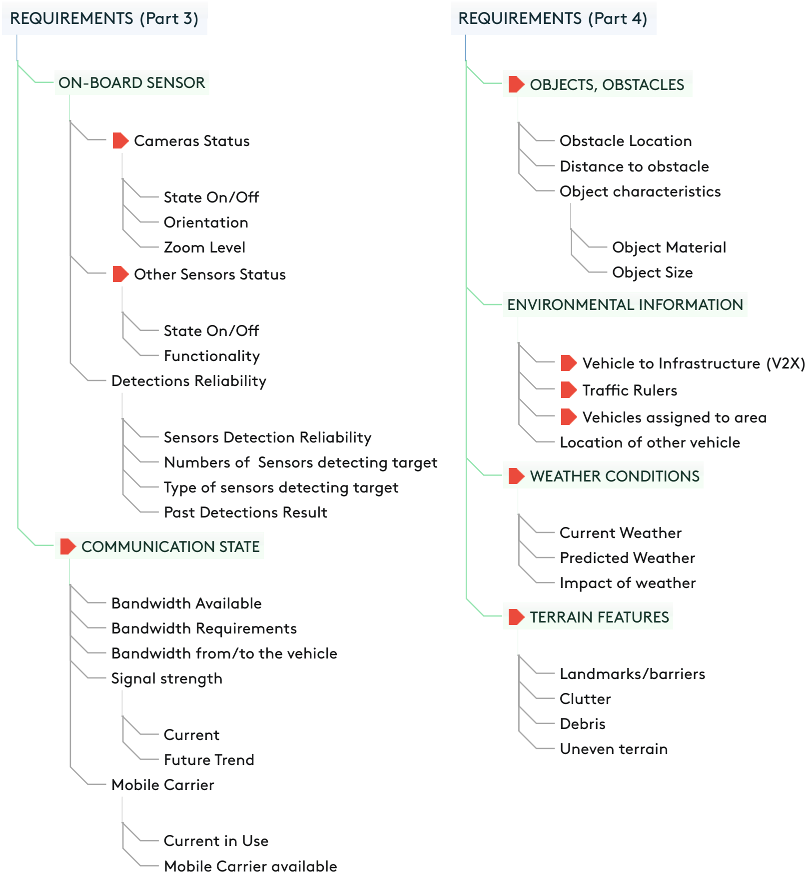


Figure 5.4: In some cases, a requirement category (e.g., vehicle position) has been selected in its entirety. While in other cases, only one category requirement was selected (e.g., vehicle speed).

5.3 Requirements Clustering and Rating

After analyzing the results obtained from the GDTA, our focus shifted towards exploring potential configurations for a teleoperator workstation. In both academic and industrial settings, a variety of workstations exist, broadly classified into three categories [224].

The first category involves workstations where teleoperators have a clear line of sight over their displays, with controls and monitors positioned below their horizontal sightline. This setup, historically featuring tilted displays and movable or fixed controllers like keyboards [224], is advantageous when multiple operators share additional displays for efficient information retrieval.

Building on the first category, the second introduces a vertical extension, incorporating multiple displays covering most of the operator's vertical field of view (cf. Figure 5.5). In Tilley [224], the control commands are positioned below the sightline for easy access, while non-manual controls, like auxiliary panels, are located above. Eventually, the third category expands horizontally, enveloping operators with additional displays and controllers.

Drawing inspiration from Lockwood et al. [126] and Tilley [224], our envisioned workstation is a compound (sit or stand) configuration with two display areas (Figure 5.5). The primary vertical display area provides essential visual information, while a secondary horizontal area serves as both input and output devices. This allows operators to receive visual and haptic feedback and send touch and control commands. Consequently, remote vehicle assessment and control can be accomplished through touch gestures or more complex maneuvers using direct steering controllers, such as joysticks.

5.3.1 Procedure and Tasks

The second study was designed to ascertain the importance and preferences of the requirements previously collected. Thus, participants were asked to cluster and rate a collection of 80 requirements (Figure 5.3, and Figure 5.4). We began the interviews by describing the logic of teleoperation and the theoretical workstation (Figure 5.5). In particular, by bringing to their attention that the workstation would have two display devices (cf. Section 5.3).

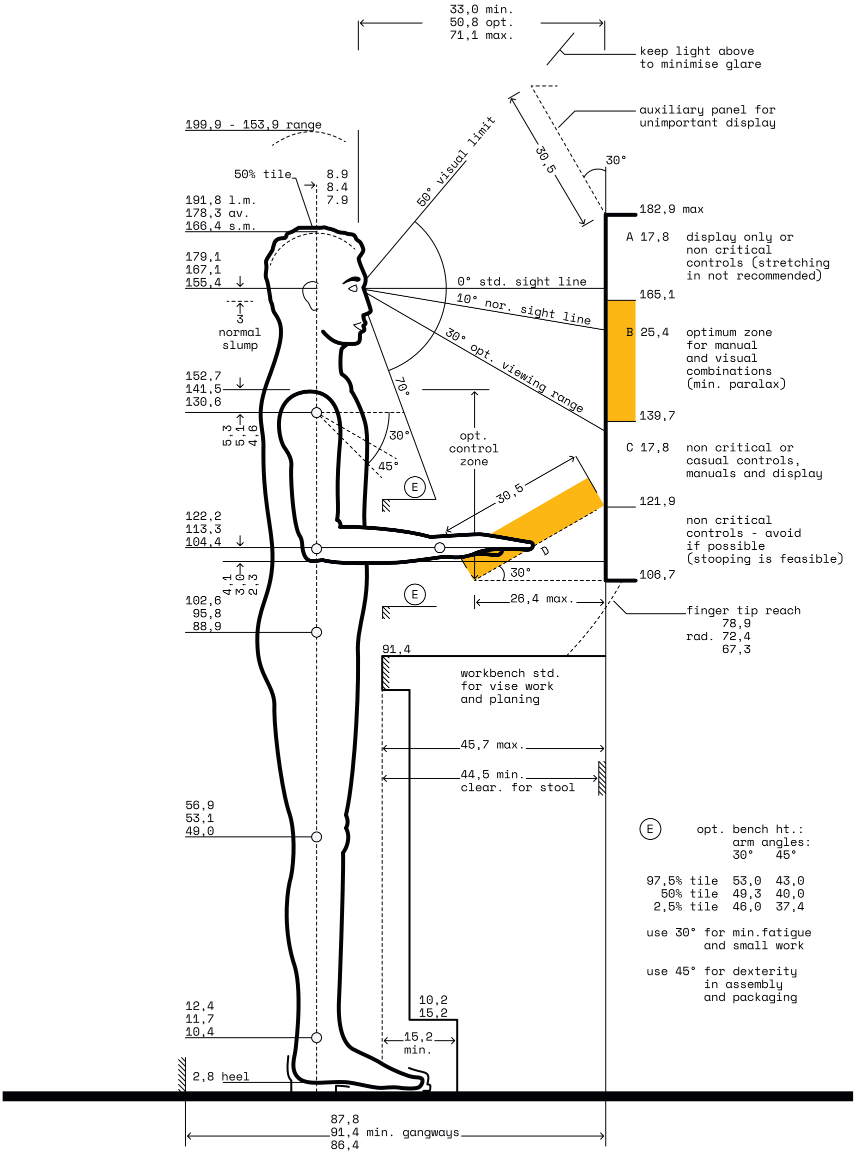


Figure 5.5: An adapted illustration of Tilley's [224] anthropocentric data for adult males at the workstation. In yellow, we highlighted the areas where we envision the displays to be, i.e., a primary display arranged vertically and a second horizontally, both in front of the operator.

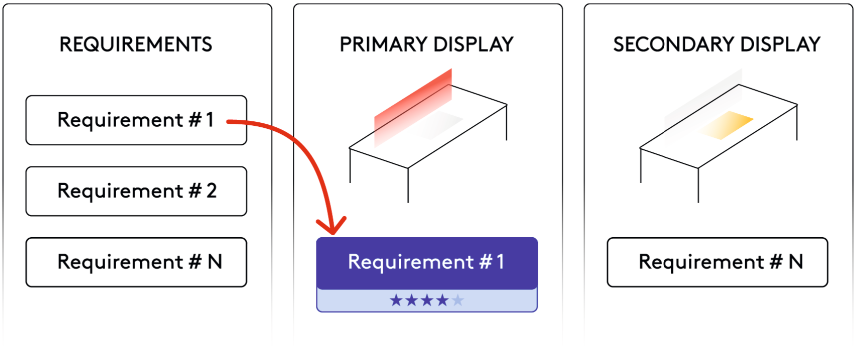
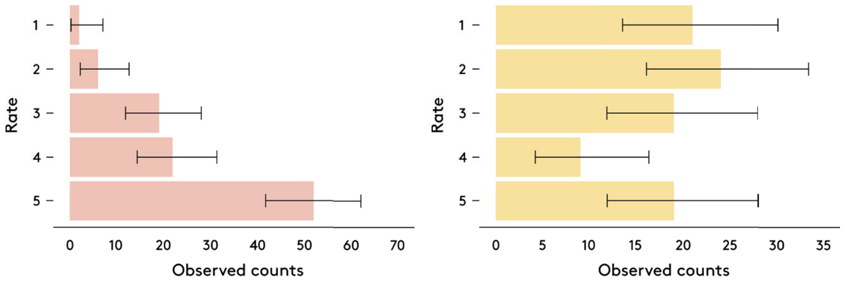


Figure 5.6: In this illustration, the activities for the second interview session are shown. The first activity was the requirement sorting. Each participant was asked to choose from the same set of requirements the information they thought they needed to conduct an AV safely (Figure 5.3, and Figure 5.4), and decide which requirements they would like to display on a primary or secondary display device. The second activity was the rating task. Each participant rated the chosen requirement on a 5-point Likert scale (1 = “unimportant” and 5 = “important”).

Then the participants were presented with a scenario. The scenario showed an AV blocked by an obstacle in a four-way intersection, and around there were cars and pedestrians. Then, participants were asked to complete two activities. The first activity was card sorting (Figure 5.6). Each participant was asked to choose from the same set of requirements, in Figure 5.3 and Figure 5.4, the information they thought they would need to conduct an AV safely and decide which requirements they would like to have on a primary or secondary display device. Card sorting is a technique that requires participants to sort certain pieces of information into various categories to help to understand what they think about content and categories [209]. The second activity was the rating task. Each participant rated the chosen requirement on a 5-point Likert scale (1 “unimportant” and 5 “important”). During the whole card sorting activity, experts were required to verbalize their thinking to know why they needed specific information.



(a) **Primary Display Descriptive Plot** (b) **Secondary Display Descriptive Plot**

Figure 5.7: The descriptive plot displays the reported count frequency and the corresponding confidence intervals (set to 95%) for every rate. (a) The descriptive plot for the primary and (b) for the secondary display devices in shown.

5.3.2 Participants

We conducted a second interview with novel and professionals IT users working in the automotive industry. We recruited 10 participants (2 male) ranging from 22 to 36 years (M = 28, SD = 4.23). On average, participants held a driving license for about nine years (SD = 4.17), and almost all the participants (N = 8) had already remotely controlled drones and similar vehicles. However, none of the subjects have relevant experience in remote controlling AVs.

5.3.3 Apparatus and Materials

Interviews were conducted in 2020, however due to external situation the whole study had to be executed online. We designed the requirements clustering and the rating activity using Trello[1], an online tool that facilitates the organization and rating of lists and cards. The communication with the subjects was facilitated by Zoom[2], a video and chat service.

5.3.4 Results

In these interviews, we investigated the participants' reported ratings of the requirements. Participants selected a total of 20 requirements. In some cases, a

---

[1] https://trello.com/

[2] https://zoom.us/

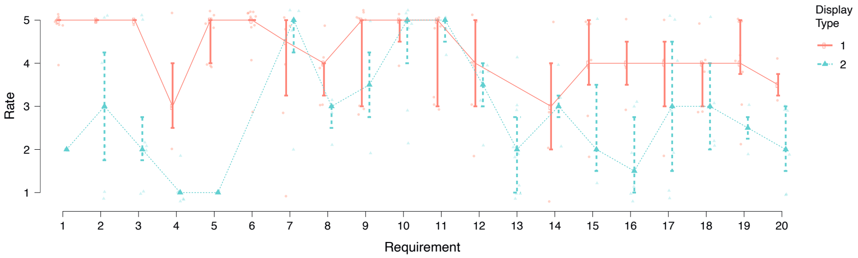


Figure 5.8: In this plot, on the y-axis, the participant rating and the requirements are on the x-axis. Each requirement interval (categorical predictors) is shown with a vertical line and displays a dot or a triangle for the median. Thereupon: (1) 360° Remote view, (2) Charge/Fuel level, (3) Communication state, (4) Vehicle assigned to an area, (5) Vehicle to infrastructure, (6) Objects/Obstacles, (7) Vehicle issues, (8) Projected other road user trajectory, (9) Potential for collision, (10) Other sensors status, (11) Vehicle speed, (12) Projected control actions, (13) Task objectives, (14) Terrain features, (15) Traffic rules, (16) Vehicle characteristics, (17) Overall vehicle damage, (18) Vehicle position, (19) Projected vehicle position considering communication latency, (20) Weather condition. On the top left, the display type indicates the requirements selected for the primary and secondary display devices.

requirement category, as *Vehicle Position* was chosen in its entirety. In other instances, only one requirement of the whole category was selected, e.g., as *Vehicle Speed* (cf. Figure 5.3). We conducted a multinomial test to reject the hypothesis that the reported rate frequency was equally distributed over the requirements. Figure 5.8 offers a graphical representation of the data. The Chi-square was used to test whether the pattern of the requirements rating differed from randomness.

Yet, participant's ratings differed significantly across the requirements χ^2 (4, N = 193) = 36.92, p < .001, Cramer's V = 0.44. The analysis shows that the ratings of the requirements selected for the primary device (Figure 5.7a), significantly differ χ^2 (4, N = 101) = 76.67, p < .001, Cramer's V = 0.93. Whereas, the ratings of the requirements selected for the secondary device (Figure 5.7b), did not statistically differ χ^2 (4, N = 92) = 6.91, p = 0.141, Cramer's V = 0.27. Participants did not have much preference over a specific rating for the requirement selected for the secondary device. Last, we tested the relationship between the variables, i.e., rating and display devices. The contingency table shows that there is a significant relationship between the requirements and the rating χ^2 (76, N = 193) = 112.97, p = .004, Cramer's V = 0.38, as well as between the requirements and the display devices χ^2 (19, N = 193) = 60.08, p < .001, Cramer's V = 0.56.

5.4 Discussion

The research endeavors to present a comprehensive SA requirements framework and analysis for AV teleoperation-based interfaces. Employing a combination of qualitative and quantitative research methods, we sought to understand the user's perspective, resulting in the identification of 80 requirements clustered across 12 categories. These requirements might play a pivotal role in maintaining a general SA, encompassing the ability to perceive, comprehend, and project all future status elements.

The detailed analysis of participant preferences, as evidenced by plots, offers valuable insights into the nuanced dynamics of user requirements in the context of AV teleoperation interfaces. Notably, the prioritization of requirements for the primary display device over the secondary device, as indicated by higher ratings and skewed distributions, underscores the significance of display allocation in enhancing user interaction and information absorption.

In this frame of references, we discerned a clear inclination towards information pertaining to the vehicle's position, remote view, vehicle speed, infrastructure details, objects/obstacles, potential collisions, and sensor status. These specific preferences highlight the multifaceted nature of user needs during teleoperation, ranging from spatial awareness to real-time information about the vehicle and its surroundings. This detailed understanding of user preferences lays a solid foundation for conceptualizing overarching requirements. At a conceptual level, it becomes evident that an effective AV teleoperation interface should prioritize and provide a comprehensive information about the vehicle's spatial context, potential obstacles, and critical system status. The emphasis on the primary display device suggests a need for a well-organized and visually intuitive presentation of this information, ensuring that operators can quickly and accurately interpret the data provided. Moreover, the plots not only reveal individual preferences but also point towards a collective trend in user expecta-

		Rate						Type		
	Item	1	2	3	4	5	Total	1	2	Total
1	360° Remote view	0	1	0	1	8	10	9	1	10
2	Charge/Fuel level	2	1	2	1	4	10	2	8	10
3	Communication state	2	4	0	0	3	9	1	8	9
4	Vehicle assigned to area	4	2	1	0	1	8	3	5	8
5	Vehicle to infrastructure	1	0	0	3	5	9	8	1	9
6	Objects/Obstacles	0	0	1	1	8	10	10	0	10
7	Vehicle issues	1	1	1	1	6	10	6	4	10
8	Projected other road user trajectory	0	1	4	3	1	9	6	3	9
9	Potential for collision	0	1	3	0	6	10	8	2	10
10	Other sensors status	0	1	1	1	7	10	3	7	10
11	Vehicle speed	0	0	3	1	6	10	7	3	10
12	Projected control actions	0	2	3	4	2	11	5	6	11
13	Task objectives	4	3	2	1	0	10	0	10	10
14	Terrain features	1	2	3	2	1	9	5	4	9
15	Traffic rules	1	2	1	2	4	10	7	3	10
16	Vehicle characteristics	3	1	3	1	1	9	3	6	9
17	Overall vehicle damage	2	2	1	2	3	10	3	7	10
18	Vehicle position	0	2	3	3	2	10	5	5	10
19	Projected vehicle position	0	2	2	3	3	10	8	2	10
20	Weather condition	2	2	4	1	0	9	2	7	9
	Total	23	30	38	31	71	193	101	92	193

Table 5.1: The contingency table shows the frequency distribution of two variables, rate and type. Here type indicates the requirements selected for the primary (Type 1) and secondary (Type 2) display devices.

tions. Recognizing this collective inclination enables us to formulate a set of general principles for designing AV teleoperation interfaces. These principles should encompass intuitive spatial representation, real-time feedback on critical vehicle parameters, and a thoughtfully organized interface that aligns with user expectations.

Eventually, our specific findings serve as building blocks for a more comprehensive understanding of user requirements, guiding the development of future teleoperation interfaces that not only meet individual preferences but also adhere to broader conceptual principles essential for ensuring effective and user-centric designs.

5.5 Summary

The chapter presents a comprehensive SA requirements framework and analysis for AV teleoperation-based interfaces. It adapts and extends prior research of user requirements for unmanned vehicles to support the development of UIs to help the operator achieve and maintain high SA levels. The results of two studies using in-depth interviews and traditional statistical analysis are presented, showing that the requirements collected focus on the vehicle and its surrounding environment. The requirements include information on vehicle position, vehicle status, vehicle issues, vehicle characteristics, vehicle operations, task objectives, onboard sensors, communication state, and objects/obstacles.

On-Boarding: Remote Takeover Request

6

Remote Takeover

Based on the Subject Matter Expert (SME) analysis described in the previous chapter, we designed and evaluated an initial interface for remote AV operation. In particular, we started the process by designing a Takeover Request (ToR) HMI. Issued by the vehicle that requires additional assistance from a human driver, the ToR delivers the prime substantial information to the teleoperators. A body of literature on shared control systems for AVs and human drivers has been attempting to lay the groundwork in building secure ToR systems. To this end, prior works studied ToR strategies considering the driver being inside the vehicle, however not so many researched the remote ToR. In the following sections, we present a ToR interface for remote operators that we evaluated over the course of one experiment.

6.1 Related Work

Teleoperation yields numerous safety benefits, yet it grapples with challenges, notably the detachment of the driver from the vehicle control loop. A comprehensive examination spanning four decades underscores the considerable hurdles faced by human operators in tandem with automated systems [63].

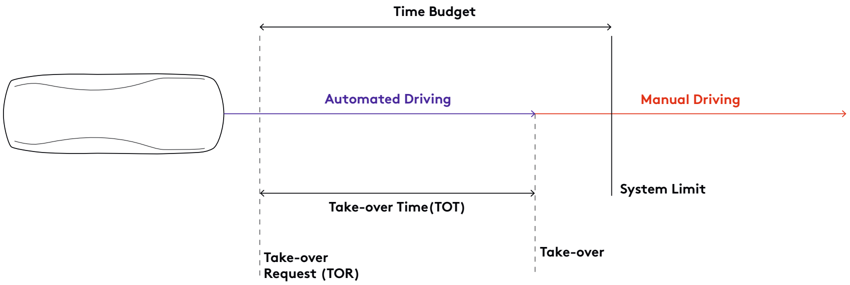


Figure 6.1: An adapted version of the Takeover Request (ToR) diagram of Gold et al. [78]. Having detected a system limit, the ToR starts from the AV that sends a ToR to a human operator and ends when the operator takes control of the vehicle. The time between the vehicle handover control and the ToR is called Takeover Time (ToT), and the time between the system limit and the ToR is called Time Budget.

The challenge is particularly evident among remote operators, as suggested in the operation of UAVs, where comprehension of the automated system's underlying plans is often elusive [185]. Empirical studies by Stockert et al. [218] and Beller et al. [14] posit that the inclusion of supplementary information, such as system uncertainty in Adaptive Cruise Control systems, holds promise for improving user interaction, trust, and acceptance. However, the persistent issue of transparency emerged in surveys of AV systems, revealing a recurrent concealment of system decisions from end-users [40].

As operator trust in system predictions decreases within low-transparency systems [57], it becomes essential to highlight the importance of trust and collaboration. This emphasis gains particular significance when considering the initiation of a ToR as depicted in Figure 6.1, where the vehicle prompts human drivers to assume control as needed [78]. Underlining the critical handover process, Flemisch et al. [67] emphasizes that both human and automated systems must align their intentions based on their perceptions for cooperative actions to transpire. Establishing this alignment not only solidifies the foundation for effective collaboration in ToR scenarios but also directly ties into the ToT dynamics.

The ToT, explored in numerous studies [76–78, 143, 231, 248], exhibits variability contingent upon factors such as driver state, traffic complexity, and

Figure 6.2: In this image, a screenshot from the nuScenes scenario we showed to the participants [19]. In particular, the AV detected an unexpected non-moving obstacle lying on its path, and since it is not allowed to drive over the solid marking lane, it sends a ToR to the teleoperator.

non-driving task attributes. In a comprehensive review by Eriksson and Stanton [65], the average stabilization time during ToT was approximately 2.47 seconds. However, ToT values span a spectrum, ranging from seconds to as high as 15 seconds in specific contexts [13, 156, 173]. Intriguingly, the presence of secondary tasks influences ToT, with participants requiring extended time for regaining control when engaged in such tasks.

While prior investigations predominantly focused on ToR within passenger cars with the driver present, a notable gap persists concerning ToR systems assuming remote operator connections. Although existing studies contribute valuable insights to the design of remote ToR HMI, they inadequately address the holistic spectrum of essential aspects. Hence, the primary objective of this chapter is to establish the foundational discourse for the discussion of our ToR HMI within an academic context.

6.2 Our Approach

Building upon the insights gained from the related work presented earlier and the exploration of teleoperation in Chapter 2, our approach stems from a comprehensive understanding of how teleoperators can effectively engage with remote vehicles.

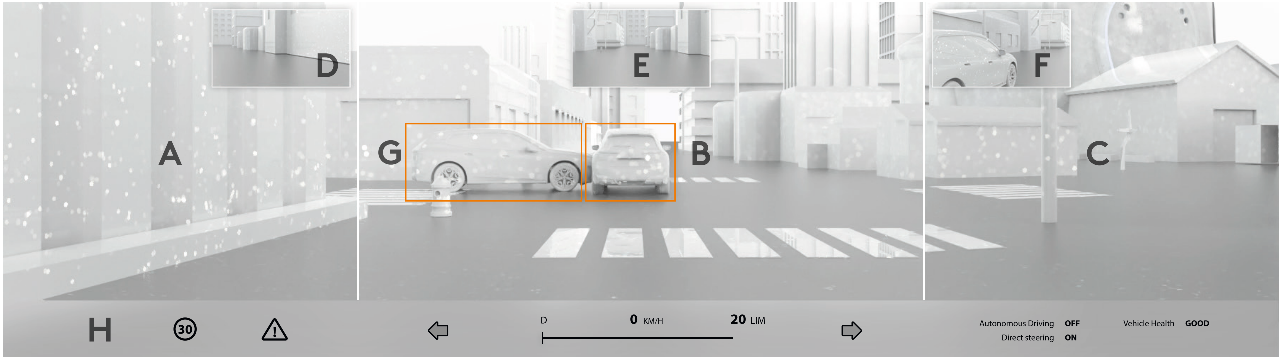


Figure 6.3: This illustration shows an egocentric 360° view of the remote vehicle. In particular, besides the front cameras (A, B, C), three additional back cameras are added (D, E, F). Other data, overlayed onto the video images, is available to the teleoperator, as the highlighted objects/obstacles of interest (G). Moreover, in a dedicated space (H), vehicles to infrastructure data may also be available to the operators, e.g., upcoming road construction or traffic light information. Here also we see status information, such as vehicle speed, vehicle charge/fuel level, communication state, i.e., the latency from and to the vehicle, other sensors status, and traffic rules signs, e.g., traffic-restricted zone or speed limit signs.

In addition to elucidating various teleoperation scenarios and use cases, we highlighted common situations, such as the detection of unexpected non-moving obstacles by the AV. In compliance with traffic regulations, when faced with obstacles on its path, the AV refrains from driving over solid lane markings and initiates a ToR to alert the teleoperator.

Yet, when designing ToR HMI for remote operations, it is essential to understand how these differ from known ToR scenarios. As stated, in a remote ToR, the teleoperator is detached from the actual traffic situation, whereas in a classic ToR, the driver is physically presented in the vehicle. Thus, in contrast to the teleoperator, the driver is aware of the vehicle history prior to the request and of the request motivation. Further factors that characterize the differences of a remote ToR are, for example, the handling of the limited field of view, camera viewpoint, orientation in a foreign environment, depth perception, and signal latency (cf. Section 2.2). These are crucial distinctions from a classic ToR that must be considered for a design of a secure takeover. As described, in Chapter 5, we let the SMEs choose and evaluate critical requirements, i.e., design assets, that they would need to secure the operation of a remote vehicle (Figure 5.3 and Figure 5.4). Eventually, SMEs divided the requirements between two display areas, i.e., among the display areas available in the workstation we envisioned (Figure 5.8 and 5.5). Thus, based on this clustering and selection, we design the HMI as follows.

6.2.1 Primary Display Design

In the vertical display in front of the operator (Figure 6.3), we deliver an egocentric 360° remote view uncovering the vehicle position. Besides the front cameras (A, B, C), three additional back cameras are added (D, E, F). The cameras D and F cover the back-left and back-right sides of the vehicle, mimicking the side mirror mounted on the vehicle, whereas camera E simulates the rear-view mirror. Other data, overlayed onto the video images, is available to the teleoperator, as the highlighted objects/obstacles of interest (G). Indirectly, teleoperators are aware of the weather condition and the terrain features from the same camera images. Lastly, in a dedicated space (H), vehicles to infrastructure data may also be available to the operators, e.g., upcoming road construction or traffic

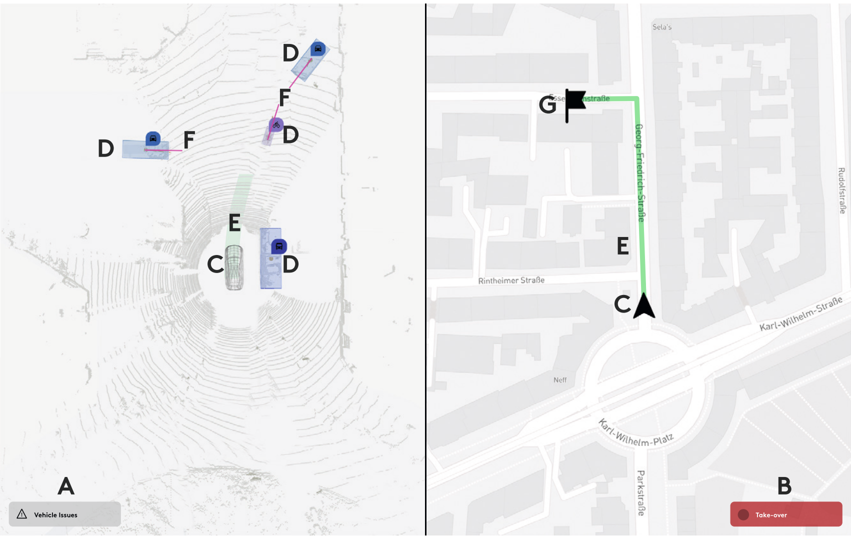


Figure 6.4: This illustration shows a 360° view of the remote vehicle. In particular, on the left, we see LiDAR point-cloud, and on the right, map-based information. At the bottom left, teleoperators, if required, can call a more detailed view of the ToR (A), and at the bottom right, operators can start and end the ToR (B). More other information is available to the teleoperator, as the objects/obstacles of interest (D), vehicle characteristics and position in relation to other objects (C), target destination (G), projected trajectory of other road users (F), e.g., of D.

light information. Here also we see status information, such as vehicle speed, vehicle charge/fuel level, communication state, i.e., the latency from and to the vehicle, other sensors status, and traffic rules signs, e.g., traffic-restricted zone or speed limit signs.

6.2.2 Secondary Display Design

The other display available to the teleoperator is positioned horizontally on the workstation, and it is touch-sensitive (Figure 6.4). Although the display size could scale according to the needs and tasks delegated to the operator, in the illustration we report here, we envision a larger screen that operators can conveniently manage. The HMI shows an exocentric viewpoint of the remote environment, on the left showing sensor-driven data (i.e., point-cloud LiDAR), and on the right map-based information. Thus, as suggested, a dual screens strategy has been applied to improve SA [97]. At the bottom left teleoperators, if required, can call a more detailed view of the ToR (A), and at the bottom right, operators can start and end the ToR (B). Further information is available to the teleoperator, such as: the objects/obstacles of interest (D), vehicle characteristics and position in relation to other objects (C), target destination (G), projected trajectory of other road users (F), e.g., of D.

6.3 User Study

We conducted a user study to assess the impact of designed interfaces on operator SA in scenarios without time constraints and secondary tasks. Measurements included participant ToT, SA level, and subjective cognitive load. The study aimed to address the research questions:

How long does it take for an operator to feel confident enough in their understanding of the remote environment to take control? At what level of SA does the operator find themselves when initiating a takeover voluntarily? Furthermore, do the HMIs impact participants' cognitive load during the takeover phase?

We answered the questions using a within-subject design [28], so that participants would go through all conditions and scenarios. The order that participants went through the conditions and scenarios followed a Latin square design [33].

6.3.1 Participants

We recruited a diverse group of 18 participants for our study, comprising 8 males. The participants had a mean age of 25 years (SD = 3.52), with an age range from 19 to 32 years. The participants' driving experience varied, with an average ownership of a driving license for six years (SD = 3.58). Among the licensed participants (N = 12), the distribution of driving frequency was as follows: two participants drove daily, two drove weekly, six drove monthly, and two rarely drove. Eight out of the 18 participants possessed experience in remote control activities, such as operating drones. Specifically, one participant had controlled a remote vehicle only once, and another participant a few times. The remaining subjects (N = 6) had several instances of remotely moving a vehicle. However, it's important to note that none of the participants had prior experience with remote takeover or remote control of autonomous vehicles.

6.3.2 Apparatus and Materials

For our user study, we carefully selected a combination of hardware to create a realistic assessment environment. The primary setup involved a local portable computer, equipped with a 48.9-inch ultra-wide display (3840 × 1080 pixels), providing participants with a comprehensive view of the remote environment.

Additionally, a Microsoft Surface Pro (5th Gen) was placed horizontally in front of each participant, featuring a 12.3-inch touch-screen display (2736 × 1824 pixels). This secondary device aimed to simulate the interface for interacting with the autonomous vehicle's HMI.

The choice of a large ultra-wide display and a touch-screen interface was deliberate, aiming to mimic the real-world set-up conditions of an AV control center. This setup allowed participants to engage with the remote environment and the HMI in a manner reflective of practical usage scenarios.

To design the scenarios, we leveraged the nuScenes dataset[1] [19], a publicly available dataset offering accurate AV sensor data for dynamic urban scenarios. The dataset provided crucial inputs, including data from three cameras for a 180° remote view, LiDAR data, and GPS/map information.

The combination of these hardware components and dataset sources was chosen to create a study environment that authentically reflects the challenges and interactions associated with assessing and taking over control in an AV scenario.

6.3.3 Study Design

Our analysis considered a robust set of independent and dependent variables, capturing the nuances of participant interactions and system responses. In particular, the study featured two independent variables, allowing participants

---

[1] https://github.com/nutonomy/nuscenes-devkit

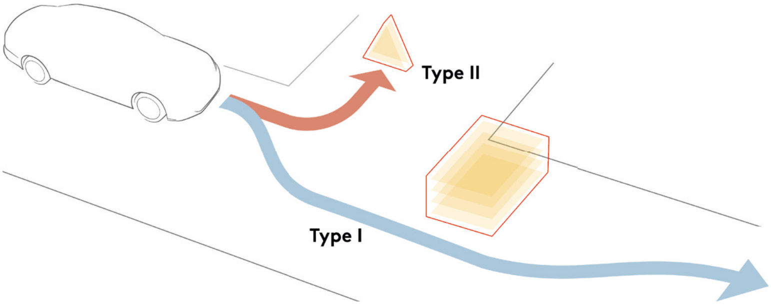


Figure 6.5: The illustrations depict the two types of scenarios we planned to deploy during the user study. In particular, the type I scenario displays an unexpected non-moving obstacle lying on the vehicle path. In contrast, type II scenario shows a road closures sign due to, e.g., an upcoming construction site.

to assess the remote environment with or without the dedicated HMI across four distinct scenarios. To ensure a comprehensive evaluation, scenarios were categorized into Type I and Type II, each presented through a video sequence.

6.3.3.1 Independent Variables

1. **HMI:** Participants interacted with the remote environment in the absence or presence of the HMI.
    a) **Presence:** Participants engaged with the remote environment utilizing the dedicated HMI.
    b) **Absence:** Participants engaged with the remote environment without the HMI, using this condition as the baseline for comparison. The absence of the HMI was deliberately configured to provide a fundamental reference point in performance evaluation and to understand the pivotal role of the HMI in the dynamics of the assigned task.
2. **Scenarios:** Scenarios were classified into two types.
    a) **Type I:** Featured unexpected non-moving obstacles on the vehicle path. In both scenes, the AV planned to go straight, but detected obstacles (fallen cases or standing vehicles), triggering the ToR system.

b) **Type II:** Involved road closures due to an upcoming construction site. In these scenarios, the AV was planning to turn left or go straight, but detected a road-closure sign before the intersection, initiating the ToR.

6.3.3.2 Dependent Variables

1. **Situation Awareness:** Assessed through the SAGAT questionnaire [58]. Participants provided responses after each task, thus removing the freezing period. The SAGAT was quantified as a percentage and averaged across questions targeting a demanding level of SA.
2. **Cognitive Load:** Evaluated using the NASA-TLX questionnaire [92]. The raw version of NASA-TLX was employed, eliminating the weighting process. Participants completed the questionnaire after each scenario.
3. **Takeover Time:** Recorded in seconds after each task, providing insights into the speed of participant takeover in different scenarios.

6.3.4 Procedure

Our user study followed a carefully designed procedure to ensure systematic data collection and participant engagement. Upon arrival, participants were warmly welcomed and asked to sign a consent form, providing demographic details to better contextualize the study findings.

After a brief introduction to the study's aim, participants were acquainted with the interface. Any queries regarding the interface were addressed at this stage to ensure a clear understanding. This initial interaction aimed to set the stage for participants to engage effectively with the upcoming tasks.

The experimental sessions consisted of four tasks, each designed to assess the participant's response to remote scenarios with and without the dedicated HMI. The order of tasks and the presence of HMI were counterbalanced to mitigate potential order effects.

Participants were instructed to assess the remote environment and, upon comprehension, touch the takeover button. The scenarios presented were derived from the nuScenes dataset, featuring both Type I and Type II scenarios with unexpected obstacles and road closures, respectively.

Post each task, participants completed the Situation Awareness Global Assessment Technique (SAGAT) questionnaire [58] and the NASA Task Load Index (NASA-TLX) questionnaire [92]. The SAGAT questionnaire aimed to measure the participant's level of SA by comparing their understanding with factual simulation data. The NASA-TLX questionnaire assessed perceived cognitive load, utilizing the raw version to eliminate the weighting process.

The study concluded with a debriefing session, where participants were provided with additional insights into the study's goals and methodology. Expressions of gratitude were extended, acknowledging their valuable contribution to advancing our understanding of AV interaction scenarios.

6.3.5 Results

In the following section, we present the ToT results, then discuss the SA results obtained via the SAGAT questionnaire [58], and those of the perceived cognitive load assessed via the NASA-TLX questionnaire [92]. We conclude the result section by exposing the correlation and linear regression analysis results.

6.3.5.1 Results of the ToT:

A paired-samples t-test was conducted to determine whether the ToT varied when participants were asked to evaluate the remote environment with the proposed HMI. Preliminary data screening showed that the data do not deviate from normality. The groups did differ, t (28) = -4.63, p < .001, 95% C.I. [10.05, 25.61], d = -.87. The ToT mean score with HMI (M = 35.09, SD = 19.32) was statistically significantly different than without HMI (M = 16.97, SD = 11.30) with an large effect size (d = .87). These findings suggest that the subjects required more time to take control of the vehicle when the proposed HMI was on.

	SAGAT		TOT		NASA-TLX	
	w/ UI	w/o UI	w/ UI	w/o UI	w/ UI	w/o UI
Valid	36	35	32	33	36	34
Missing	0	1	4	3	0	2
Mean	0.514	0.309	35.094	16.970	51.851	53.571
Std. Deviation	0.197	0.171	19.318	11.298	12.753	16.151
Shapiro-Wilk	0.959	0.921	0.930	0.907	0.984	0.943
P-value of Shapiro-Wilk	0.200	0.016	0.040	0.008	0.875	0.077
Minimum	0.100	0.050	7.000	3.000	26.190	26.190
Maximum	0.850	0.800	70.000	44.000	80.950	83.330

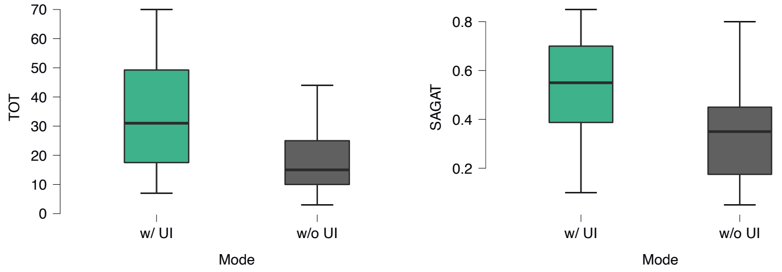
Table 6.1: The descriptive statistics of the variable tested during the first user study, in particular, those of the SAGAT queries [58], Takeover Time (ToT), and NASA-TLX questionnaire [92] are reported.

6.3.5.2 Results of the SAGAT:

A paired-samples t-test was conducted to determine whether the level of SA assessed via the SAGAT questionnaire varied when participants were asked to evaluate the remote environment with the proposed HMI. Preliminary data screening showed that the data do not deviate from normality. As showed by the paired-samples t-test, the groups did differ, t (34) = -4.95, p < .001, 95% C.I. [-.29, .12], d = -.84. The SAGAT mean score without HMI (M = .31, SD = .17) was statistically significantly different than with HMI (M = .51, SD = .2) with an large effect size (d = -.84). These findings suggest that a greater SA can be obtained when showing a dedicated HMI.

6.3.5.3 Results of the NASA-TLX:

A paired-samples t-test was conducted to determine whether participants subjective cognitive load varied when asked to evaluate the remote environment with the proposed HMI. Preliminary data screening showed that the NASA-TLX does not deviate from normality. The groups did not differ, t (33) = 1.17, p = n.s., 95% C.I. [-2.37, 8.82], d = .2. hese findings suggest that the participants perceived an equal cognitive load with or without HMI.



(a) **Takeover Time** (b) **Situation Awareness**

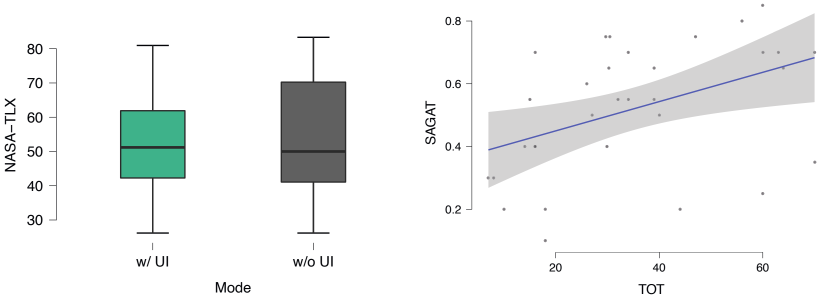
Figure 6.6: (a) The descriptive boxplots of the participant's takeover Time (TOT). (b) The descriptive boxplots of the SAGAT questionnaire.

6.3.5.4 A Correlation & Linear Regression Analysis:

A Pearson's rho correlation was conducted to assess the relationship between the ToT and SA level obtained via the SAGAT questionnaire. Preliminary analyses showed a linear relationship with both variables normally distributed, as assessed by Shapiro-Wilk's test, and some outliers were found and eliminated (cf. Table 8.1). There was a significant positive correlation between the perceived SA and ToT, $r = .44$, $p < .01$. The linear regression established that the ToT could statistically significantly predict the SA level, $F(1,30) = 7.15$, $p < .01$; and ToT accounted for 19% of the explained variability in SA. The regression equation was: $SA = .36 + (.005 \times ToT)$.

6.3.5.5 Qualitative Results

Having administrated all the conditions, we interviewed the participant asking their opinion about the proposed HMI. The interview was conducted in a semistructured manner that inquired about the advantages, disadvantages, and participant suggestions. To quickly assess the remote environment, helpful information perceived by the participants was a) the visualization that highlighted the obstacle, b) the predicted trajectory of the other road users, and c) the LiDAR visualization. On the contrary, the map was found too prominent and did not help solve the proposed tasks. Subjects suggested having a degree



(a) **Subjective workload** (b) **Linear Regression**

Figure 6.7: (a) The descriptive boxplots of the subjective workload assessed via NASA-TLX questionnaire, and (b) the linear regression that predicts the SA index based on ToT values.

of danger report, i.e., an additional message informing about the risk level and the warning signal. Participants felt the need to have further information about the remote AV initial state, a visualization highlighting approaching pedestrians, and the street boundaries. Lastly, participants suggested displaying a takeover countdown.

6.3.6 Discussion

The analysis results lay the foundation for an initial discussion on how to design future ToR HMIs for vehicle teleoperation. When we compared the ToT data with those previously observed by our peers (cf. Section 6.1), we noted overall higher ToT subject response ($M = 35.09$, $SD = 19.32$), which to some extent approach those observed by Eriksson and Stanton [65], with the difference that the subjects of our study were asked to operate a vehicle remotely, i.e., physically separate from the vehicle. We further recorded higher ToTs when participants were required to assess the remote environment when the HMI was active ($MD = 17.83$); however, obtaining significantly higher SA levels (+20%). Moreover, the linear regression equation had explained that subjects who spent 35.09 seconds (mean average) analyzing the ToR had scored 53% of SAGAT correct answer.

Thus, if we would like to designate a minimum SA threshold equal to 60%, we need to consider that an average operator can take up to 50 seconds before taking complete control over the AV.

Lastly, the subjective perceived cognitive load evaluation does not show any significance, suggesting that the participants with or without HMI perceived an equal cognitive load.

These are positive results that suggest that the HMI tested, could increase the operator's SA, a critical requirement for vehicle teleoperation, yet high ToT values remain. Future works should present the design of ToR HMIs introducing approaches that include the prediction of the AV disengagement, as in Kuhn et al. [115], to allow operators more time to assess the remote environment. We discuss such an approach in Chapter 7.

Addressing the limitation of this chapter, the methodological concerns regarding the use of the SAGAT, particularly in relation to timing versus freeze frame, warrant thorough discussion. The choice of SAGAT methodology was made with careful consideration, aiming to capture participants' cognitive processes at specific points in time. Acknowledging the complexities inherent in this method, we employed the SAGAT after the assessment by the participants of the remote situation, thus removing the freezing period. Here, the high ToT might not be directly attributed to the SAGAT, as it might be influenced by factors beyond its scope.

Furthermore, we are aware that with this work, we can only compare the effect on participant SA and perceived cognitive load with and without the designated interfaces to understand whether it serves its purpose correctly or not. This limits the possibility of discussing how effective the HMI is compared to others. Nevertheless, ToT data shows how the proposed concepts are positioned with similar works reported in Section 6.1. A further limitation is the few scenarios we tested, making it hard to generalize the result to more situations. Thus, future studies should consider integrating and validating a broader set of scenarios to guarantee a comprehensive understanding of how to model remote ToR HMIs. More studies should also be performed to reduce remote ToT while improving SA. Nonetheless, validations of other interface concepts should not be neglected.

6.4 Summary

The chapter presents a ToR interface for AV remote operation. As described, previous research has focused on ToR strategies for in-vehicle drivers, yet less for remote operators. In this chapter, a first ToR interface were designed and evaluated to improve transparency and trust in the system for remote operators.

The proposed HMI was effective in improving SA and overall performance, yet it has increased the time it takes for a user to take control of a vehicle. The study also found a positive correlation between SA and ToT, suggesting that increasing SA increases the time. Finally, participants suggested enhancements for the proposed HMI, recommending features like displaying the degree of danger and providing additional information about the initial state of the remote AV.

7

Detecting and Visualizing Anomalies

The Chapter 6 introduced a ToR HMI for remote vehicle operations. We noticed that although the HMI could increase the operator SA, high ToT values continued to be recorded. To solve this issue, in this chapter, we present one approach that includes the prediction of the AV disengagement to the ToR, allowing operators more time to assess the remote environment. In particular, this method aims to increase the teleoperator's SA by highlighting which areas of the remote environment are responsible for the ToR. Thus, to complement the existing ToR HMI presented in Chapter 6, hereafter, we show how the operator can quickly infer *where* the vehicle finds uncertainty in its environment.

7.1 Related Work

To facilitate the teleoperator's understanding of the remote environment and to reduce ToTs, it is essential to design intuitive HMIs that do not cause excessive cognitive work. Merat et al. [143], in a study on AVs ToR, have found that it can take up to 40 seconds for a subject to fully stabilize the vehicle in a non-

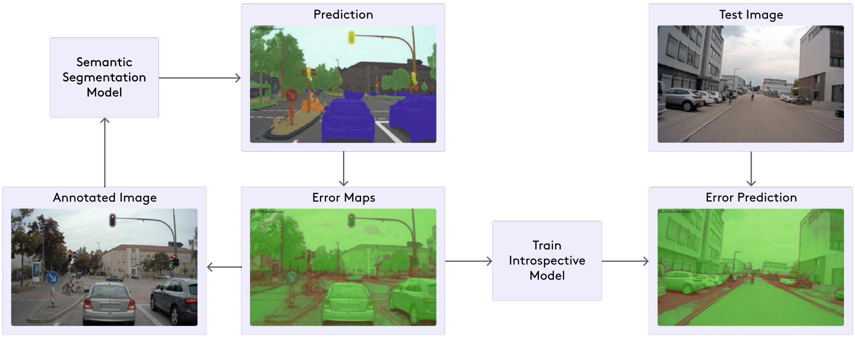


Figure 7.1: An adapted overview of the introspective failure prediction approach for semantic segmentation proposed by Kuhn et al. [114].

critical scenario. We acknowledge that a quick takeover, which requires a direct understanding of the request call, is even more essential in remote hazardous situations.

A prior body of work has proposed approaches to predict future vehicle disengagements of semi-autonomous [71] or autonomous vehicles [115] up to seven seconds, allowing for more time to assess the remote environment. These methods mainly focus on detecting incorrect steering angles by predicting disengagements in a binary fashion. For instance, Fridman et al. [71] have developed a method they call *"arguing machines"* that adds a secondary artificial intelligent system to a primary one to detect future disagreements. The disagreement between the two systems is used as a predictor for the ToR. To evaluate the proposed method, the author implemented the *"arguing machines"* to the Tesla Autopilot software and could predict 5 seconds early whether a disengagement would have happened with 90.4% accuracy.

Other approaches have also been proposed to early vehicle disengagement, for example, by applying the concept of introspective perception. Initially developed by Daftry et al. [41], the concept of introspection was first applied in vision systems that built a perception system with introspective behaviors, i.e., in systems that know when they make mistakes. Inspired by this idea, Kuhn et al. [115] proposed an introspection model that uses previously detected failures to predict future ones. With data collected from a BMW AV in their evaluation, the authors could predict failures up to seven seconds early with more than 80% accuracy.

In a follow-up paper, to locate the reason for the control request Kuhn et al. [114] proposed an introspection model used in semantic segmentation to predict the pixel-wise failures (Figure 7.1). Thus, instead of classifying the entire input image as a failure, the model could determine which part of the input image is incorrectly classified. The authors trained a dedicated failure prediction model using the failures generated by a pre-trained semantic segmentation model. The suggested approach outperforms, by 3.2% and 6.7%, the prior estimation uncertainties methods while requiring fewer resources during inference, predicting when and where the semantic segmentation will fail.

Other existing works have also focused on detecting anomalies based on image inputs [123, 249]. For example, Xia et al. [249] addressed the pixel-level failure prediction task in two steps. Firstly, by synthesizing an image from a semantic label, secondly, comparing the difference between the generated image and the original image to obtain the anomaly areas. The authors could validate and improve current anomaly detection methods in different challenging research fields, from autonomous driving to medical image analysis, improving current existing methods.

Differently, Lis et al. [123] base their method on the idea that a network generates false labels in regions representing unexpected objects. The authors resynthesized the image starting from the resulting semantic map, obtaining significant appearance differences from the input image. Thus, instead of detecting unknown classes, Lis et al. [123] shift the problem to the identification of mis-resynthesized regions, showing that the method defeats those based on both uncertainty and autoencoder designs.

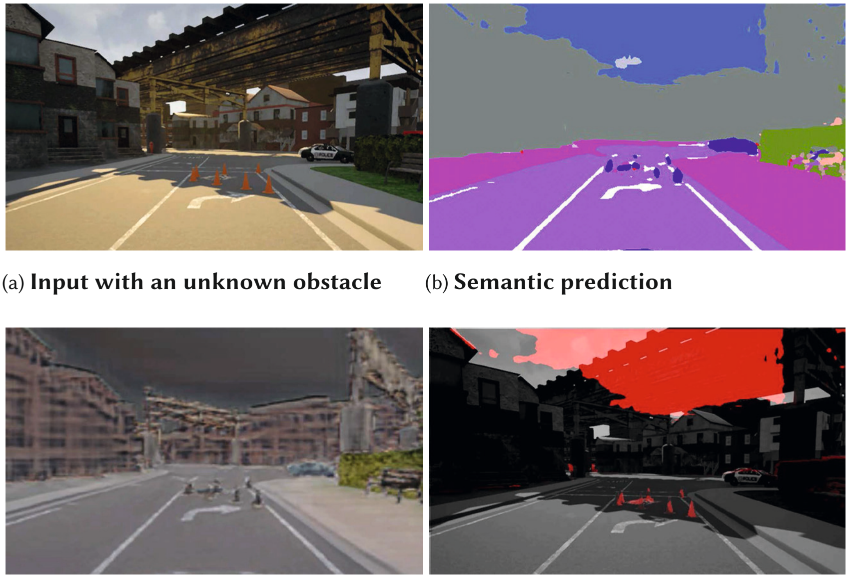
To support the information acquisition process, our idea is to apply the aforementioned methods to alert the teleoperator visually to imminent criticalities. We extend these methods in two ways. First, our method uses the spatial nature of fault prediction to visualize for the teleoperator where the fault is located in the environment. Second, we propose to use the image-based failure detection method in a predictive manner; this is essential for the teleoperator, where early reactions to problems are required. Thus, the teleoperator can focus directly on the anomaly to take quick countermeasures.

7.2 Our Approach

Our prototype uses failure prediction methods to reveal anomalies in the remote environment based on the visual inputs we receive from the cameras mounted on the remote vehicle. Prior work has shown that inverse image synthesis is an accurate technique in reporting anomalies of a video image [123, 249]. However, in addition to the prior body of work, we use inverse synthesis to calculate how many faults the semantic segmentation has made. In our approach, this data is processed to calculate the average error per image used for early anomaly prediction. The final output of the back-end computation is twofold. On the one hand, the resultant triggers the request for a human takeover. On the other hand, we can show the teleoperator the areas that caused the remote ToR.

This may be the case of an AV entering an area that does not meet the standards, e.g., due to some unknown objects on the vehicle way, such as traffic cones (Figure 7.2a). In such a situation, the semantic prediction will assign a specific class to the traffic cones, in Figure 7.2b shown as dark blue. The prediction synthesization result (Figure 7.2c), derived from the semantic prediction (Figure 7.2b), provides us with a photo-realistic approximation of the remote environment that the system received as input (Figure 7.2a). For the prediction synthesization task, we implemented a Generative Adversarial Network [79] that compares the original image with the one generated to highlight differences [123]. Next, we employed the pixel-level error detection method to generate a heatmap, highlighting misclassifications in red (Figure 7.2d). Eventually, misclassifications, such as the bridge identified as the sky, are automatically omitted.

This process allows the teleoperator to be alerted up to seven seconds in advance about the misinterpreted/misclassified areas and to be informed about



(a) **Input with an unknown obstacle** (b) **Semantic prediction**

(c) **Synthesized input** (d) **Critical objects detected**

Figure 7.2: To detect anomalies in the remote environment, an input image (a) is used to deliver a semantic prediction (b). Using a dedicated neural network, the semantic prediction can be synthesized (c) and compared to the input image (a). The comparison is shown in (d), which presents the semantic prediction errors. The results of an image-based failure prediction are used to visualize the anomalies for the teleoperator. The critical and the misclassified objects on the road are highlighted, allowing the teleoperator to be immediately aware of the AV challenge.

the critical safety issues known to the AV [115]. Thus, compared to the previous HMIs we evaluated, we radically reduced the information that the teleoperator has to process.

For this reason, along with an egocentric view of the environment, a visualization that embeds a symbolic representation of danger, has been integrated into the simulation (Figure 7.3a). The message is aligned with the real world, placed with spatial relation to the anomaly of interest, i.e., 2D registered. Therefore, by integrating a critical threshold to evaluate the difference between the result obtained by the prediction synthesization task and the remote camera video images, challenging scenarios can be routed to the teleoperator in advance, highlighting the uncertainties discovered.

(a) **Registered visualization** (b) **No visualization**

Figure 7.3: In this image, the explanatory scenario that includes the two independent variables of the user study, i.e., the registered visualization (a) and its absence (b).

7.3 User Study

To evaluate the proposed method for anomaly detection and visualization, we performed a lab-based driving simulator study recording the operator's perceived cognitive load and SA. The user study was designed to answer the research questions:

Does the anomaly detector improve the teleoperator SA? Moreover, will the anomaly detector influence the teleoperator cognitive load?

We answered the questions using a within-subject design [28], so that participants would go through all conditions. The order that participants went through the presence or absence of HMI experienced over two different durations (7 and 10 seconds) followed a Latin square design [33].

7.3.1 Implementation Set-Up

The whole project is built and based on CARLA[1] [47]. CAR Learning to Act (CARLA) is an open-source simulator for urban environments, developed for training, prototyping, and validating AV models, including perception and control. The project architecture relies on a client-server concept that provides a flexible Python Application Programming Interface (API) on the client-side (Figure 7.4). And on the server-side, the rendering allowing for the association

---

[1] https://github.com/carla-simulator/carla

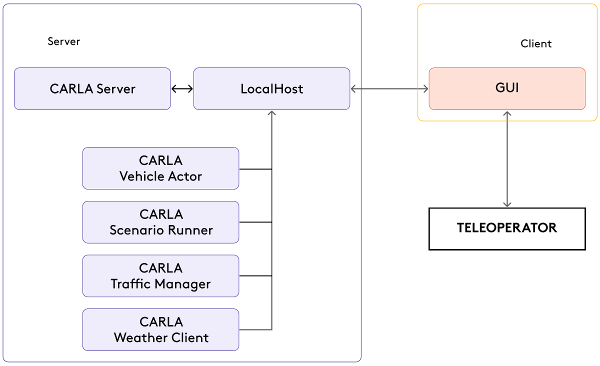


Figure 7.4: The project architecture based on CAR Learning to Act (CARLA) [47]. The system relies on a client-server concept that provides on the client-side a flexible Python Application Programming Interface (API), and on the server-side, the rendering allowing for the association of multiple clients.

of multiple clients, e.g., CARLA Vehicle Actors, CARLA Scenario Runner, CARLA Traffic Manager, CARLA Weather Client. The different clients are connected with the Localhost, which communicates with the CARLA server. The Graphical User Interface (GUI) module simulates the image video streaming by providing three front cameras to the user (cf. Figure 7.5).

The decision to transition to CARLA from the nuScenes dataset that we discussed in Chapter 6, is rooted in the unique advantages that CARLA offers in comparison to static datasets such as nuScenes. CARLA serves as a dynamic and flexible simulator, providing us with unparalleled customization capabilities in scenario design. Unlike static datasets, CARLA empowers us to tailor and adapt scenarios to our specific research needs, offering a level of flexibility that is crucial for the nuanced exploration of AV models. This flexibility is especially valuable in the context of training, prototyping, and validating AV models for perception and control. In contrast, static datasets inherently pose limitations in customization, constraining the range of scenarios that can be explored. The shift to CARLA aligns with our pursuit of a dynamic and adaptable research environment, enhancing the depth and breadth of our investigations in the field of AV.

Figure 7.5: The driving scenario we administrated during the user study. In particular, in this image, we see the remote environments augmented with a warning signal close to each hazardous area.

7.3.2 Participants

Fifteen participants (5 female) were recruited for this study, with a mean age of 27 years (SD = 2.73), ranging from 24 to 35. The participants, on average, held a driving license for 7.5 years (SD = 3.47). A significant majority of participants, 73%, reported previous experience in remote driving activities, such as steering drones, controlling RC cars, and similar endeavors. Within this group, 27% had controlled a remote vehicle several times, 55% had done so on occasion, and 18% had only one instance of remote control experience. It is noteworthy that none of the participants had prior experience in remote controlling autonomous vehicles (AVs). This diverse participant pool, with varying levels of remote driving experience, contributes to the study's ability to capture a range of perspectives and responses.

7.3.3 Apparatus and Materials

The materials employed during the user study included a local computer with an Intel i7 CPU, NVIDIA GeForce GTX-1080, and 8 GB memory. This computer ran the failure prediction models within the open-source autonomous driving simulator CARLA [47] (cf. Section 7.3.1). The same computer provided the virtual environment through three monitors, each 23.8 inches (1920 × 1080 pixels), covering a total horizontal and vertical area of 161.7 × 49.8 cm (Figure 7.5).

7.3.4 Study Design

The study design encompassed two independent variables: the presence or absence of registered visualization and its duration. Participants experienced these conditions over one scenario depicting a four-way intersection with multiple lanes per road (Figure 7.5). In this scenario, the AV encountered an obstacle—an unidentified road construction site—blocking its intended path, prompting the teleoperation request. Participants assessed the remote situation by reviewing the driving scenario for seven and ten seconds, with two trials per participant.

7.3.4.1 Independent Variables

1. **HMI:** Participants interacted with the remote environment in the absence or presence of the HMI.
    a) **Presence:** Participants engaged with the remote environment utilizing the dedicated HMI.
    b) **Absence:** Participants engaged with the remote environment without the HMI, using this condition as the baseline for comparison. The absence of the HMI was deliberately configured to provide a fundamental reference point in performance evaluation and to understand the pivotal role of the HMI in the dynamics of the assigned task.
2. **Duration:** Participants evaluated the remote environment during two distinct time intervals.
    a) **Duration of 7 seconds:** In accordance with the recommendation by Kuhn et al. [115], a duration of 7 seconds was chosen for the first time interval. This specific duration was selected based on their model's ability to accurately predict failures up to seven seconds in advance, achieving an accuracy rate exceeding 80%. Consequently, participants were instructed to assess the remote environment for a period of 7 seconds.
    b) **Duration of 10 seconds:** For the second time interval, a duration of 10 seconds was selected. This duration was chosen as it repre-

sents an increment of approximately 40% from the prior interval suggested by Kuhn et al. [115]. Thus, during this interval, participants were instructed to assess the remote environment for a period of 10 seconds.

7.3.4.2 Dependent Variables

1. **Situation Awareness:** Assessed through the SAGAT questionnaire [58]. Participants provided responses after each task, thus removing the freezing period. The SAGAT was quantified as a percentage and averaged across questions targeting a demanding level of SA.
2. **Cognitive Load:** Evaluated using the NASA-TLX questionnaire [92]. The raw version of NASA-TLX was employed, eliminating the weighting process. Participants completed the questionnaire after each scenario.

7.3.5 Procedure

Upon participants' arrival, a warm welcome awaited them, accompanied by a request to sign a consent and demographics form, along with an introduction to the study's aim. Participants were tasked with assessing the remote environment during the experiment, detecting anomalies that triggered the teleoperation request. The task was repeated twice, both with and without the registered visualization, experienced for durations of 7 and 10 seconds.

After each task, participants completed the NASA-TLX questionnaire [92] and the SAGAT questionnaire [58]. The NASA-TLX measured perceived cognitive load, employing the "raw" version to eliminate the weighting process. The SAGAT questionnaire determined the participants' SA level, comparing their understanding with factual simulation data. The study concluded with a debriefing session, expressing gratitude for their valuable participation.

7.3.6 Result

The results are presented and divided into two main categories. We first discuss the SA results obtained via the SAGAT questionnaire [58], then those of the perceived cognitive load assessed via the NASA-TLX questionnaire [92].

	SAR7	SAR10	SAN7	SAN10	NTR7	NTR10	NTN7	NTN10
Valid	5	5	5	4	5	3	5	5
Missing	0	0	0	1	0	2	0	0
Mean	0.400	0.800	0.100	0.375	39.668	41.667	37.734	39.332
Std. Dev.	0.285	0.209	0.137	0.144	11.018	1.665	13.798	12.671
Shapiro-Wilk	0.961	0.881	0.684	0.729	0.938	1.000	0.907	0.979
Shapiro-Wilk P-value	0.814	0.314	0.006	0.024	0.654	0.997	0.448	0.927
Minimum	0.000	0.500	0.000	0.250	28.330	40.000	20.000	23.330
Maximum	0.750	1.000	0.250	0.500	55.000	43.330	51.670	55.000

Table 7.1: Descriptive statistics of the subjective assessment of workload assessed via NASA-TLX and Situation Awareness (SA) via the SAGAT questionnaire. Note: SA stands for the SAGAT questionnaire, NT for the NASA-TLX, R for registered, N for none, 7 and 10 is the duration.

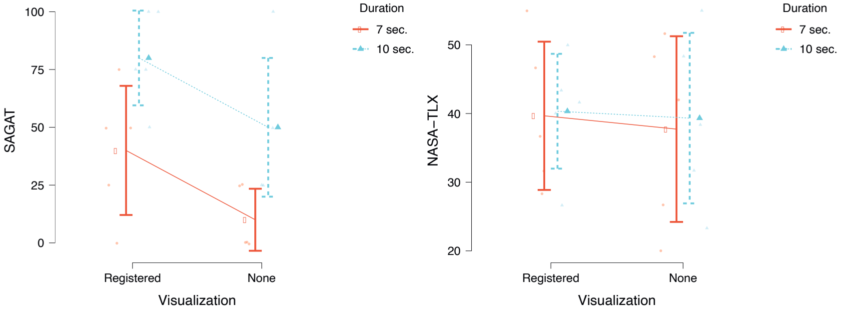
7.3.6.1 Analysis of Variance on Situation Awareness

An ANOVA was conducted to determine whether the level of SA assessed via the SAGAT questionnaire [58] varied with and without the visualization and the time. Preliminary data screening showed that the data did not deviate from normality. One outlier was found as we screened the data for assumptions which we removed.

The visualizations had a significant effect on SA, F (1,3) = 121, p = .002, ηp^2 = .32. Bonferroni post hoc analysis revealed a significant increment on participant's SA with the presence, i.e., registered visualization (MD = .34), p = .002, d = 5.5. Similarly, although moderately, time had a significant effect on participant's SA, F (1,3) = 10.37, p = .049, ηp^2 = .32. Bonferroni post hoc analysis revealed a significant increment of participants' SA when assessing the environment for ten seconds (MD = .34), p = .049, d = 1.6. However, no statistically significant differences were recorded between the time and the visualization F (1,3) = .3, p = n.s., ηp^2 = .003.

7.3.6.2 Analysis of Variance on the perceived cognitive load

Moreover, an ANOVA was conducted to determine whether the conditions differ in perceived cognitive load. Preliminary data screening showed that the data did not deviate from normality. Two outliers were found as we screened the data for assumptions which we removed.



(a) **SAGAT over time**

(b) **NASA-TLX over the visualizations**

Figure 7.6: In these descriptive plots, (a) we placed the SA as assessed by the SAGAT questionnaire [58] in relation with time and visualizations. Similarly, we compared (b) the perceived cognitive load assessed via the NASA-TLX questionnaire [92].

There was no statistically significant difference neither for the visualizations F (1,2) = .68, p = n.s. ηp^2 = .05, nor between time, F (1,2) = 1.3, p = n.s., ηp^2 = .1.

7.4 Discussion

From the human factor perspective, the teleoperator faces numerous challenges, from high latency [159] to low remote SA [160]. Related work has shown that driver assistance systems for teleoperated autonomous vehicles [32, 35] have improved operators' performance and SA. Likewise, in the study we reported in this chapter, subjects could achieve higher SA levels when the anomaly detector was active. In particular, we noticed that participants could gain 40% of SA with the help of the dedicated HMI after seven seconds, whereas only 10% without. However, significant SA improvements have been recorded when we let the subject assess the remote environment for ten seconds. Here, 80% of SA has been acquired after ten seconds with anomaly detector, and 50% without. These results suggest that simple additional visual cues, such as the warning signals overlayed onto the video images, may improve the SA and reduce ToTs. We interpret these findings as a suggestion favoring a reduction of the ToR HMI.

However, no significant changes were recorded when assessing participants' cognitive load, as the anomaly detector could not significantly reduce the operational workload. We believe that these results derive from the user study design since participants could only experience the scenarios for a few seconds, too little for a subjective assessment of workload.

Unlike prior works, our prototype adds an important safety component over existing anomaly detection methods, namely showing where the problem is located. Although these are only the first results and further studies must be conducted to confirm our findings, we are confident that our proposed approach will reduce the ToT and improve the usability of the control of teleoperated AVs. Eventually, future work may consider supplementary visualization, such as polygons that delineate the risk area or corridors that explain viable paths to drive; this may further improve the quality and efficiency of the anomaly detector we showed.

7.5 Summary

This chapter presents a new approach to increase the teleoperator's SA. The method included the prediction of AV disengagement and highlighted areas of the remote environment responsible for the ToR. The method is an improvement over existing approaches, which only alert the operator that the AV is in a difficult situation and does not provide information about the specific cause of the request. By predicting the disengagement and highlighting the relevant areas of the remote environment, the operator could quickly understand the situation and take appropriate action. The approach has been evaluated and found to improve the teleoperator's SA and reduce ToT.

En-route: teleoperator assistance systems

8

The Predictive Corridor

In Chapter 6 and Chapter 7 we introduced remote ToR HMI concepts and presented one approach that includes the prediction of the AV disengagement by highlighting where the vehicle found uncertainty in the remote environment. Thus, having assessed the situation and the risks of the request, the teleoperator is ready to drive the vehicle.

This chapter provides a detailed exploration of the following publication:

Gaetano Graf, Yomna Abdelrahman, Hao Xu, Yasmeen Abdrabou, Dmitrij Schitz, Heinrich Hußmann, and Florian Alt. "The Predictive Corridor: A Virtual Augmented Driving Assistance System for Teleoperated Autonomous Vehicles." In: *ICAT-EGVE 2020 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. Ed. by Ferran Argelaguet, Ryan McMahan, and Maki Sugimoto. The Eurographics Association, 2020. ISBN: 978-3-03868-111-3. DOI: `10.2312/egve.20201260`

Gaetano Graf, Hao Xu, Dmitrij Schitz, and Xiao Xu. "Improving the Prediction Accuracy of Predictive Displays for Teleoperated Autonomous Vehicles." In: *6th International Conference on Control, Automation and Robotics*. IEEE, 2020, pp. 440–445. DOI: `10.1109/ICCAR49639.2020.9108011`

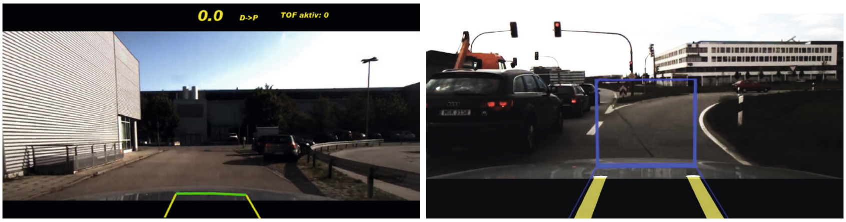
In this chapter, we would like to introduce an Advanced Teleoperator Assistance System (ATAS) to mitigate low depth perception and time latency called Predictive Corridor. The Predictive Corridor combines a novel Predictive Display with the Free Corridor. The final result is an interface that shows the predicted position of the remote vehicle and the vehicle braking to the teleoperator.

8.1 Related Work

Direct controls are the most widespread method to conduct AVs remotely. By using control inputs such as the steering wheel, pedal, or joysticks, the teleoperator during the task is responsible the whole time for observing the remote environment and deciding on an appropriate strategy, i.e., in the case of direct driving, the execution of steering and acceleration/deceleration [32, 75]. The telepresence of operators at this stage is mandatory as they are the ones that could close the control loop [203] and also the ones that can stabilize the vehicle through the driving maneuver.

To guide the vehicle safely and to guarantee robustness during the teleoperation against time delays, the operator can be supported by ATAS. For instance, the Free Corridor (FC) proposed by Chen [32] shows a braking path in which the AV will continue to travel in cases of unexpected network losses (Figure 8.1a). In such a scenario, an automatic emergency braking would be triggered, as the remote vehicle can not update the control commands from the operator. Thus, in the FC, the vehicle would follow the predetermined trajectory and stop if no further connection with the operator is available [32].

Another ATAS is the Predictive Display (PD) [35]. The PD, as the name implies, predicts the movements of the remote vehicle and the movements of other road users to forecast the vehicle position to overcome the time delay (Figure 8.1b). The prediction of the remote vehicle is shown to the teleoperator as an overlay on the delayed video images. On the one hand, we note that the FC provides a solution in case of communication loss, yet it does not address the problem with time latency. On the other hand, the PD allows overcoming time latency, but no emergency concept is provided in situations of communication



(a) **The Free Corridor** (b) **The Predictive Display**

Figure 8.1: Two remote driver assistance systems for vehicle teleoperation. In both concepts, the teleoperator's prediction of the remote vehicle position is shown as an overlay on the delayed video images. In the Free Corridor (a), the teleoperator sees a path of the full braking [32]. The Predictive Display (b) forecasts the movements of the remote vehicle and the movements of other road users to predict the vehicle position.

failure. It has been stated that a combination of both concepts would benefit the teleoperator since strategies to mitigate the time delays and the required emergency concept would be present [96].

In recent works, Hosseini [96] explored this opportunity. The author combines the FC and the PD, adding an emergency braking system. In the concept, called Predictive Brake Assistance (PBA), we first see the PD, and seamlessly attached to it, the FC. That means that the calculation and plotting of the FC do not begin at the current vehicle position but at the end of the predicted position of the PD. In the author's concept, the PBA would be triggered when other road users overlook the AV driving path.

Despite being a valuable concept to mitigate low depth perception and time latency, the PBA predicts the vehicle position considering, exclusively, the remote vehicle inputs, i.e., the velocity and yaw rate. The effectiveness of the PBA is therefore dependent on the remote vehicle inputs and communication delay. To solve this issue, we propose a novel method that considers the remote vehicle inputs and the local operator inputs to predict the vehicle position (Figure 8.2). Thereby, our method, referred to as Predictive Corridor (PC) [87], offers two key advantages: it is less reliant on vehicle input and more robust against time latency in mobile networks.

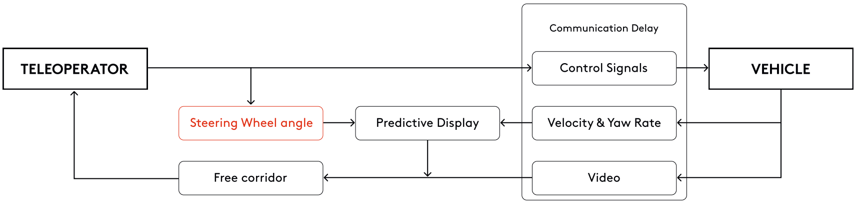


Figure 8.2: The novel curvature model proposed in this chapter. In prior works, authors exclusively consider the remote vehicle inputs, i.e., the velocity and yaw rate, to predict the remote vehicle position. In contrast, we propose a novel curvature model that considers both the teleoperator inputs (in red) and the remote vehicle states to calculate the curvature value.

8.2 Our Approach

In our prototype, the predicted trajectory is generated using the clothoid method, a spline with linear changing curvature values that can smoothly combine arcs with different radii. Prior works had calculated the curvature by using the yaw rate and speed of the remote vehicle [35, 96]. However, as introduced, those methods rely on a curvature model that highly depends on time delays as the AV input signals are transferred wirelessly via cellular networks. We address this problem by introducing a new model that integrates the operator's input into the current prediction state-of-the-art method [243].

In an initial open-loop evaluations we demonstrated that our approach has, overall, the best performance in low speeds scenarios compared to the state-of-the-art prediction method [87]. In particular, in an experiment, we accelerated the vehicle to the desired velocity of 15 km/h and performed an avoidance maneuver, i.e., a double lane change maneuver [73]. We designed the evaluation into SPIDER, a Software Programming Interface for Distributed Real-time Driving Simulation [219], where we set a constant communication latency. We sent the control commands of the teleoperator to the remote vehicle with a 200 ms delay, and we sent back the vehicle state feedback with an additional delay of 200 ms. Therefore, a 400 ms time delay affected the whole control loop, i.e., Round-Trip-Time. The vehicle state (e.g., speed, steering wheel angle) was updated every 0.01 ms, whereas the predicted path every 400 ms. Compared to the state-of-the-art prediction method, the performance results of

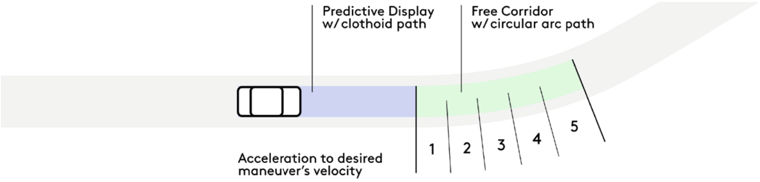


Figure 8.3: The Predictive Corridor combines into a single virtual augmented element the Predictive Display (in blue) as proposed by Graf et al. [87] and the Free Corridor (FC), in green, as proposed by Chen [32]. The prior is based on the clothoid method, the latter on the circular arc method. The total length of the FC is calculated by adding (1) video transmission delay, 2) operator's perception-reaction delay, 3) control delay, 4) brake activation, 5) brake distance.

our approach indicated that the model had 6.2% less vertical average absolute deviation, 9.4% less lateral average absolute deviation, and 7.3% less euclidean absolute deviation.

These first positive results have created the foundation to develop our concept further. Thus, having evaluated the novel curvature model, we attached to it the FC [32]. The FC presents a path in a circular form with one fixed center point and a steady curvature, which according to Chen [32], is suitable for permanent curvature path prediction. The final total length of the FC was calculated as proposed by Hosseini [96] by adding 1) the video transmission delay, 2) the operator's perception-reaction delay, 3) the control delay, 4) the brake activation, and 5) the brake distance (Figure 8.3). We considered the video transmission delay and control delay to be 400 ms, whereas the teleoperator perception/reaction time and brake activation time to be 1.15 s [96]. We calculated the braking distance as $S_b = v^2/a_{max}$, where the deceleration a_{max} was directly obtained from the SPIDER vehicle dynamic model.

Based on the recent development described above, further analysis regarding the combination of the FC and the PD is required. In this regard, we conducted a user study with 32 participants, where we evaluated the PC by employing subjective and objective performance measures for workload assessment.

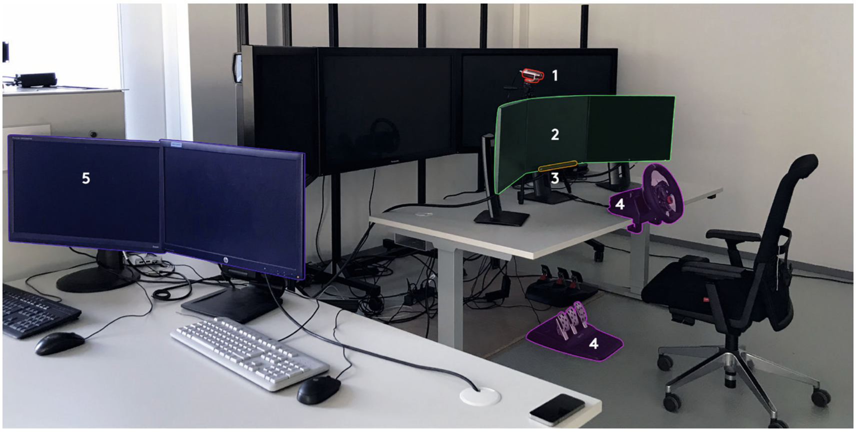


Figure 8.4: The experimental setup. In this image, we see (1) thermal camera Optris Xi 400, (2) three 23,8-inch LCDs, (3) Tobii Eye Tracker 4C, (4) Logitech G29 steering wheel and pedals, and (5) experimenter computer.

8.3 User Study

To evaluate the effectiveness of the Predictive Corridor, we performed a lab-based driving simulator study (Figure 8.6). We recorded the participant's driving performance, the subjectively perceived cognitive load, eye movements, and facial temperatures for an objective analysis of the operator workload. The user study was designed to answer the research questions:

Does the Predictive Corridor influence the cognitive load, visual attention, and the performances of the teleoperator?

To address our research question, we employed a within-subject study design. In addition to the baseline recording, where facial temperature was measured, the sequence in which participants undertook the two test courses, with and without the PC, followed a Latin square design [33].

8.3.1 Participants

We recruited a total of 32 participants (5 female). The participants had a mean age of 27 years (SD = 5,06), ranging from 18 to 44 years. On average, participants owned a driving license for eight years (SD = 4,25). More than one-third of the participants, i.e., 37,5%, declared driving a vehicle daily, whereas 25% drive weekly. The rest of the participants are divided into 18,8% monthly, 15,6% rarely, and 3,1% never. Additionally, we assessed the participant's remote driving experience and frequency of use. Most of the participants, 75%, already had remote driving experience, e.g., drones steering, RC car, and similar. Of those participants, 37,5% controlled a remote vehicle several times, 29,2% sometimes, yet 33,3% only once.

8.3.2 Apparatus and Materials

Our experimental setup consisted of three monitors, each 23,8-inch big (1920 × 1080 pixel) covering a total horizontal and vertical area of 161,7 cm × 49,8 cm. A Logitech G29 steering wheel was installed to control the vehicle. The steering wheel provided force feedback and a 900-degree rotation. One computer enabled the communication of input and output signals between the simulator and the automotive control elements (i.e., steering wheel and pedals). Furthermore, three other computers, each per monitor, generated the virtual environment. To record the x- and y-coordinates of the participant gaze, we attached a commercial eye tracker, Tobii Eye Tracker 4C[1], to the middle monitor, operating with a frequency of approximately 70 Hz, connected via USB. Additionally, a compact infrared camera was installed on a tripod beyond the screens to assess the participant's facial temperature. The Optris Xi 400[2] camera measures temperatures between -20 and 900°C, and an optical resolution enables a spot-distance ratio of up to 390:1. The optical resolution of the camera is 382 × 288 pixels, with a frame rate of 80 Hz and thermal sensitivity NETD of 80 mK. The ambient temperature was kept at 25°C constant throughout the experiment to avoid noise in the data set. The Optris PI[3] connects the software with the camera. We annotated the regions of interest, i.e., forehead and nose, and used the built-in data extraction function to store the temperature values [2].

---

[1] https://gaming.tobii.com

[2] https://optris.com/optris-xi-400

[3] https://optris.com/software-development-kits

8.3.3 Study Design

The study design included two independent variable the presence or absence of PC, and two scenarios. In particular, participants experienced the PC related to a LC and an EBS. Dependent variables encompassed subjective and objective perceived cognitive load assessed via the NASA-TLX questionnaire and thermal imaging. Since teleoperating involves controlling a vehicle under some time latency, we simulated a constant time delay of 400 ms in this study. This was suggested as prior research had shown that a constant latency is easier to manage than a variable one [44, 124, 131]. The gas and brake pedals were adjusted for consistent linear acceleration and deceleration.

8.3.3.1 Independent Variables

1. **HMI:** Participants interacted with the remote environment in the absence or presence of the HMI.
    a) **Presence:** Participants engaged with the remote environment utilizing the dedicated HMI.
    b) **Absence:** Participants engaged with the remote environment without the HMI, using this condition as the baseline for comparison. The absence of the HMI was deliberately configured to provide a fundamental reference point in performance evaluation and to understand the pivotal role of the HMI in the dynamics of the assigned task.
2. **Scenarios:** Two test courses were created to assess the performance of the Predictive Corridor (Figure 8.6). These courses were modeled after the pre-crash typology defined by the National Highway Traffic Safety Administration [155]. The objective for both test courses was to experience longitudinal and lateral control of the vehicle under clear weather conditions, in daylight, with a determined speed limit of 15 km/h [96].
    a) **Lane-Change (LC):** The LC course simulates an obstacle avoidance maneuver without prior action [73].

    b) **Emergency Braking Stop (EBS):** The EBS test course simulates an emergency braking stop on a straight path (50 m).

8.3.3.2 Dependent Variables

1. **Cognitive load:** Cognitive load has been evaluated through both objective and subjective assessments.
    a) **Objective assessment of cognitive load:** Evaluated using an infrared camera to monitor changes in forehead and nose temperature (difference to the baseline) [1, 256].
    b) **Subjective assessment of cognitive load:** Evaluated using the NASA-TLX questionnaire [92]. The raw version of NASA-TLX was employed, eliminating the weighting process. Participants completed the questionnaire after each scenario.
2. **Visual attention:** We recorded the gaze points falling in the PC area (± 50 pixels) by using a commercial eye tracker. This allowed us to assess whether participants might have been influenced by the presence of the PC, providing valuable insights into their attention and interaction with the interface.
3. **Effect on Performances:** We assessed the impact of using the PC on the driver's performance, examining both longitudinal and lateral effects.
    a) **Performance in Lane-Change Maneuver:** We evaluated the driver's performance in the LC maneuver by quantifying and comparing the number of cone collisions and the deviation from the optimal path. The evaluation employed RSME as a metric, and the number of collisions.
    b) **Performance in Emergency Braking Stop:** To gauge the longitudinal impact of the PC, we conducted an EBS where the distance to the stop line was considered statistically relevant for analysis.

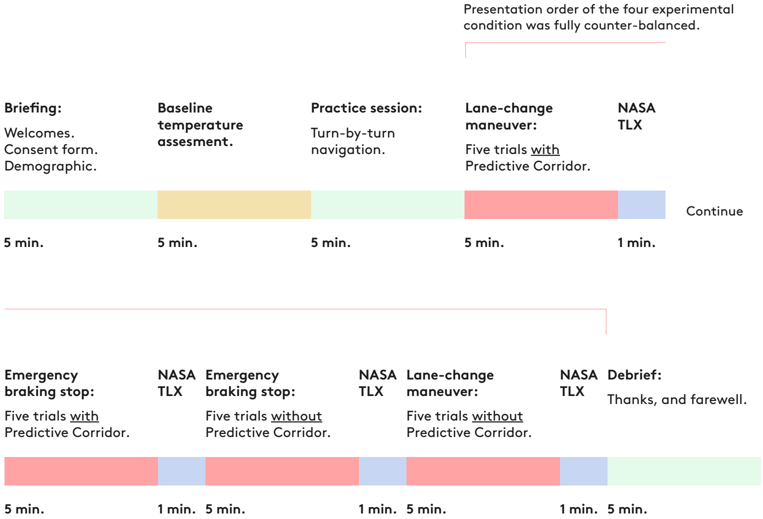


Figure 8.5: User studies procedure line-up. The whole user study consisted of three blocks. In the first part, participants were briefed, and the temperature baseline was assessed. In the second part, we administered two tasks. Lastly, we debriefed the participants.

8.3.4 Procedure

The whole study lasted about 45 minutes. Upon arrival, participants were welcomed and asked to sign a consent and demographics form and were informed about the study aims. Then, we asked participants to relax for 5 minutes while listening to a white noise sound as a calibration condition for sensing. This allowed us to collect their physiological data in a state of relaxation. In the second part of the study, we introduced the prototype and the driving simulator to the participants and let them familiarize themselves with the driving simulator for five minutes. After the practice session, we administrated two tasks, the Lane-Change (LC) maneuver according to the Fritzsche et al. [73] and the Emergency Braking Stop (EBS), both with and without PC. After each task, we asked participants to complete the NASA-TLX [92] questionnaire to assess the perceived cognitive load. We concluded the study by debriefing the participant. During the entire experiment, we recorded the participant's facial

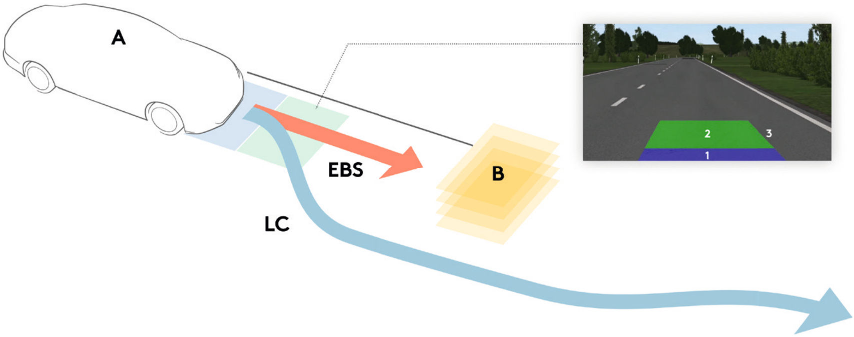


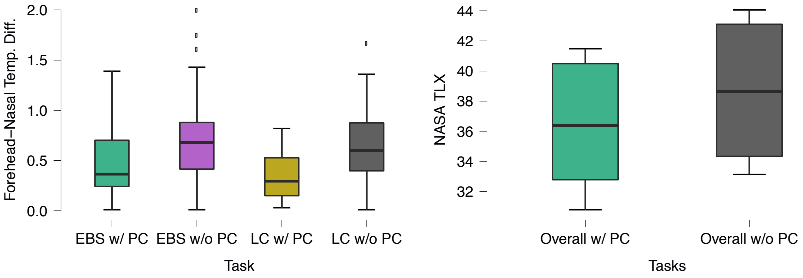
Figure 8.6: In this illustration, we see the two tasks administrated, namely the Emergency Braking Stop (EBS) and the Lane-Change (LC) maneuver. The PC, on the top right, shows to the teleoperator the predicted position (1) of the remote AV and the path of the full braking (2). The width of the PC (3) corresponds to that of the AV.

temperature and eye gaze coordinates. The study was recorded using an RGB thermal camera (cf. Section 8.3.2) while maintaining a room temperature of 25°C.

8.3.5 Results

The results are presented and divided into three main categories. Firstly, the results of the effect of the PC on the participant's cognitive load. Secondly, the results of the effect of the PC on the participant's visual attention, and lastly, on the participant's performances.

The analysis includes one independent variable, the presence of the PC, and two tasks, the LC maneuver and EBS, and both experienced five times with and without PC. Hence, the total number of measurements has been $5 \times 4 \times 32 = 640$. We counterbalanced all conditions to reduce any potential carryover effects and avoid interference and learning effects. The dependent variables considered were the eye movement, facial temperature, and the subjective assessment of the perceived level of mental workload. The participant's performance was logged and considered as a dependent variable. Therefore, we recorded the number of cones collisions for the LC maneuver, whereas the distance to the



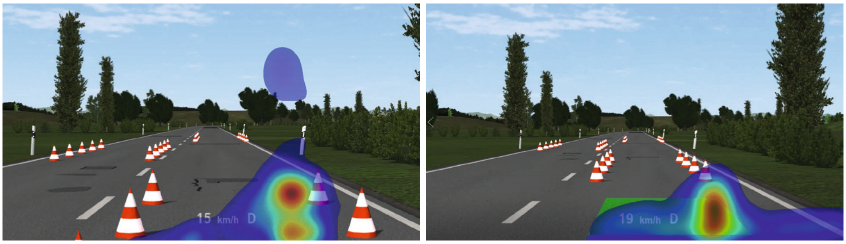
(a) **Forehead/nose Δ temperature** (b) **NASA-TLX**

Figure 8.7: Image (a) shows the forehead and nose temperature (difference to the baseline) over the Lane-Change (LC) maneuver and the Emergency Braking Stop (EBS) task. Both are experienced with and without Predictive Corridor (PC). Image (b) shows the subjective assessment of workload via NASA-TLX.

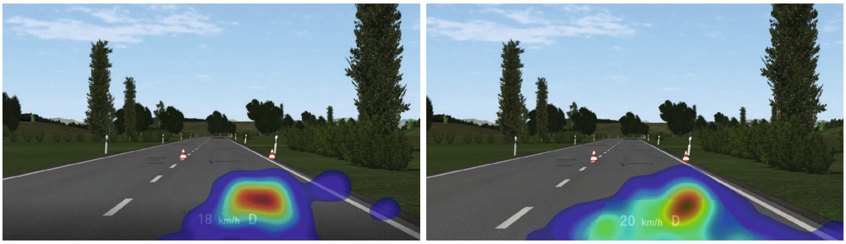
stop line for the EBS task. Lastly, a post-hoc power analysis with $\alpha = .05$, d = .55, and a sample size of n = 32 found an observed power of (1 - β) = .85 for the two-tailed paired sample T-Test analysis.

8.3.5.1 Effect on Cognitive Load

According to the literature, the cognitive load could be assessed by monitoring the facial temperature [1, 256], namely the changes in forehead and nose temperature (difference to the baseline). The increase in the temperature would indicate a higher cognitive load. In this work, we analyzed the effect of using the PC on the forehead-nasal temperatures as an indicator of the experienced cognitive load. During the LC task, on average participant's forehead-nasal temperature increased by 0.31°C (SE: 0.06°C) when driving without the PC. The t-test showed this to be significant, t (31) = 5.18, p < .001, 95% C.I. [.19, .43], d = .92. Similarly, during the Emergency Braking Stop (EBS), the mean forehead-nasal temperature increased by 0.24°C (SE: 0.06°C) when participants performed without the PC. This increment in temperature showed to be significant, t (31) = 4.13, p < .001, 95% C.I. [.12, .36], d = .73. These results have also been confirmed by subjective assessment of cognitive load reported by the participants. According to the NASA-TLX rating scale, on average, participants



(a) **LC without PC** (b) **LC with PC**



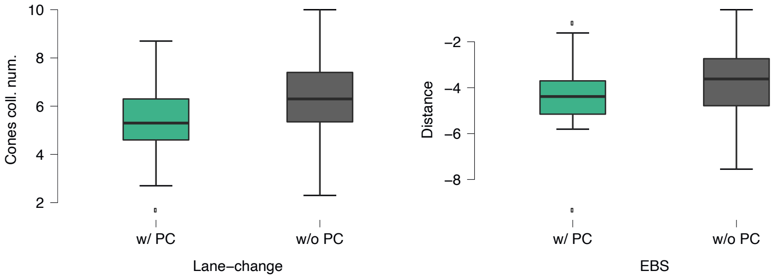
(c) **EBS without PC** (d) **EBS with PC**

Figure 8.8: The image a) shows the gaze patterns of one participant during the Lane-Change (LC) maneuver when performing without the Predictive Corridor (PC). Furthermore, the b) image shows the gaze pattern of the same participant during the LC using the PC. In the image c), the gaze patterns during the Emergency Braking Stop (EBS) task without PC, and in the image d), the gaze plot during the EBS task with the PC. To provide clarity and contextual understanding, it is important to explicitly mention that the images depict a typical snapshot of a random participant's gaze patterns during the maneuver. This clarification ensures that readers understand that the images serve as representative examples of the gaze patterns observed during the maneuver.

experienced 22.5% less workload (SE: 0.59) when performing with PC. The t-test showed this decrease to be significant, t (5) = -3.81, p = .006, 95% C.I. $[-\infty, -1.06]$, d = -1.56. These outcomes indicate that the users experienced less cognitive load when performing with the PC, as the cognitive load was higher in the task without PC (Figure 8.7).

8.3.5.2 Effect on Visual Attention

We investigated differences in gaze distribution by evaluating the gaze data from all participants. We recorded the gaze points falling within a defined PC



(a) **Cones collision number** (b) **EBS stop distance**

Figure 8.9: The image (a) displays the descriptive boxplots of the cones collision number for the Lane-Change (LC) maneuver. In image (b), the distance to the stop line on the Emergency Braking Stop (EBS) task can be seen.

area, which spanned ± 50 pixels from the center of interest, using a Tobii Eye Tracker 4C at an operating frequency of approximately 70 Hz. The recorded x- and y-coordinates of the participants' gaze allowed us to assess the influence of the PC on participants' visual attention and interaction with the interface. Heatmaps were generated from this data to visually represent the concentration and distribution of gaze points across the PC area. The t-test did not show a statistically significant difference in the LC maneuver when operating with and without PC, t (17) = 1.27, p = n.s., 95% C.I. [-207.74, 834.30], d = .3. As seen in Figure 8.8, the difference between the gaze data distribution and the fixation length in both conditions is almost equal. Similarly, over the EBS task, the t-test did not reveal a statistical significance mean effect with and without PC t (17) = -.19, p = n.s., 95% C.I. [-924.21, 768.43], d = -.05. However, there is a difference in gaze data distribution between both conditions, i.e., there is more gaze data in the PC area when it is enabled (Figure 8.8).

8.3.5.3 Effect on Performances

Ultimately, we evaluated the effect of using the PC on the driver's performance. Longitudinal and lateral tests were administrated.

	Cones Collision		Distance	
	w/ PC	w/o PC	w/ PC	w/o PC
Valid	158	159	149	153
Missing	2	1	11	7
Mean	5.892	6.403	-4.551	-3.874
Std. Deviation	2.353	2.442	1.308	1.690
Shapiro-Wilk	0.973	0.973	0.989	0.987
P-value of Shapiro-Wilk	0.003	0.003	0.291	0.161
Minimum	0.000	1.000	-7.743	-7.924
Maximum	11.000	11.000	-1.446	-0.337

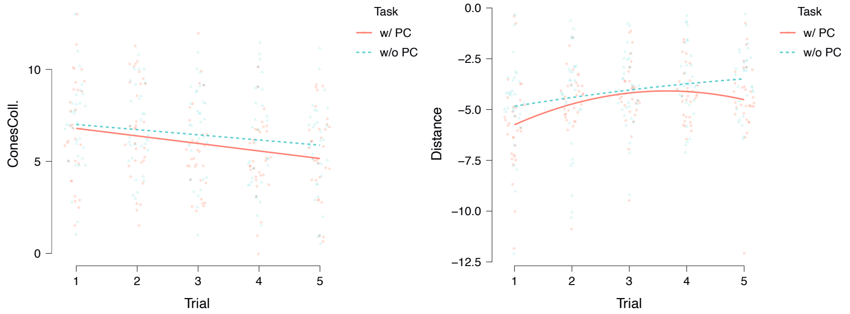
Table 8.1: The descriptive statistics of the participant's driving performance with and without the Lane-Change (LC). In particular, we recorded the number of cones collisions for the Lane-Change (LC) maneuver, whereas the distance to the stop line for the Emergency Braking Stop (EBS) task.

Performances on the Lane-Change maneuver: To evaluate the PC, the LC test course was assigned [73]. The performances of the LC maneuver were tested by counting and comparing the number of cones collisions and the deviation of the optimal path, i.e., the Root Square Mean Error (RSME).

We first start examining the number of cones collisions. On average participants hit 0.74 fewer cones (SE: 0.40) when performing with PC. The t-test showed this decrease to be significant t (31) = -1.87, p < .05. A linear regression was calculated to predict the number of cones collisions based on the number of trials (Figure 8.10a). A significant regression equation was found when performing with PC, r^2 = 0.80, F (1, 3) = 12.24, p < .05, and without PC, r^2 = 0.90, F (1, 3) = 26.19, p = .01. These models showed that the number of cones collisions decreased to 13 cones (trial/participant) with PC and nine cones without PC.

Further analysis of the lateral control was conducted considering the two changes of lanes that the participant was asked to perform for each trial. In particular, the first lane change occurred when the participant had to steer the vehicle to the left. The second lane change occurred when the participant returned to their original lane by steering the vehicle to the right.

Using RSME as a metric, a t-test was conducted to assess participants' deviation from the optimal path (Figure 8.9). On average, participants on the first lane change were deviating 0.09 m (SE: 0.05). The t-test showed this



(a) **A linear regression on the LC** (b) **A linear regression on the EBS**

Figure 8.10: In image (a), a linear regression predicts the number of cones collisions based on the number of trials during the Lane-Change (LC) maneuver. In image (b), a linear regression predicts the distance to the stop line based on the number of trials during the Emergency Braking Stop (EBS) task.

difference to be non-significant, t (28) = 1.83, p = n.s., 95% C.I. [-.01, .19], d = .34. On the contrary, on the second lane change, the mean difference registered was 0.16 m (SE: 0.04). The t-test showed this difference to be significant, t (28) = 3.89, p < .001, 95% C.I. [-.075, .24], d = .72. Pearson's rho correlation analysis, shows significant results between the number of collision and the RSME (of the whole track) with PC, r = 0.41, p < 0.02, and without PC, r = 0.67, p < .001.

Performances of Emergency Braking Stop: To evaluate the PC longitudinally, an EBS was performed. Here the distance to the stop line was considered statistically relevant to the analysis. On average, participants stopped closer to the stop line (0.56 m, SE: 0.24) when performing without PC. The t-test showed this increment to be significant, t (31) = -2.29, p < .01, 95% C.I. [$-\infty$, -.15], d = -.41. A linear regression was calculated to predict the distance to the stop line based on the number of trials (Figure 8.10b). A significant linear regression equation was found without PC, r^2 = 0.97, F (1, 3) = 91.99, p = .002. This has been confirmed also when performing with PC, r^2 = 0.95, F (2, 2) = 19.48, p < .005. As depicted in Figure 8.10b, the linear regression analysis reveals a positive effect without PC and a negative effect with PC.

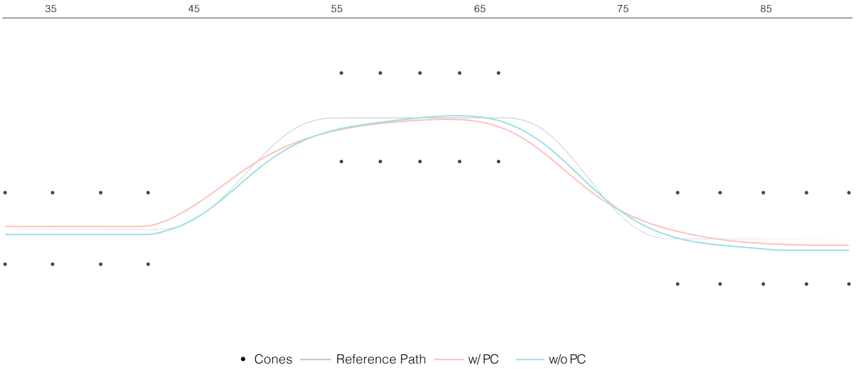


Figure 8.11: In this illustration, we can see the graph of the lateral control conduction with two changes of lanes. In particular, the first lane change occurred when the participant had to steer the vehicle to the left. The second lane change occurred when the participant returned to their original lane by steering the vehicle to the right.

8.4 Discussion

To overcome the challenge of time latency and mitigate low depth perception, the design of ATAS, as the PC, will offer a strategy to support the teleoperator for operating the remote AV securely and ergonomically. In this chapter, we were interested whether the PC could improve the teleoperator user experience. In particular, whether the PC could influence the teleoperator's cognitive load, locus of attention, and operative performances.

When it comes to the subjective measurement of cognitive load, the analysis conducted on all NASA-TLX indicates a significant difference when driving with and without the PC. During the LC maneuver and the EBS task, evidence suggests that with the PC, the participants perceived less mental demand. Inline, the inferred cognitive load from the facial temperature shows that the PC causes less cognitive load, as it led to fewer temperature changes.

However, the assumption of whether the PC affected the participant's visual attention cannot be confirmed. By adding the PC for both situations, LC and EBS, we noticed that the participant's gaze data scattered/diverged over a larger area. We verified this observation for participants who had higher gaze data in the PC area and others whose gaze data did not differ. We saw that the spread of the gaze data is similar across all participants and situations, indicating that adding the PC changes and affects the gaze behavior. However, it does not correlate with the increase of the visual attention around the PC area, not even in the whole scene.

Ultimately, we tested whether the PC influenced the operative performances. Results suggest that the presence of the PC led to fewer collisions. As the linear regression analysis shows, a faster negative flow of the data point was recorded when performing with the PC than without the PC (Figure 8.10a), advising evidence in favor of the PC. Similarly, the analysis on the EBS task shows positive results in favor of the PC. Without the presence of the PC, participants tended to shorten the distance to the stop line by each trial, which did not happen when the PC was shown (Figure 8.10b). Keeping the proper distance is essential to prevent collisions and react accordingly to dynamic objects, as (a) it mitigates low depth perception (cf. Section 2.2.5), and (b) one of the causes of about 50% of the lead-vehicle-stopped crashes is caused by the short distance between the vehicles [155].

In line with prior ATAS research, namely the Predictive Display [35] and Free Corridor [32], overall, the analysis advises evidence in favor of the PC. However, despite this first positive result, the limited finding of the implications of the PC requires more investigations, for example, considering evaluating the PC in dynamic environments, including real test scenarios. This study is the first attempt in this direction.

8.5 Summary

This chapter proposes an ATAS called Predictive Corridor for mitigating low depth perception and time latency in remote vehicle control. The PC combines a novel curvature model with the FC, a prior ATAS. The final result is an augmented reality feature that shows the predicted position of the remote vehicle and the vehicle braking position to the teleoperator. The Predictive Corridor addresses the limitations of previous ATASs by providing a novel curvature model that considers both the teleoperator inputs and the remote vehicle states to improve the accuracy of the prediction. This system could improve the safety and efficiency of remote vehicle control. The PC was found to improve driving performance and reduce cognitive load.

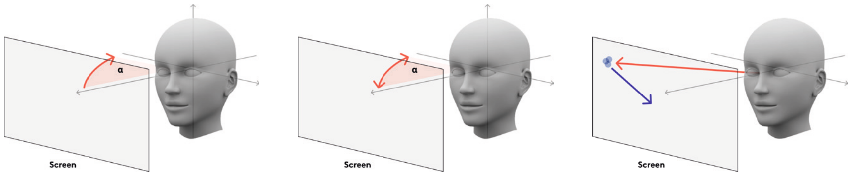
9

Dynamic Camera Control

The previous chapter discussed methods to assist the teleoperator during the driving task. However, in most HMIs for vehicle teleoperation, a common issue is the handling of multiple control devices simultaneously. This might be the case when the teleoperator must steer the vehicle while controlling the remote cameras. On such occasions, the teleoperator must switch between different control devices that could affect subjects' SA and cognitive workload [258]. A solution to this problem has been provided by prior research that implemented interaction gazing techniques in "hands-busy" situations. Inspired by Hild et al., Latif, Zhu et al. [94, 120, 257], ultimately, we focus on developing and evaluating two eye-tracking mechanisms for camera viewpoint control.

9.1 Related Work

Assessing remote environments is one of the most critical tasks in many teleoperation scenarios. By observing the live streaming images the teleoperator can obtain an adequate Situation Awareness (SA). Thus, during the driving maneuver, the operator is responsible for monitoring and understanding of the remote environment and deciding on an appropriate strategy for conducting the



(a) **Zhu et al. [257]** (b) **Zhu et al. [257]** (c) **Zhu et al. [258]**

Figure 9.1: The images show the gaze-based remote camera control designs by Zhu et al. [257] and Zhu et al. [258]. In the image (a), the pan and tilt of the camera are mapped along with the head movement. In image (b), the double arrow stands for the rapid movement the user has to perform to turn on/off the camera to move in the desired direction. Image (c) shows that the teleoperator's gaze can move the image to the center of the display.

AV. In this regard, the teleoperator must perform several tasks simultaneously, driving the vehicle and, if needed, operating the cameras, which may cause physical and cognitive overload.

An extensive body of literature covering the effect of multitasking on the remote operator suggests mitigating the overlapping of motor or sensory interference [144, 145]. The limitations of cognitive processes usually do not allow multiple operations [164] unless the subject already has sufficient practice in performing the tasks [144, 145]. To solve this issue, researchers proposed several "hands-free" camera control concepts by leveraging natural patterns of interactions [226].

For instance, Zhu et al. [257] have adopted different head gestures to control and interact with mining machinery remote cameras. In particular, the authors have introduced two methods for pan, tilt, and zoom. The first method replicates the operator's head movement by panning and tilting the remote camera along with the user's head movement (Figure 9.1a). The interaction stops by repositioning the head and gaze to the center of the screen. Moreover, the user can change the zoom level by bringing the torso closer to screen or farther away from it.

The second method that Zhu et al. [257] proposed refers to quick head movements to switch on or off the pan and tilt function (Figure 9.1b). The process starts when the user moves the head towards one direction and ends when they move back to the original position. The cameras will keep panning and tilting until the user replicates the same movement in the opposite direction. The cameras will stop then at the desired position.

To assess the efficiency of the proposed methods, the authors have led an experiment with a total of ten participants. The subjects were asked to play a table soccer game by controlling two handles. The participant could not see the game directly and neither in its entirety. Instead, it was visible only partially in a monitor placed in front of the operator. The authors set up the camera to have a narrow field of view, and only through the use of the methods described above was it possible to navigate the entire table game. The validation of the proposed methods was obtained by comparing the average of the scores. Results show that the first method provided reliable and comparable performance when operating with more traditional input devices as the keyboard. In contrast, the second method was significantly worse.

In a further study, Zhu et al. [258] proposed a different gaze-driven control method. Based on the position of the teleoperator's gaze, the system will bring the fixated region of interest to the screen center (Figure 9.1c). To assess the feasibility of the introduced method, the authors conducted an experiment with 24 participants. The study design was very similar to the prior and intended to be employed for remote mining machinery. Thus, participants were asked to play a table soccer game with both hands to simulate the hands-busy paradigm. The authors gained the empirical outcomes by comparing the gaze-driven method with a standard gaming control. Again, participants' performance leaned towards the gaze-driven control in this experiment, as a significant difference in scoring goals was achieved.

Similarly, though with different objectives, Hild et al. [94] developed a gaze-based interaction technique for moving target tracking. The interaction technique implemented was similar to the method described by Kotus et al., Zhu et al. [112, 257]. Namely, the video navigation started when the participant's gaze or the computer mouse was located off the display center. To evaluate their method, the authors conducted an empirical evaluation. In the study, 28 participants were asked to watch a three-minute video and keep track of two persons. Subjects could navigate the 360° video data, captured by a drone at an altitude of 30 m, using a computer mouse (baseline condition) or with their gaze. Although the NASA-TLX result shows low scores for both conditions, the statistical analysis reveals that the gaze-based control technique caused significantly less workload than the manual mouse control technique. The authors conclude that a gaze input technique may cause less workload when surveying dynamic targets.

Unlike the literature reported above, other works focus on developing HMI for gaze interaction rather than an eye-tracking mechanism. For instance, Latif [120] proposed the design of an interface for robot control. In their concept, the display was divided into active and non-active regions. The functional areas performed as buttons and were designed to be transparent, not to obstruct the video feedback below. Thus, controlling the robot and monitoring the remote environment were achieved in the same space. In particular, the action of moving the robot was separated from the camera movement. Furthermore, users could enhance their spatial awareness by inferring information from the downward-facing camera visible on the top left corner and gathering more details from the remote environment by turning off the interactive area.

Latif [120] employed three different control devices to evaluate the suggested HMI. Thus, besides the participant's gaze, a navigational task was conducted using a joystick and a computer mouse. In this experiment, Latif [120] observed a degraded participant's performance when asked to only interact with the gaze. When comparing the gaze approach with the joystick condition, participants needed 33% more time and 37% when compared to the computer mouse.

Gaze interaction methods seem to improve operational performances, as long as the teleoperator has to deal with multiple control devices and tasks, e.g., when maneuvering the remote AV and controlling the cameras [94, 257, 258]. Inspired by the prior literature, we present two different gaze-based concepts that adopt eye-tracking mechanisms presented by Hild et al., Latif, Zhu et al. [94, 120, 257]. That is, the prototypes are based on the most recent neural network for Head Pose Estimation, the Fine-Grained Structure Aggregation Network (FSA-Net), which provided top performances in head angle predictions [252]. The following section will introduce the two prototypes, referred to as continuous and discrete (Figure 9.2), and discuss the evaluation results.

9.2 Our Approach

Teleoperation involves more than merely maneuvering a vehicle, and sure enough, it requires interacting with the remote environment. One crucial aspect that makes teleoperation a unique task is that operators are typically engaged in dynamic situations with rapidly changing *"actors"* and problems to solve Hennessy et al. [93]. Drivers adjust decisions based on received sensor data, such as that from remote cameras, depending on their driving experience.

In this regard, supplementary camera viewpoints may enhance operational awareness. Also, on-demand camera viewpoints could be served for other functions, such as avoiding arousal and cognitive overload or enhancing the awareness of the proximity to other artifacts, which according to Marsh and Collett [135], will directly influence the driving and interpersonal driver's behavior.

Thus, besides the three front and the one rear-view mirror camera, two additional side cameras facing the vehicle back left and right sides are provided. The added cameras will close a 360 remote view degree, which we suppose will enhance operators' awareness. The side cameras are not directly visible to the teleoperator, yet they can be called on-demand by turning and pointing the head and gaze on the left- or right-hand screen side. The head gesture is intended to mimic the natural driving behavior when, situated in the vehicle, the driver looks at the rear-view mirrors.

We developed two different control interfaces to call the additional cameras, which we refer as discrete and continuous. In the continuous mode (Figure 9.2a), the camera control is mapped along the yaw head angle, i.e., based on the distance between the user's current head position and the screen center. Might the gaze be not over the central area, the camera view will start moving toward the operator gaze direction. Thus, when the user turns his/her head to the left or the right screen, the remote camera view will move along the same direction showing the further left or right side camera. Additionally, at the bottom center of each video image, operators can see the camera name, i.e., its spatial relationship with the vehicle, e.g., back left.

(a) **Continuous Control Interface**

(b) **Discrete Control Interface**

Figure 9.2: This figure illustrates a comparison of two control interfaces for additional video feedback in teleoperation, i.e., continuous vs. discrete mode. Figure (a) visualises the continuous control mode: here the video feedback is dynamically linked to the operator's head movement relative to the centre of the screen. As the user turns their head to the left or right, away from the central area, the camera view intuitively pans to follow the direction of the operator's gaze, providing a continuous view of the corresponding side camera, like a smoothly guided tour of the environment. A label indicating the camera name and its spatial orientation relative to the vehicle (e.g. 'rear left') is provided at the bottom centre of each video feed for orientation. Figure (b) highlights the discrete control mode: this interface divides the display into specific active zones. When the operator's gaze falls on one of these zones, such as the top left for the rear side camera, the designated camera feed is superimposed on the current front camera view.

In the discrete control mode (Figure 9.2b), the display area is subdivided into active regions that enable the teleoperator to send control commands. For instance, the respective side-back camera will appear when moving the gaze over the screen upper left-hand (or right-hand) side.

The image is assembled on top of the active front cameras, showing, therefore, at the same time the front and the side back cameras. This concept is known as picture-in-picture, and it has been demonstrated to improve operative performance in teleoperation settings [21].

9.3 User Study

To evaluate the proposed gaze interaction, we performed a lab-based driving simulator study. We recorded the participant's driving performance, the perceived cognitive load, and the operator's SA. The user study was designed to answer the research questions:

Do the continuous and the discrete control interface influence the Situation Awareness, cognitive load, and the teleoperator's performances?

To address our research question, we employed a within-subject study design. In addition to the baseline recording, where facial temperature was measured, the sequence in which participants experienced the discrete and the continuous camera control interface, as well as the two scenarios, followed a Latin square design [33].

9.3.1 Implementation Set-Up

We integrated both prototypes into the open-source autonomous driving simulator CARLA [47]. On the server-side, excluding the different clients connected with the Localhost discussed in Chapter 7 (Section 7.3.1), in this implementation set-up, we include the Robot Operating System (ROS) module [178].

ROS, an open-source robotics middleware suite, connects the driver's Head Pose Estimation module with the CARLA environment. The Head Pose Estima-

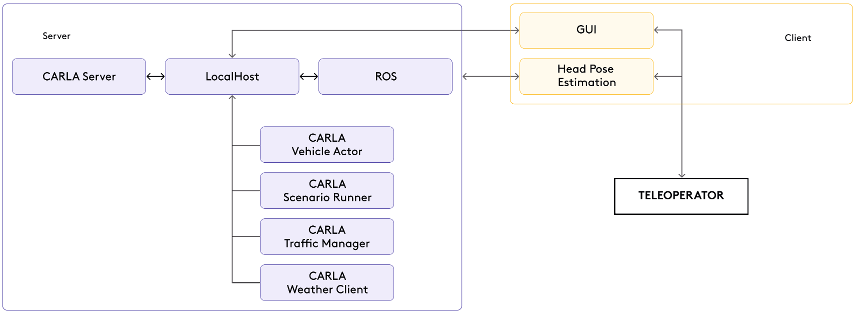


Figure 9.3: The project architecture based on CARLA [47]. The system relies on a client-server concept that provides a flexible Python Application Programming Interface (API) on the client-side, and on the server-side, the rendering allows for the association of multiple clients.

tion module is based on the FSA-Net[1] that calculates the driver's head posture for each video frame, following the roll, yaw, and pitch value in real-time [252], necessary for the control of cameras. The video frame is captured by a camera mounted on top of the central display (cf. Section 9.3.3).

On the client-side, the Graphical User Interface (GUI) module simulates the image video streaming by providing six cameras to the user. In particular, the operator can see three front cameras, one back camera that mimics the rear-view mirror, and two side cameras facing the vehicle left and right sides. Furthermore, the GUI module reads the operator commands from the control devices and applies the relative haptic feedback to the steering wheel (cf. Section 9.3.3). The steering wheel key values were mapped to the vehicle actuators using Pygame[2], a built-in python package.

9.3.2 Participants

For this study, a total of 28 participants (6 female) were actively involved. The participants, with an average age of 28 years (SD = 5.83), ranging from 18 to 48. The participants' driving experience was notable, with an average possession of a driving license for ten years (SD = 5.57). When it came to driving frequency, 43% of participants declared daily driving, 25% drove weekly, 21% drove monthly,

---

[1] https://github.com/aoru45/FSANet.Pytorch

[2] https://pygame.org

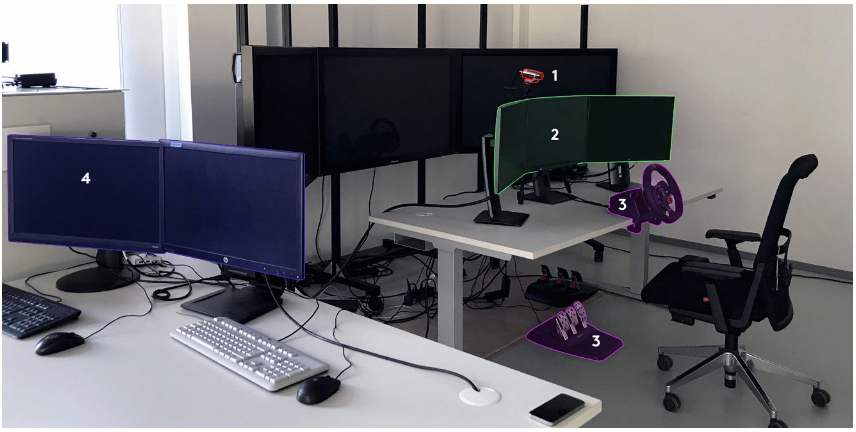


Figure 9.4: The experimental setup. In particular, in this image, we see (1) thermal camera Optris Xi 400, (2) three 23,8-inch LCDs, (3) Logitech G29 steering wheel and pedals, and (4) experimenter computer.

and 11% drove rarely. In addition to driving habits, we assessed participants' remote driving experience and frequency of use. A significant portion, 68%, had prior remote driving experience, including activities such as steering drones, operating RC cars, and similar endeavors. Among these participants, 21% had controlled a remote vehicle several times, 32% had done so on occasion, and 14% had experienced it only once.

9.3.3 Apparatus and Materials

The project is built in CARLA [47]. One computer enabled the transfer of input and output signals between CARLA and the control devices, in our case, the Logitech G29 steering wheel. The same computer rendered the virtual environment visible through three monitors, each 23.8 inch (1920 × 1080 pixel) covering a total horizontal and vertical area of 161.7 × 49.8 cm. Other equipment includes a camera, LarmTek, placed on top of the central display. The LarmTek camera provides a Full Hd resolution (1080p), with a rate of 30 fps and a covering angle of 90 degrees. Through this camera, it was possible to estimate the participant's head position.

Additionally, a compact infrared camera was installed on a tripod beyond the screens to assess the participant's facial temperature. The Optris Xi 400[1] camera measures temperatures between -20 and 900°C, and an optical resolution enables a spot-distance ratio of up to 390:1. The optical resolution of the camera is 382 × 288 pixels, with a frame rate of 80 Hz and thermal sensitivity NETD of 80 mK. The ambient temperature was kept at 25°C constant throughout the experiment to avoid noise in the data set. The Optris PI[2] connects the software with the camera. We marked the areas of interest (forehead and nose) and employed the built-in data extraction function to save temperature values [2].

9.3.4 Study Design

The analysis includes two independent variables, the discrete and the continuous camera control interface, and two scenarios. In this exploration of our research, we scrutinize pivotal components that form the foundation of our study. We examine the discrete and the continuous camera control interface carefully manipulating them to observe their specific impacts, and delve into the corresponding dependent variables.

9.3.4.1 Independent Variables

1. **HMIs:** Participants interacted with the remote environment with the discrete and the continuous camera control interface.
    a) **Continuous camera control:** In continuous control mode, users can smoothly shift their gaze to either side of the screen (e.g., left-hand side), initiating corresponding video feedback movement.
    b) **Discrete camera control:** In discrete control mode, users can move their gaze to a specific area, such as the left, and it will appear as a supplementary camera overlaying the existing video feedback.
2. **Scenario:** The test scenarios were designed to simulate an obstacle avoidance maneuver and intended to experience longitudinal and lateral control

---

[1] https://optris.com/optris-xi-400

[2] https://optris.com/software-development-kits

Figure 9.5: The driving scenario we administrated during the user study. In particular, in this image, the ego-vehicle encounters a broken standing vehicle ahead.

of the vehicle under clear weather conditions and daylight. In both scenarios, the ego-vehicle encountered an obstacle, and thus the teleoperator was asked to perform an avoidance maneuver. These courses were modeled after the pre-crash typology defined by the National Highway Traffic Safety Administration [155].

a) **Scenario (a):** In one scenario, participants saw a truck blocking the street sideways. The situation required the participant to avoid the obstacle by steering to the left, briefly establishing the left lane, and then returning to the right lane. This scenario was experienced five times during the study (5 trials/participant).

b) **Scenario (b):** In the other scenario, participants encountered a broken standing vehicle ahead (Figure 9.5). The situation required the participant to navigate around the obstacle by steering to the left, briefly establishing the left lane, and then returning to the right lane. This scenario was also experienced five times during the study (5 trials/participant).

9.3.4.2 Dependent Variables

1. **Cognitive load:** Cognitive load has been evaluated through both objective and subjective assessments.

   a) **Objective assessment of cognitive load:** Evaluated using an infrared camera to monitor changes in forehead and nose temperature (difference to the baseline) [1, 256].

    b) **Subjective assessment of cognitive load:** Evaluated using the NASA-TLX questionnaire [92]. The raw version of NASA-TLX was employed, eliminating the weighting process. Participants completed the questionnaire after each scenario.

2. **Situation Awareness:** Assessed through the SAGAT questionnaire [58]. Participants provided responses after each task, thus removing the freezing period. The SAGAT was quantified as a percentage and averaged across questions targeting a demanding level of SA.
3. **Camera usage:** We recorded and analyzed the usage of supplementary cameras throughout the study. Specifically, we counted the instances when participants actively utilized the supplementary cameras to enhance their situational awareness and inform their driving decisions.
4. **Performances:** To assess participant performance, we monitored and quantified two driving errors. This included instances of crossing solid lines and any collisions that occurred during the experimental tasks. By documenting these errors, we aimed to gain insights into the impact of different conditions on participants' driving proficiency and the effectiveness of the supplementary cameras in mitigating errors.

9.3.5 Procedure

The entire study duration was approximately one hour, as illustrated in Figure 9.6. Upon participants' arrival, they received a warm welcome and were invited to complete a consent and demographics form, accompanied by an overview of the study's objectives. Subsequently, participants were introduced to the two prototypes, experiencing both the discrete and continuous camera control interfaces. A brief relaxation period of 3 minutes, featuring white noise, served as a calibration condition and served as the baseline condition for thermal imaging. This allowed us to collect physiological data in a relaxed state.

In the second phase of the study, participants engaged in two test courses, each utilizing either the discrete or continuous camera control interface. These tasks were administered in distinct periods, counterbalanced to ensure a comprehensive and unbiased evaluation. Following each task, participants provided

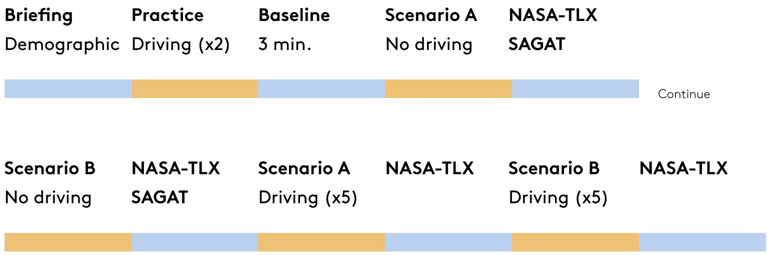


Figure 9.6: Overview of the user study flow. The user study was conducted in a single session and comprised a series of consecutive activities. Initially, participants received an orientation briefing and engaged in a practice session to familiarize themselves with the prototypes. Following this, participants underwent baseline temperature assessments for physiological workload measurement. Subsequently, the main evaluation consisted of two driving scenarios, one for each camera control interface (discrete and continuous). These scenarios were administered in a counterbalanced order using a Latin Square design to control for order effects and participants shifted between driving and non-driving tasks within the same session. After completing the tasks, a debriefing session concluded the study.

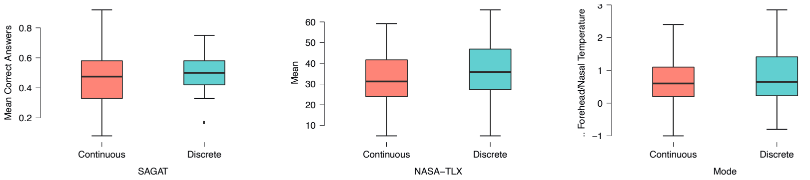
feedback through the NASA-TLX questionnaire [92] and the SAGAT questionnaire [58] to assess perceived cognitive load and SA, with the latter being administered after the initial two tasks. The study concluded with a debriefing session, expressing gratitude to the participants for their valuable contribution.

9.3.6 Results

The results are presented and divided into two main categories. We will first discuss the cognitive load and SA results before the subjects maneuvered the vehicle, namely when they were asked to assess the remote environment. Then we presented the analysis of the perceived cognitive load, the camera usage, and participants' performance while controlling the vehicle.

9.3.6.1 Before the Driving Maneuver

Results of the SAGAT: A paired-samples t-test was conducted to determine whether the level of SA assessed via the SAGAT questionnaire varied between the discrete or continuous camera control prototype. Preliminary data screening showed that the data did not deviate from normality. The groups did not differ, t (27) = 0.34, $p > .05$, 95% C.I. [-.09, .12], d = .06. The SAGAT mean score for



(a) **SAGAT's answers** (b) **NASA-TLX** (c) **Forehead/nose Δ temp.**

Figure 9.7: The image shows the boxplots summarising the pre-task assessments of Situation Awareness and workload. In particular, Figure (a) shows the distribution of scores from the SAGAT questionnaire, which was administered before participants began the driving task to assess their baseline situational awareness. Figure (b) shows the subjective workload ratings provided by participants using the NASA-TLX questionnaire, which was also administered prior to the driving task to assess participants' perceived workload in anticipation of the activity. Figure (c) shows the results of an objective workload assessment using thermal imaging to analyse physiological markers of stress, such as changes in facial temperature patterns, measured before participants engaged in the driving task.

the discrete camera control prototype (M = .50, SD = .15) was not statistically significantly different than the continuous camera control (M = .49, SD = .2) and the effect size was trivial (d = .06). These findings suggested that an equal level of SA was obtained with both prototypes introduced, the discrete and continuous camera control.

Results of the NASA-TLX: A paired-samples t-test was conducted to determine whether participants subjective cognitive load varied when interacting with the discrete or the continuous camera control prototype. Preliminary data screening showed that the NASA-TLX did not deviate from normality. The groups did not differ, t (27) = 1.20, p > .05, 95% C.I. [-3.21, 12.14], d = .23. The mean for the NASA-TLX when using the discrete camera control prototype (M = 37.14, SD = 14.99) was not statistically significantly different than the continuous camera control (M = 32.68, SD = 13.37) and the effect size was small (d = .23). These findings suggested that the two camera control modes were perceived as equally demanding.

Results of the thermal imaging: A paired-samples t-test was conducted to determine whether participants' objective cognitive load varied when interacting either with the discrete or the continuous camera control prototype. Preliminary data screening showed that the forehead-nasal temperature (difference to the baseline) did deviate from normality; however, not sufficient to justify the data transformation, Shapiro-Wilks test (p > .02), and outliers were eliminated. The groups did not differ, t (21) = 0.2, p > .05, 95% C.I. [.03, -], d = .005. The mean for the forehead-nasal temperature when using the discrete camera control prototype (M = 0.78, SD = .96) was not statistically significantly different than the continuous camera control (M = .69, SD = .89) and the effect size was small (d = .005). These findings suggested that the two camera control modes were objectively equally cognitive demanding.

9.3.6.2 During the driving maneuver

Results of the camera usage: A paired-samples t-test was conducted to determine whether participants preferred to interact more with the discrete or continuous camera control. Preliminary data screening showed that usage of the cameras began to deviate from normality yet not sufficient to justify the data transformation, Shapiro-Wilks test (p > .002). The groups differed significantly, t (17) = 3.95, p = .001, 95% C.I. [2.03, 6.69], d = .93. The mean for the discrete camera control usage (M = 4.78, SD = 4.84) was statistically significantly higher than the continuous camera control (M = 0.42, SD = 0.42) and the effect size was large (d = .93). These findings supported the idea that the discrete camera control, when driving, was more used than the continuous camera control mode.

Results on performances: A paired-samples t-test was conducted to determine whether participants made more errors when driving with the discrete or continuous camera control prototype. Preliminary data screening showed that the errors did not deviate from normality. The groups differed significantly, t(27) = 2.32, p = .03, 95% C.I. [.03, .46], d = .44. The mean error when using the discrete camera control prototype (M = .18, SD = .26) was statistically significantly lower than the continuous camera control (M = .42, SD = .44) and the effect size was

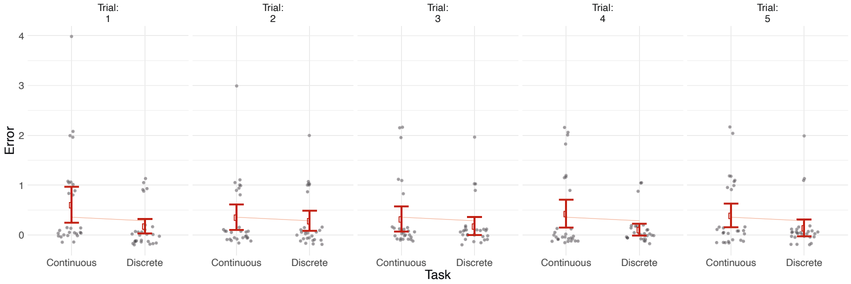
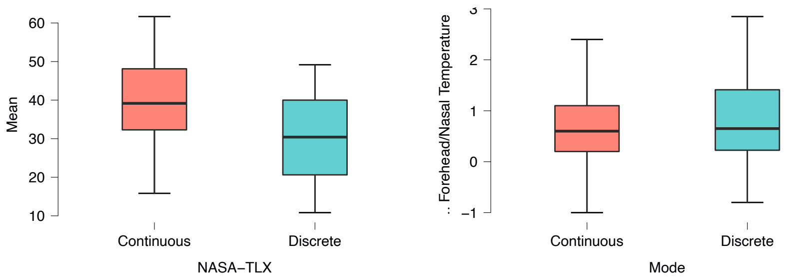


Figure 9.8: This plot shows, for both control interfaces (continuous and discrete), the driving errors made by the participants during the five trials. The red intervals show the standard errors with a dot for the participants' driving error mean.

medium (d = .44). These findings supported the idea that participants who drove with the discrete camera control tended to make fewer driving errors than those driving with the continuous camera control method.

Results of the NASA-TLX: A paired-samples t-test was conducted to determine whether participants perceived more cognitive load when interacting with the discrete or continuous camera control prototype. Preliminary data screening showed that the NASA-TLX did not deviate from normality. The groups differed significantly, t (27) = 2.95, p = .006, 95% C.I. [3.12, 17.36], d = .56. The mean for the NASA-TLX when using the discrete camera control prototype (M = 29.82, SD = 11.77) was statistically significantly lower than the continuous camera control (M = 40.06, SD = 12.43) and the effect size was medium (d = .56). These findings supported the idea that the discrete camera control was perceived as less cognitively demanding than the continuous camera control method, at least when driving.

Results of the thermal imaging: A paired-samples t-test was conducted to determine whether participants' objective cognitive load varied when interacting either with the discrete or the continuous camera control prototype. Preliminary data screening showed that the forehead-nasal temperature (difference to the baseline) did not deviate from normality; however, outliers were found and



(a) **NASA-TLX** (b) **Forehead/nose Δ temp.**

Figure 9.9: Image (a) shows the descriptive boxplots of the subjective workload was evaluated via NASA-TLX questionnaire during the driving task. In image (b), the descriptive boxplots of the objective workload was evaluated via thermal imaging during the driving task.

eliminated. The groups differed significantly, t(21) = 2.03, p > .03, 95% C.I. [.04, -], d = .43. The mean for the forehead-nasal temperature when using the discrete camera control prototype (M = .15, SD = 1.13) was statistically significantly lower than the continuous camera control (M = .58, SD = .89) and the effect size was medium (d = .43). These findings supported the idea that the discrete camera control was objectively less cognitively demanding than the continuous camera control method, at least when driving.

A correlation & linear regression analysis: A Pearson rho correlation was conducted to assess the relationship between the subjective and objective cognitive load before and during the driving task when operating with the discrete control prototype. Preliminary analyses showed a linear relationship with both variables normally distributed, as assessed by Shapiro-Wilk test (Table 9.1). There was a significant positive correlation between the NASA-TLX administrated before and after the driving task, r = .63, p < .001. Also, a significant positive correlation between the forehead/nasal temperature administrated before and after the driving task, r = .51, p = .02.

A linear regression established that the NASA-TLX administrated before the driving task could statistically significantly predict the cognitive load during the driving task, F(1,26) = 17.26, p < .001; and NASA-TLX administrated before

	NASA-TLX		Temperature	
	Before	During	Δ Before	Δ During
Valid	28	28	26	22
Missing	1	1	3	7
Mean	37.143	29.821	0.781	0.150
Std. Deviation	14.985	11.773	0.962	0.716
Shapiro-Wilk	0.988	0.950	0.973	0.937
P-value of Shapiro-Wilk	0.978	0.195	0.694	0.169
Minimum	5.000	10.830	-0.800	-1.200
Maximum	65.830	49.170	2.850	1.500

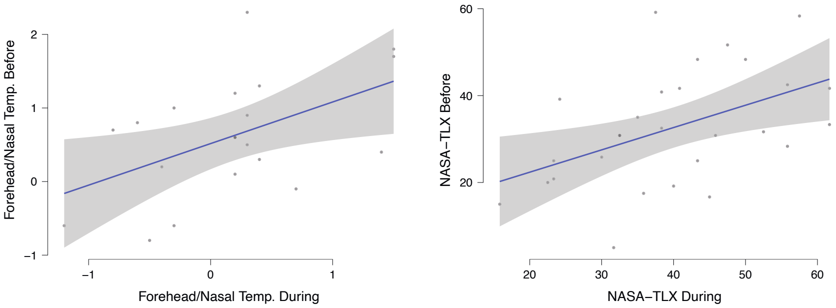
Table 9.1: Descriptive statistics of the subjective and objective assessment of workload assessed via NASA-TLX and thermal imaging.

accounted for 38% of the explained variability in cognitive load during the driving task. The regression equation was: predicted driving NASA-TLX During = 11.39 + (0.5 x NASA-TLX Before).

Similarly, a linear regression established that the forehead/nasal temperature recorded before the driving task could statistically significantly predict the forehead/nasal temperature during the driving task, F(1,18) = 6.23, p = .02; and forehead/nasal temperature recorded before accounted for 22% of the explained variability in cognitive load during the driving task. The regression equation was: Forehead/Nasal Temperature During = -.11 + (0.46 x Forehead/Nasal Temperature Before).

9.4 Discussion

Motivated by the existing hands-busy problem in teleoperation, we developed and evaluated two gaze-driven HMI for remote AVs controlling. Based on the operator's head position and the region of interest where he or she is staring, we could enable/disable additional camera images. In this regard, the method allowed us to interact naturally with supplementary viewpoints. Also it will allow to reduce bandwidth if the design objective requires it, as a constraint of the real-world, namely, the unavailability of high-bandwidth mobile network, is a common problem in telerobotics.



(a) **Thermal Imaging Linear Regression** (b) **NASA-TLX Linear Regression**

Figure 9.10: Image (a) and (b) shows the linear regression that predicts the cognitive index when performing the driving task. In image (a), the outcome is based on the participant's forehead/nasal temperature difference when asked, before driving, to examine the traffic scene. In image (b), the outcome is based on the participant's NASA-TLX mean when asked to examine the traffic scene before driving.

The results of our experimental investigation provide a comprehensive overview of how different HMIs fare in terms of performance, usage, cognitive load, and SA. Our findings reveal a distinct preference for the discrete control interface, especially when it comes to tasks that require guiding a vehicle. It seems that prior to the vehicle maneuvering activities, users did not show a bias toward any particular prototype. However, the process of maneuvering allowed us to observe a marked difference—usage patterns shifted significantly, performance heightened, and cognitive load altered, although SA remained unaffected.

Furthermore, through empirical evaluation, we gained valuable insights into multiple aspects of each HMI. Performance assessments and cognitive load evaluations pointed to the superiority of the discrete control interface over its continuous counterpart. Users tasked with maneuvering the vehicle found this interface allowed for a smoother experience, with fewer errors and a reduced mental strain. In contrast, the continuous interface, which might appear more intuitive at first glance, led to disorientation among the subjects, particularly when controlling cameras and the vehicle. This interface was associated with higher rates of driving errors and an increased cognitive workload, hinting at why its usage was less frequent.

Interestingly, the discrete control interface did not necessitate the same demand for orientation capacity, making it the more user-friendly option between the two. This finding echoes the conclusions drawn in previous studies [21, 50], highlighting the efficacy of the Picture-In-Picture technique, as it allows another video channel to appear as an overlay on the screen, providing users with additional context without overwhelming them.

Interestingly, the regression equation reveals that the subjects who experienced a high cognitive load level before driving had recorded 38% less workload while conducting the vehicle with the discrete control camera prototype. This finding has also been confirmed by the thermal imaging analysis, which counted 22% less of facial temperature changing. We hypothesize that participants who have more thoroughly examined the remote traffic situation before the driving task had less mental strain during the driving task. We hypothesize that the sum of factors such as (a) the stress caused by the ToR, (b) the elaboration of the information of an unfamiliar environment, (c) making driving decisions based on perceived information, appeared to generate more cognitive load, then stabilize the vehicle during the driving maneuver. An alternative interpretation of this finding could be that participants who had more thoroughly examined the remote traffic situation before the driving task had less cognitive workload during the driving task. These assumptions are also supported by Stapel et al. [215], demonstrating that monitoring automation sets a significant driver mental workload.

Our study provides initial insights that serve as a promising foundation for the development of improved HMIs for camera control in scenarios where the user's hands are otherwise engaged. Recognizing that usability and efficiency are paramount in such situations, the outcomes of our research shed light on the design characteristics that merit further refinement. This could involve integrating more intuitive feedback mechanisms, simplifying control schemes, or enhancing the interface's responsiveness to user input.

Further, it is critical to augment our preliminary results with extensive data collection. Further research should extend beyond our initial review, encompassing a broader set of variables that might influence HMIs performance in real-world conditions. An in-field evaluation, for instance, could not only corroborate our findings but also unearth nuanced challenges and opportunities that laboratory conditions might not reveal. Such practical insights will be vital to iteratively refine our proposed approach, moving us closer to an HMIs that is both effective and user-friendly under a wider array of circumstances.

Eventually, the vision ahead is to create an interaction paradigm that intuitively aligns with the operator's cognitive processes by preventing errors before they occur. With this continuous iteration to researching and experimenting with innovative designs, we are yielding HMIs that are reliable, efficient, and ready to meet the demanding dynamics of of teleoperation.

9.5 Summary

In this chapter, we presented two eye/head-tracking mechanisms for camera viewpoint control for remote vehicle operations. These mechanisms have been designed to alleviate the cognitive workload on the operator by allowing them to control the cameras with their head movements, rather than having to manipulate multiple control devices simultaneously. The mechanisms have been inspired by previous research on "hands-free" camera control and have been tested and found to be effective in reducing cognitive workload and improving SA. The first mechanism maps the pan of the camera to the movement of the user's head, while the second mechanism allows the user to switch on and off the superimposed camera views. The user study showed that the second camera control mode was preferred by subjects and led to fewer errors when driving a remote vehicle. This suggests that the discrete mode may be more effective for controlling the vehicle in a remote environment.

Reflections

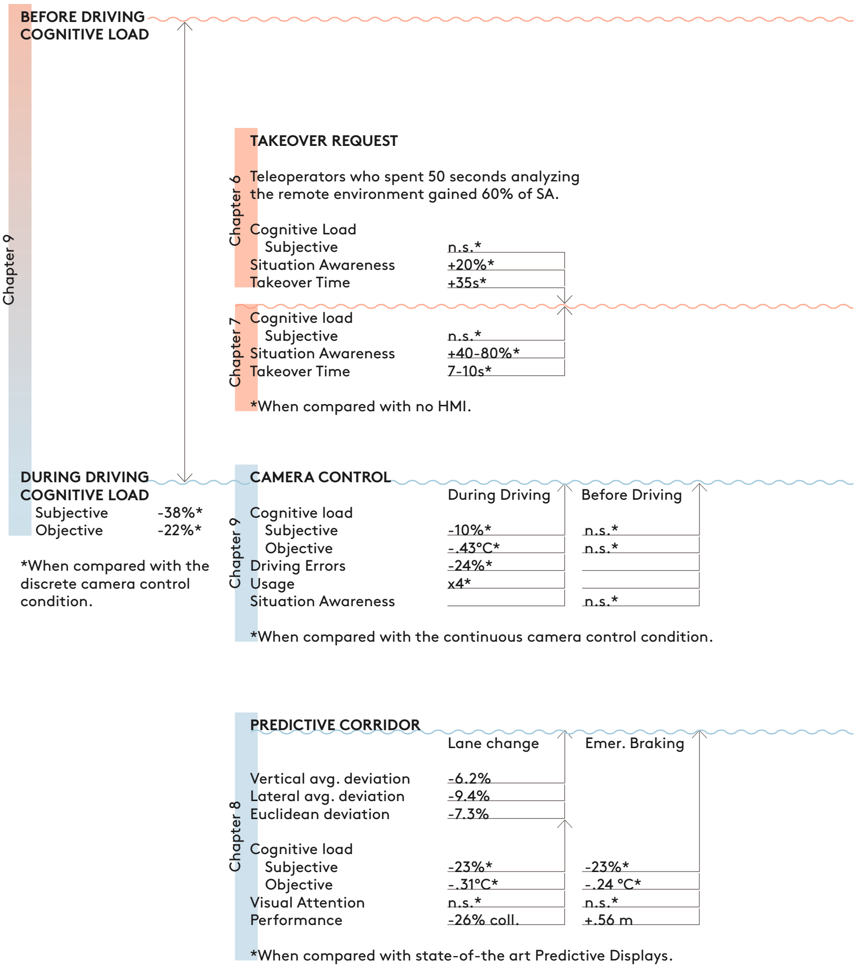
10

Conclusion and Outlook

This thesis aims to enhance the safety and ergonomics of teleoperator control in AV operation by developing user-interface and interaction concepts that improve spatial perception and decision-making. Our work, grounded in cognitive science and particularly focused on SA and cognitive workload, provides a theoretical basis for our empirical research. A summary of our contributions, reflections on their limitations, and future research directions are outlined in this chapter, offering insight and direction for other researchers in the field.

10.1 Summary

Our design journey for an optimal teleoperator workplace began with reviewing existing literature, complemented by Subject Matter Expert analysis to extract practical user requirements. Chapter 5 presents a comprehensive framework for SA requirements and its application for AV teleoperation interfaces, guiding our design process.



BEFORE DRIVING COGNITIVE LOAD

DURING DRIVING COGNITIVE LOAD

Subjective	-38%*
Objective	-22%*

*When compared with the discrete camera control condition.

TAKEOVER REQUEST

Teleoperators who spent 50 seconds analyzing the remote environment gained 60% of SA.

Cognitive Load	
Subjective	n.s.*
Situation Awareness	+20%*
Takeover Time	+35s*

Cognitive load	
Subjective	n.s.*
Situation Awareness	+40-80%*
Takeover Time	7-10s*

*When compared with no HMI.

CAMERA CONTROL

	During Driving	Before Driving
Cognitive load		
Subjective	-10%*	n.s.*
Objective	-.43°C*	n.s.*
Driving Errors	-24%*	
Usage	x4*	
Situation Awareness		n.s.*

*When compared with the continuous camera control condition.

PREDICTIVE CORRIDOR

	Lane change	Emer. Braking
Vertical avg. deviation	-6.2%	
Lateral avg. deviation	-9.4%	
Euclidean deviation	-7.3%	
Cognitive load		
Subjective	-23%*	-23%*
Objective	-.31°C*	-.24 °C*
Visual Attention	n.s.*	n.s.*
Performance	-26% coll.	+.56 m

*When compared with state-of-the art Predictive Displays.

Figure 10.1: A visual summary of the research findings.

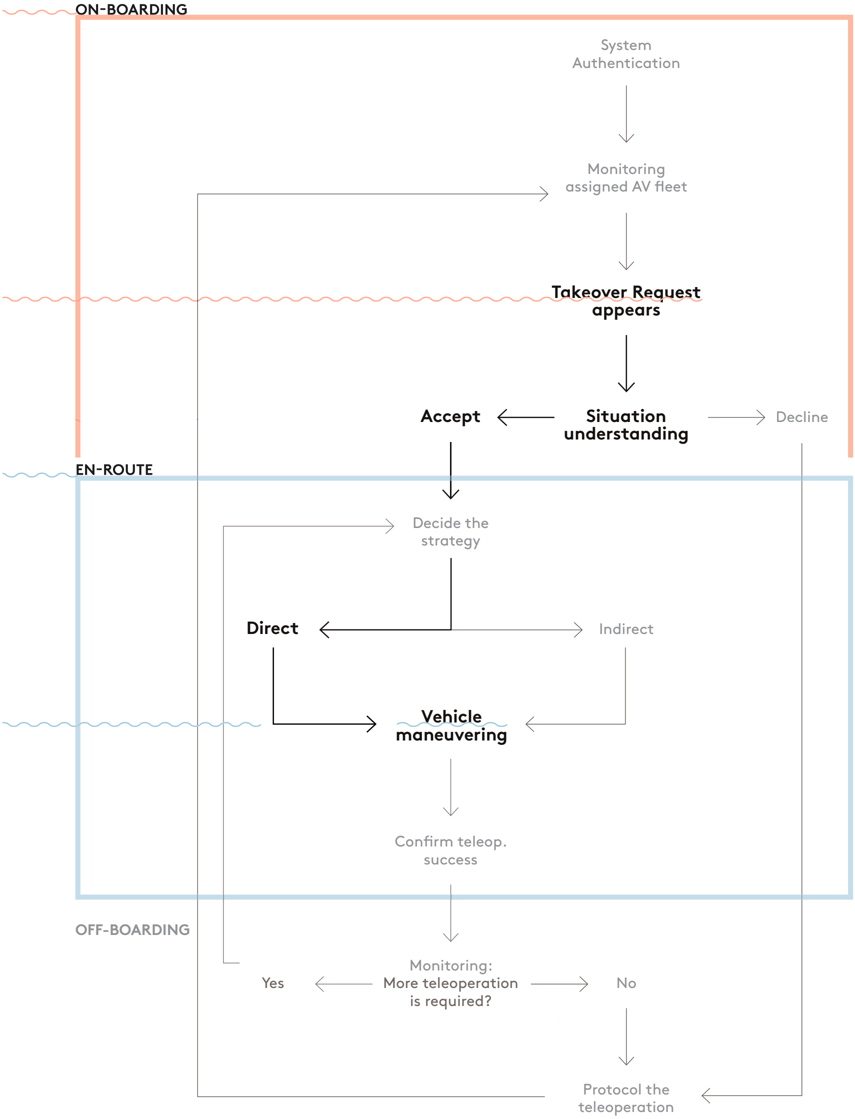


Figure 10.2: The chart flow diagram of the system we developed. In bold are the topics handled in this research.

10.1.1 On-Boarding: Takeover Request

In Chapter 6 and Chapter 7, based on the Subject Matter Expert analysis we laid an initial basic interface layout for remote a Takeover Request (ToR) Human-Machine Interface (HMI). The HMIs we proposed were designed over two display areas, arranged vertically and horizontally in front of the teleoperator. Both areas performed as input and output devices. In particular, teleoperators received visual feedback from both areas; yet, they could send touch commands from the horizontal display area. We presented an egocentric view of the remote environment in the vertical display. Whereas in the horizontal display area, the HMI showed an exocentric perspective of the remote environment, providing sensor-fusion data information (i.e., point-cloud LiDAR and map) to the teleoperator.

In this regard, in Chapter 6, we were interested in understanding how long it takes an operator to understand the remote environment and what level of SA and cognitive load can be achieved. Overall, we learned that the user necessitates a high Takeover Time (ToT) when dealing with remote vehicles, i.e., a mean 35 seconds, based on the HMI of interest (cf. Figure 10.1). Comparing the results we reported in Chapter 6 with those previously observed by Eriksson and Stanton [65], we noted higher subject responses. We assume that this difference has caused due to the operational mode of our subjects, as they were asked to operate the vehicle remotely, i.e., detached from the actual environment. Moreover, we could not reduce the subjective perceived cognitive load, as the findings suggested that an equal cognitive demand was perceived. However, despite these findings, we witnessed a +20% higher SA when participants were asked to assess the remote environment with our proposed HMIs. We viewed these results as promising, although the limited research of remote ToR restricts our ability to compare our findings with other authors.

In Chapter 7, we tried to approach the problem of high ToTs from a different direction. That is, by employing methods that have been developed to predict the AV disengagement, we could request the takeover up to seven seconds in advance. We extended this method by controlling the spatial nature of the fault prediction showing the teleoperator where the fault is located in the environment. This strategy allowed operators to focus directly on the anomaly and take quick countermeasures. Thus compared to the previous HMIs we evaluated, we radically reduced the information processed by the operator by showing, only the fault prediction and the egocentric view of the environment. Eventually, we observed that subjects could achieve significantly higher SA levels, +40%, within the first seven seconds and +80% in ten seconds. These results suggest that limited functionality to guarantee simplicity could improve SA and ToTs. However, similarly to the results reported during the ToR analysis in Chapter 6, we could not observe a significant decrease in cognitive load as participants could only experience the ToR scenarios for seven and ten seconds, which might have been insufficient to record subjective perceived cognitive load changes.

10.1.2 En-Route: Teleoperator Assistance Systems

In Chapter 6 and Chapter 7, our attention was devoted to the development of HMIs that could effectively communicate the ToR to the operator in an intuitive fashion. The goal of these chapters was to support the operator in gathering necessary information prior to initiating a driving maneuver. Contrastingly, Chapter 8 and Chapter 9 shifted focus to address issues related to time latency, depth perception, and challenges that arise when the operator's hands are occupied, commonly referred to as the 'hands-busy' problem. Here, we discussed user interface and interaction concepts designed to aid the teleoperator during the driving maneuver.

A clear distinction between these sets of chapters also lies in the methods adopted for measuring the operator's cognitive load. In Chapter 6 and Chapter 7, we employed subjective assessment tools, utilizing the NASA-TLX questionnaire to capture cognitive load. However, this approach had limitations as we were unable to detect notable fluctuations in cognitive load. To address this, Chapter 8 and Chapter 9 introduced a novel approach using thermal imaging to measure cognitive load. This technique, as substantiated in Chapter 4, allowed us to estimate changes in cognitive load by observing variations in facial temperature, which offers an unobtrusive, objective, and real-time assessment. This advancement represents a significant step forward in our ability to understand and measure the cognitive demands placed on operators within these contexts.

10.1.2.1 Predictive Corridor

To support the teleoperator in complex driving maneuvers, in Chapter 8, we presented the Predictive Corridor (PC), an Advanced Teleoperator Assistance System (ATAS). The PC offers a strategy to overcome the challenge of time latency and mitigate operators' depth perception. In particular, the PC builds upon a new curvature model we proposed and the Free Corridor (FC) described by Chen [32]. In prior works, authors rely on a curvature model that highly depends on time latency, as they consider the remote vehicle inputs only (e.g., speed and yaw rate). In contrast, we suggested a novel curvature model that assesses both the teleoperator inputs and the remote vehicle states to calculate the curvature value. When we compared it with the state-of-the-art prediction method, the performance results of our approach indicated that the curvature model had 6.2% less vertical average absolute deviation, 9.4% less lateral average absolute deviation, and 7.3% less euclidean absolute deviation.

Based on these results, we attached to the novel curvature model the FC [32]. The PC has been shown as a semi-transparent layer on top of the delayed video image. In particular, the PC presents the predicted vehicle position, considering the time delay, and displays the vehicle braking path. Eventually, we were interested in whether the PC could improve the teleoperator's user experience, i.e., whether the PC could influence the teleoperator's cognitive load, locus of attention, and operative performances. The subjective cognitive load assessment suggested that the PC could reduce cognitive load by 23%, and the thermal imaging assessment recorded fewer temperature changes, -.31°C for the Lane-Change maneuver and -.24°C for the Emergency Braking Stop task. The participants' performance also confirmed these results. We documented fewer cones collision (-26%) and a significant increase in distance to the stop line (+.56m), when the PC was active. However, the hypothesis tested to assess whether the PC affected the participants' visual attention could not be confirmed. Overall, we could conclude that the analysis advises evidence favoring the PC, as long as we can create a stable latency connection from and to the AV.

10.1.2.2 Dynamic Camera Control

As mentioned, dedicated control devices have been envisioned to manage complex driving maneuvers, i.e., teleoperators have been supported by a steering wheel and pedals to assess hazardous situations in challenging environments. However, adopting additional input devices implies that operators have to control multiple devices simultaneously, e.g., when the teleoperator must steer the vehicle while controlling the remote cameras. Thus, motivated by this problem, known in the literature as the *hands-busy* problem, in Chapter 9, we developed and evaluated two gaze-driven HMIs for remote AVs controlling. Based on the operators' head position and the region where they are staring, we could enable/disable additional camera images. In particular, in the continuous control mode, when users move their gaze over one or the other side of the screen, e.g., over the left-hand side, the view starts smoothly to move, showing the related video feedback. In the discrete control mode, when users move their gaze over a specific area, e.g., over the left, it will appear on top of the existing video feedback a supplementary camera. Both methods allow the interaction with supplementary viewpoints naturally and, if required, save bandwidth when the cameras are inactive.

The experimental results to determine whether the proposed methods differed significantly across cognitive load, performance, usage, and SA favor the discrete control mode. In particular, we could reduce subjective cognitive load by 10% and record fewer changes in forehead-nasal temperature by -.43°C. We could also observe fewer driving errors, 24%, e.g., crossing solid lines and collisions, and four times of usage of the discrete control mode. These statements are true, at least when guiding the vehicle. Findings show that an equal SA and cognitive load level was obtained before the vehicle maneuvering, i.e., when the vehicle was standing, and the operator assessed the situation. It appears that when users are stressed with multiple tasks, such as driving and camera controlling, the discrete control mode was overall superior. Eventually, the correlation and regression analysis we conducted revealed that the subjects who experienced a high cognitive load before the driving task had recorded 38% less subjective workload while conducting the vehicle and 22% less of facial temperature change. We hypothesize that the sum of factors such as (a) the stress caused by the ToR, (b) the elaboration of the information of an unfamiliar environment, (c) making driving decisions based on perceived information, appeared to generate more cognitive load, then stabilize the vehicle during the driving maneuver. An alternative interpretation of this finding could be that participants who had more thoroughly examined the remote traffic situation before the driving task had less cognitive workload during the driving task. These assumptions are also supported by Stapel et al. [215], demonstrating that monitoring automation sets a significant driver mental workload. We would, however, note that these conclusions are valid when assessing the remote environment without additional ToR HMI, yet with the discrete control camera prototype.

Moreover, we would like to emphasize once again that despite the compensatory methods discussed in this thesis and more generally in the literature concerning teleoperation, concepts to improve operators' spatial perception, state awareness, decision-making processes, i.e., concepts to improve Situation Awareness, remain a challenge when developing HMIs for teleoperated AV.

10.2 Contributions

Now that we have reached the end, we would like to summarize the contribution of this research. Overall our contribution can be reviewed into three points: First, we strived to uncover user interface requirements for remote teleoperation. Second, we conceptualized, designed, and developed novel user interfaces and interactions for teleoperated vehicles. Third, we gathered results that will help with the design of future user interfaces for remote teleoperation.

The first contribution is the extensive qualitative and quantitative requirement analysis. These requirements have yet to be explicitly tailored for the use-cases to which we refer. Instead, the requirements we have collected from the literature analysis and the experts cover an area of development that is not necessarily task-dependent nor dependent on a specific teleoperation model (manual, supervised, or automatic). In this regard, the requirement analysis was meant to assist the teleoperator in making appropriate decisions and facilitate the content transformation into design assets.

The second contribution is the conceptualizing, designing, and developing novel user interface and interaction concepts for teleoperated vehicles. We focused on developing HMIs that could intuitively explain the Takeover Request to the operator. Then, we designed an Advanced Teleoperator Assistance System to overcome the time latency and low depth perception. In particular, the system forecasts the vehicle position considering the time delay from/to the vehicle and shows the vehicle braking path in an emergency stop. Finally, as managing a remote vehicle might demand various input devices and feedback displays, we designed a method to interact naturally with supplementary viewpoints.

The third contribution is the empirical evidence we have generated. The experimental results of our study provide insights that will hopefully influence the design of future user interfaces for remote teleoperation. These findings invite a variety of investigations that will build on the knowledge we have amassed.

This work, while conclusive in its current form, indeed sets the foundation for an exciting and innovative vision for the future. We envisage a new generation of HMIs, characterized by interfaces that are not only responsive and predictive but that also facilitate a higher level of situational awareness and operational safety. The advanced teleoperation systems of tomorrow will likely feature adaptive interfaces that adjust in real-time to operator needs, augmented reality elements for enhanced depth perception, and even AI-driven predictive models that anticipate operational challenges before they arise. This vision, inspired by the critical insights we have drawn through our research, aims to delineate a realm of possibility where teleoperation is more intuitive, efficient, and aligned with human cognitive processes—pushing the envelope of what is currently achievable in remote vehicular control.

10.3 Limitations

This thesis set the focus on the development of interface and interaction concepts for remote ToR and vehicle management during the driving maneuver. To some extent, the most apparent limitation that might be encountered is the design of too specific applications that may carry the danger of result misinterpretations that are predisposed to be generalized. This might be even more true when testing with a limited set of scenarios. Our investigations did not always include dynamic environments and never real test scenarios. This limitation cannot ensure a complete conclusion of how we could improve remote operational awareness. We are aware that this work only allows us to compare the effect that the design proposals have had on the operator's SA and cognitive load and whether these ultimately serve their purpose correctly or not. This limits the possibility of discussing how effective our design is compared to others. However, in some measure, the results discussed in this work might provide valuable insights within the field of automotive teleoperation research.

Then, we would like to critically reflect upon the small number of subjects who participated in our evaluations. In all the qualitative and quantitative studies we conducted, we were able to evaluate and analyze data from a total of 121 subjects. Participants recruited during the studies were all students or IT specialists. Furthermore, the data skewed towards male subjects, as only 31% were female and none of others genders (83 male and 38 female). Therefore, the statistical verifiability of the results could have been biased.

In closing, the most explicit limitations that might have been formed during the course of this thesis could be summarized in the following key point: First, erroneous inferences arising from specific application contexts. Second, the resulting limited generalization of the results. Third, the relatively small number of subjects who participated in our evaluations. These limitations can unquestionably be covered in future work.

10.4 Future Work

The progress we pursue in AV teleoperation opens new design spaces to explore, along with new opportunities. Most urgent works should consider integrating and validating a broader set of scenarios, including dynamic environments and real test scenarios, to guarantee a comprehensive understanding of how to model remote Situation Awareness. We are also confident that an evaluation in a real life situation could bring further undiscovered insights to the proposed approaches. Likewise, as we could only examine the designated interfaces in specific application contexts, validations of the HMIs along with other tasks should not be neglected. This newly gathered data, which will complement the one we reviewed, might eventually confirm our results.

Coming studies should also consider the subjects' variance and size as we identify the need to increase the number of participants to generate additional statistically verifiable results. Furthermore, we suggest including more diversity. Variation in expertise, age, gender, disabilities, inter-nationality, and technological affinity will enrich the spectrum of the design study quality. Additional data on participants' traits, such as affinity towards anxiety or personality profiles, should also be included into the analyses. Prior research has shown that stress can affect sympathetic responses, thereby influencing the results [210], as do different personality types on specific driving behaviors [176]. The addition of these elements in the data collection process could help identify additional driver effects and provide a more accurate evaluation measure.

Beyond the broader sets of scenarios and the diversity of the studies, the most evident future works that might be noticed in the Figure 10.2 are, for instance, the design of systems for coordinating and monitoring remote AVs, the development for indirect maneuvering, and strategy to handover the control to the vehicle. Eventually, strategies for direct/indirect maneuvering and takeover from/to the vehicle might adopt innovative solutions from related research fields, such as Intelligent User Interfaces.

For instance, future teleoperation design concepts might have an adaptive character, adaptive to the actual driving situation and the driver's mental state. The task success then will not depend only on the quality of the HMI but also on the operator's cognitive workload. Driver's mental state, including Situation Awareness, defines the capability to perceive and correctly interpret a complex situation and anticipate its future development [58, 60], particularly important in dynamic and complex environments, as inner-city scenarios, to act efficiently [75]. That is, to prevent the operator from making the wrong decisions [216].

A solution to this problem might be given from intelligent systems that adapt to the operator's workload. For instance, as Abdelrahman et al. [1] stated, thermal imaging may be employed to discriminate the operator's workload in real-time. The great advantage that arises by using thermal cameras is the opportunity to maintain an adequate level of cognitive workload. A high-level workload could trigger a potential transition of the commands to another operator or adopt indirect controls. In this respect, the conduction of the vehicle can be adapted to the driver's mental state. As Kohlmorgen et al. [110] underline, adaptive adjustment of authority between operators and vehicles could enhance operative performance, especially in driving contexts, since it strongly relies upon the operator's mental state.

In light of these promising developments to enhance operative performance through adaptive systems, such as using thermal imaging for workload monitoring and dynamic adjustment of control authority, it is crucial to envision how these innovations could transform the design and operation of a remote AVs. The following strategic recommendations arise naturally from the core findings of our research, providing a roadmap to apply our studies' insights in practical, impactful ways. Each recommendation draws from empirical evidence, addressing specific challenges and opportunities in teleoperation that combine to form a comprehensive, integrated approach. The aim is to build upon the strengths of adaptive technologies and intelligent systems, ensuring that operator cognitive workload is optimized and safety is paramount. By advancing these recommendations, we can shape the future of teleoperation in a manner that is not just reactive to immediate needs, but proactively prepares for the evolving landscape of AV operation:

1. **Dual-Display Touch Command Interfaces:** Implement interfaces that provide both egocentric and exocentric views of the Autonomous Vehicle environment. There is evidence that this setup improves Situation Awareness.
2. **Predictive Takeover Request Systems:** Develop systems that allow pre-emptive takeover requests to reduce Autonomous Vehicle, which has been shown to significantly improve Situation Awareness.
3. **Teleoperator Assistance Systems:** Integrate real-time thermal imaging to objectively assess and manage cognitive load, enhancing operator performance and mitigating fatigue.

4. **Advanced Predictive Corridor:** Make use of the Predictive Corridor method to address latency and perception challenges, as it has been shown to optimize operator response.
5. **Gaze-Driven Multimodal Human-Machine Interfaces:** Adopt gaze-driven Human-Machine Interfaces for efficient management of multiple controls, for reducing cognitive strain during complex maneuvers.
6. **Inclusive Design Approaches:** Embrace diversity in scenario design, reflecting a broad range of operator profiles to ensure robustness and generalizability of the teleoperation system.
7. **Interdisciplinary Research and Development:** Foster continuous cross-disciplinary collaboration that leverages insights from cognitive psychology, ergonomics, and artificial intelligence.
8. **Continuous Learning and Adaptation:** Embed machine learning and data analytics to evolve system capabilities based on operator experiences.
9. **Operator-Centric Design Validation:** Pursue validation of the Human-Machine Interface and teleoperation systems with real-world testing to ensure the designs meet actual operational needs.
10. **Investment in Future Visions:** Support research oriented toward next-generation teleoperation technologies, including indirect vehicle maneuvering and smart vehicle-operator authority sharing.

10.5 Closing Remarks

Throughout this thesis, we have explored user interfaces and interaction concepts for autonomous vehicle teleoperation. We have conducted both theoretical and practical research on this topic, including building prototypes and conducting user studies. We have emphasized the importance of sketching and prototyping to explore and formalize ideas. The work has been motivated by a desire to design initial HMI concepts for remote civil operations.

Thus the effect of different user interfaces and interaction concepts on daily teleoperator tasks has been examined and informed the design of future work. We hope our experiences and findings will inspire others interested in designing interaction techniques for vehicle teleoperation.

Bibliography

[1] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, Florian Alt. “Stay cool! understanding thermal attacks on mobile-based user authentication.” In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* 2017, pp. 3751–3763. DOI: `10.1145/3025453.3025461`.

[2] Yomna Abdelrahman, Anam Ahmad Khan, Joshua Newn, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, Albrecht Schmidt. “Classifying attention types with thermal imaging and eye tracking.” In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3.3 (2019), pp. 1–27. DOI: `10.1145/3351227`.

[3] Bettina Abendroth, Ralph Bruder. “Die Leistungsfähigkeit des Menschen für die Fahrzeugführung.” In: *Handbuch Fahrerassistenzsysteme.* Springer, 2009, pp. 4–14. DOI: `10.1007/978-3-658-05734-3_1`.

[4] Hillary Abraham, Chaiwoo Lee, Samantha Brady, Craig Fitzgerald, Bruce Mehler, Bryan Reimer, Joseph Coughlin. “Autonomous vehicles and alternatives to driving: trust, preferences, and effects of age.” In: *Proceedings of the Transportation Research Board 96th Annual Meeting (TRB’17).* 2017. DOI: `https://trid.trb.org/view/1438862`.

[5] Robert Altschaffel, Tobias Hoppe, Sven Kuhlmann, Jana Dittmann. “Simulation of Automotive Security Threat Warnings to Analyze Driver Interpretations and Emotional Transitions.” In: *International Conference on Computer Safety, Reliability, and Security.* Springer. 2014, pp. 47–58. DOI: `10.1007/978-3-319-24255-2_5`.

[6] Wendy Amai, Jill Fahrenholtz, Chris Leger. “Hands-free operation of a small mobile robot.” In: *Autonomous robots* 11.1 (2001), pp. 69–76. DOI: `10.1023/A:1011260229560`.

[7] Bernhard Anzengruber, Andreas Riener. “"FaceLight" potentials and drawbacks of thermal imaging to infer driver stress.” In: *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 2012, pp. 209–216. DOI: `10.1145/2390256.2390292`.

[8] Edoardo Ardizzone, Antonio Chella, Roberto Pirrone. “An architecture for automatic gesture analysis.” In: *Proceedings of the working conference on Advanced visual interfaces.* 2000, pp. 205–210. DOI: `10.1145/345513.345317`.

[9] Anthony Aretz. “The design of electronic map displays.” In: *Human Factors* 33.1 (1991), pp. 85–101. DOI: `10.1177/001872089103300107`.

[10] Henrik Artman. “Team situation assessment and information distribution.” In: *Ergonomics* 43.8 (2000), pp. 1111–1128. DOI: `10.1080/001401300050084905`.

[11] Henrik Artman, Christer Garbis. “Situation awareness as distributed cognition.” In: *Proceedings of ECCE.* Vol. 98. Citeseer. 1998.

[12] Gregory Bedny, David Meister. “Theory of activity and situation awareness.” In: *International Journal of cognitive ergonomics* 3.1 (1999), pp. 63–72. DOI: `10.1207/S15327566IJCE0301_5`.

[13] Cag Belderbos. “Authority transition interface: A human machine interface for taking over control from a highly automated truck.” MA thesis. Delft University of Technology, Netherlands, 2015.

[14] Johannes Beller, Matthias Heesen, Mark Vollrath. “Improving the driver–automation interaction: An approach using automation uncertainty.” In: *Human factors* 55.6 (2013), pp. 1130–1141. DOI: `10.1177/0018720813482327`.

[15] Benjamin Blankertz, Claudia Sannelli, Sebastian Halder, Eva Hammer, Andrea Kübler, Klaus-Robert Müller, Gabriel Curio, Thorsten Dickhaus. “Neurophysiological predictor of SMR-based BCI performance.” In: *Neuroimage* 51.4 (2010), pp. 1303–1309. DOI: `10.1016/j.neuroimage.2010.03.022`.

[16] Cheryl Bolstad, Mica Endsley. “The effect of task load and shared displays on team situation awareness.” In: *proceedings of the human factors and ergonomics society annual meeting.* Vol. 44. 1. 2000, pp. 189–192. DOI: `10.1177/154193120004400150`.

[17] Michael Borenstein. “Effect sizes for continuous data.” In: ed. by Harris M. Cooper, Larry V. Hedges, Jeff C. Valentine. 2nd ed. Vol. 2. New York: Russell Sage Foundation, 2009, pp. 221–235. ISBN: 978-0-87154-163-5.

[18] Patrick Bourdot, Brian Katz, Antoine Tarault, Jean-Marc Vézien. “The Use of 3D-Audio in a Multi-Modal Teleoperation Platform for Remote Driving/Supervision.” In: *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*. Audio Engineering Society. 2007.

[19] Holger Caesar, Varun Bankiti, Alex Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, Oscar Beijbom. “nuscenes: A multimodal dataset for autonomous driving.” In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 11621–11631. DOI: `10.1109/CVPR42600.2020.01164`.

[20] Gloria Calhoun, John Fontejon, Mark Draper, Heath Ruff, Brian Guilfoos. “Tactile versus aural redundant alert cues for UAV control applications.” In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 48. 2004, pp. 137–141. DOI: `10.1177/154193120404800130`.

[21] Gloria Calhoun, Heath Ruff, Austen Lefebvre, Mark Draper, Antonio Ayala. “"Picture-in-Picture" Augmentation of UAV Workstation Video Display.” In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 51. 2007, pp. 70–74. DOI: `10.1177/154193120705100204`.

[22] Pascale Carayon, Peter Hancock, Nancy Leveson, Ian Noy, Laerte Sznelwar, Geert Van Hootegem. “Advancing a sociotechnical systems approach to workplace safety–developing the conceptual framework.” In: *Ergonomics* 58.4 (2015), pp. 548–564. DOI: `10.1080/00140139.2015.1015623`.

[23] Carlos Carreto, Daniel Gêgo, Luis Figueiredo. “An eye-gaze tracking system for teleoperation of a mobile robot.” In: *Journal of Information Systems Engineering & Management* 3.2 (2018), p. 16. DOI: `10.20897/jisem.201816`.

[24] Stephen Casner. “Mitigating the loss of navigational awareness while flying with GPS and moving map displays under VFR.” In: *International Journal of Applied Aviation Studies* 6.1 (2006), p. 121.

[25] Stephen Casner. “The effect of GPS and moving map displays on navigational awareness while flying under VFR.” In: *International Journal of Applied Aviation Studies* 5.1 (2005), pp. 153–165.

[26] Jennifer Casper, Robin Murphy. “Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center.” In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 33.3 (2003), pp. 367–385. DOI: `10.1109/TSMCB.2003.811794`.

[27] Sheryl Chappell, Kendra Dunlap. “Incorporating operator situation awareness into the design process: A key ingredient for future combat unmanned rotorcraft.” In: *Annual Forum Proceedings-American Helicopter Society*. Vol. 62. American Helicopter Society, INC. 2006, p. 1927.

[28] Gary Charness, Uri Gneezy, Michael Kuhn. “Experimental methods: Between-subject and within-subject design.” In: *Journal of Economic Behavior & Organization* 81.1 (2012), pp. 1–8. DOI: `10.1016/J.JEBO.2011.08.009`.

[29] Jessie Chen, Paula Durlach, Jared Sloan, Laticia Bowens. “Human-robot interaction in the context of simulated route reconnaissance missions.” In: *Military Psychology* 20.3 (2008), pp. 135–149. DOI: `10.1080/08995600802115904`.

[30] Jessie Chen, Ellen C. Haas, Michael Barnes. “Human performance issues and user interface design for teleoperated robots.” In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37.6 (2007), pp. 1231–1245. DOI: `10.1109/TSMCC.2007.905819`.

[31] Jessie Chen, Jennifer Thropp. “Review of low frame rate effects on human performance.” In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 37.6 (2007), pp. 1063–1076. DOI: `10.1109/TSMCA.2007.904779`.

[32] Tito Lu Tang Chen. “Methods for improving the control of teleoperated vehicles.” PhD thesis. Technische Universität München, Germany, 2015.

[33] Richard Yoonsik Chin, Bruce Lee. *Principles and practice of clinical trial medicine*. eng. Amsterdam Boston: Elsevier/Academic Press, 2008. ISBN: 978-0-12-373695-6. DOI: `10.1016/B978-0-12-373695-6.X0001-4`.

[34] Roger W Cholewiak, Amy Collins. “The generation of vibrotactile patterns on a linear array: Influences of body site, time, and presentation mode.” In: *Perception & Psychophysics* 62.6 (2000), pp. 1220–1235. DOI: `10.3758/BF03212124`.

[35] Frederic Chucholowski. “Eine vorausschauende Anzeige zur Teleoperation von Straßenfahrzeugen.” PhD thesis. Technische Universität München, Germany, 2015.

[36] Charles Cohen, Lynn Conway, Dan Koditschek. “Dynamical system representation, generation, and recognition of basic oscillatory motion gestures.” In: *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition.* IEEE. 1996, pp. 60–65. DOI: `10.1109/AFGR.1996.557244`.

[37] On-Road Automated Driving Committee. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.* 2021. DOI: `10.4271/J3016_201609`.

[38] Lynn Conway, Charles Cohen. “Video mirroring and iconic gestures: enhancing basic videophones to provide visual coaching and visual control.” In: *IEEE Transactions on Consumer Electronics* 44.2 (1998), pp. 388–397. DOI: `10.1109/30.681955`.

[39] Mary Cummings, Stephanie Guerlain. “The tactical tomahawk conundrum: Designing decision support systems for revolutionary domains.” In: *SMC’03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483).* Vol. 2. IEEE. 2003, pp. 1583–1588. DOI: `10.1109/ICSMC.2003.1244638`.

[40] Luiz Marcio Cysneiros, Majid Raffi, Julio Cesar Sampaio do Prado Leite. “Software transparency as a key requirement for self-driving cars.” In: *2018 IEEE 26th international requirements engineering conference (RE).* IEEE. 2018, pp. 382–387. DOI: `10.1109/RE.2018.00-21`.

[41] Shreyansh Daftry, Sam Zeng, Andrew Bagnell, Martial Hebert. “Introspective perception: Learning to predict failures in vision systems.” In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE. 2016, pp. 1743–1750. DOI: `10.1109/IROS.2016.7759279`.

[42] Rudolph Darken, Helsin Cevik. “Map usage in virtual environments: Orientation issues.” In: *Proceedings IEEE virtual reality.* IEEE. 1999, pp. 133–140. DOI: `10.1109/VR.1999.756944`.

[43] Rudolph Darken, Kurt Kempster, Barry Peterson. “Effects of streaming video quality of service on spatial comprehension in a reconnaissance task.” In: *Proceedings of I/ITSEC.* 2001.

[44] James Davis, Christopher Smyth, Kaleb McDowell. “The effects of time lag on driving performance and a possible mitigation.” In: *IEEE Transactions on Robotics* 26.3 (2010), pp. 590–593. DOI: `10.1109/TRO.2010.2046695`.

[45] Krista DeLeeuw, Richard Mayer. “A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load.” In: *Journal of educational psychology* 100.1 (2008), p. 223. DOI: `10.1037/0022-0663.100.1.223`.

[46] Chris Dede. “Immersive interfaces for engagement and learning.” In: *science* 323.5910 (2009), pp. 66–69. DOI: `10.1126/science.1167311`.

[47] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, Vladlen Koltun. “CARLA: An open urban driving simulator.” In: *Conference on robot learning*. PMLR. 2017, pp. 1–16. DOI: `10.48550/arXiv.1711.03938`.

[48] Jennifer Virginia Draper, Steven Yale Handel, Christopher Hood, Camilla Kring. “Three experiments with stereoscopic television: When it works and why.” In: *Conference Proceedings 1991 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE. 1991, pp. 1047–1052. DOI: `10.1109/ICSMC.1991.169826`.

[49] John Draper, David Kaber, John Usher. “Speculations on the value of telepresence.” In: *CyberPsychology & Behavior* 2.4 (1999), pp. 349–362. DOI: `10.1089/cpb.1999.2.349`.

[50] Mark Draper, Gloria Calhoun, Jeremy Nelson, Heath Ruff. “Synthetic vision overlay concepts for uninhabited aerial vehicle operations: Evaluation of update rate on four operator tasks.” In: *Proceedings of the NATO RTO Human Factors and Medicine Panel Symposium, HFM-135, held in Biarritz, FR*. 2006, pp. 9–11.

[51] Mark Draper, Gloria Calhoun, Heath Ruff, David Williamson, Timothy Barry. “Manual versus speech input for unmanned aerial vehicle control station operations.” In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 47. 2003, pp. 109–113. DOI: `10.1177/154193120304700123`.

[52] David Drascic. “Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing.” In: *Proceedings of the Human Factors Society Annual Meeting*. Vol. 35. 1991, pp. 1367–1371. DOI: `10.1177/154193129103501906`.

[53] Jill Drury, Jean Scholtz, Holly Yanco. “Awareness in human-robot interactions.” In: *SMC’03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance*. Vol. 1. IEEE. 2003, pp. 912–918. DOI: `10.1109/ICSMC.2003.1243931`.

[54] Jill Drury, Stacey Scott. “Awareness in unmanned aerial vehicle operations.” In: *The International C2 Journal* 2.1 (2008), pp. 1–10.

[55] Jan Dul, Ralph Bruder, Peter Buckle, Pascale Carayon, Pierre Falzon, William Marras, John Wilson, Bas van der Doelen. "A strategy for human factors/ergonomics: developing the discipline and profession." In: *Ergonomics* 55.4 (2012), pp. 377–395. DOI: `10.1080/00140139.2012.661087`.

[56] Francis Durso, Carla Hackworth, Todd Truitt, Jerry Crutchfield, Danko Nikolic, Carol Manning. "Situation awareness as a predictor of performance for en route air traffic controllers." In: *Air Traffic Control Quarterly* 6.1 (1998), pp. 1–20. DOI: `10.2514/ATCQ.6.1.1`.

[57] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, Heinrich Hussmann. "Bringing transparency design into practice." In: *23rd international conference on intelligent user interfaces*. 2018, pp. 211–223. DOI: `10.1145/3172944.3172961`.

[58] Mica Endsley. "Design and evaluation for situation awareness enhancement." In: *Proceedings of the Human Factors Society annual meeting*. Vol. 32. 1988, pp. 97–101. DOI: `10.1177/154193128803200221`.

[59] Mica Endsley. *Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition*. 2nd. USA: CRC Press, Inc., 2011. ISBN: 9781135729493. DOI: `10.1201/9780203485088`.

[60] Mica Endsley. "Designing for situation awareness in complex systems." In: *Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment*. 2001, pp. 1–14. DOI: `10.1201/9780203485088`.

[61] Mica Endsley. "Direct measurement of situation awareness: Validity and use of SAGAT." In: *Situation awareness analysis and measurement* 10 (2000), pp. 147–173. DOI: `10.4324/9781315087924-9`.

[62] Mica Endsley. "Final report: Situation awareness in an advanced strategic mission (NOR DOC 89-32)." In: *Hawthorne, CA: Northrop Corporation* (1989).

[63] Mica Endsley. "Towards a new paradigm for automation: Designing for situation awareness." In: *IFAC Proceedings Volumes* 28.15 (1995), pp. 365–370. DOI: `10.1016/S1474-6670(17)45259-1`.

[64] Mica Endsley, Michelle Robertson. "Situation awareness in aircraft maintenance teams." In: *International Journal of Industrial Ergonomics* 26.2 (2000), pp. 301–325. DOI: `10.1016/S0169-8141(99)00073-6`.

[65] Alexander Eriksson, Neville Stanton. “Takeover time in highly automated vehicles: noncritical transitions to and from manual control.” In: *Human factors* 59.4 (2017), pp. 689–705. DOI: 10.1177/0018720816685832.

[66] Alia Fisher, Patricia McDermott, Shane Fagan. “Bandwidth allocation in a military teleoperation task.” In: *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2009, pp. 287–288. DOI: 10.1145/1514095.1514179.

[67] Frank Ole Flemisch, Klaus Bengler, Heiner Bubb, Hermann Winner, Ralph Bruder. “Towards cooperative guidance and control of highly automated vehicles: H-Mode and Conduct-by-Wire.” In: *Ergonomics* 57.3 (2014), pp. 343–360. DOI: 10.1080/00140139.2013.869355.

[68] Terrence Fong. “Collaborative control: A robot-centric model for vehicle teleoperation.” PhD thesis. Carnegie Mellon University, Pittsburgh, Pennsylvania, 2001.

[69] Terrence Fong, Francois Conti, Sébastien Grange, Charles Baur. “Novel interfaces for remote driving: gesture, haptic, and PDA.” In: *Mobile Robots XV and Telemanipulator and Telepresence Technologies VII*. Vol. 4195. International Society for Optics and Photonics. 2001, pp. 300–311. DOI: 10.1117/12.417314.

[70] Benjamin Franz, Jan Haccius, Dennis Stelzig Krombholz, Matthias Pfromm, Michaela Kauer, Bettina Abendroth. “Evaluation of the SAGAT method for highly automated driving.” In: *The Proceedings of the IEA Congress*. 2015.

[71] Lex Fridman, Li Ding, Benedikt Jenik, Bryan Reimer. “Arguing machines: Human supervision of black box AI systems that make life-critical decisions.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2019, pp. 0–0. DOI: 10.1109/CVPRW.2019.00173.

[72] Manel Frigola, Josep Fernandez, Joan Aranda. “Visual human machine interface by gestures.” In: *IEEE International Conference on Robotics and Automation*. Vol. 1. IEEE. 2003, pp. 386–391. DOI: 10.1109/ROBOT.2003.1241626.

[73] Egbert Fritzsche, Thorsten Leonhardt, Hakim Mkinsi, Fabiola Caragol Rivera. *ISO 3888-1:2018 Passenger cars – Test track for a severe lane-change manoeuvre – Part 1: Double lane-change*. 1999.

[74] Rikke Gade, Thomas Moeslund. “Thermal cameras and applications: a survey.” In: *Machine vision and applications* 25.1 (2014), pp. 245–262. DOI: 10.1007/s00138-013-0570-5.

[75] Sebastian Gnatzig. “Trajektorienbasierte teleoperation von straßenfahrzeugen auf basis eines shared-control-ansatzes.” PhD thesis. Technische Universität München, Germany, 2015.

[76] Christian Gold, Ilirjan Berisha, Klaus Bengler. “Utilization of drivetime-performing non-driving related tasks while driving highly automated.” In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 59. 2015, pp. 1666–1670. DOI: `10.1177/1541931215591360`.

[77] Christian Gold, Riender Happee, Klaus Bengler. “Modeling take-over performance in level 3 conditionally automated vehicles.” In: *Accident Analysis & Prevention* 116 (2018), pp. 3–13. DOI: `10.1016/j.aap.2017.11.009`.

[78] Christian Gold, Moritz Körber, David Lechner, Klaus Bengler. “Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density.” In: *Human factors* 58.4 (2016), pp. 642–652. DOI: `10.1177/0018720816634226`.

[79] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. “Generative adversarial networks.” In: *arXiv preprint arXiv:1406.2661* (2014). DOI: `10.1145/3422622`.

[80] Jamie Gorman, Nancy Cooke, Harry Pederson, Janie DeJoode. “Coordinated awareness of situation by teams (CAST): Measuring team situation awareness of a communication glitch.” In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 49. 3. 2005, pp. 274–277. DOI: `10.1177/154193120504900313`.

[81] Michael Goza, Robert Ambrose, Myron Diftler, Ivan Spain. “Telepresence Control of the NASA/DARPA Robonaut on a Mobility Platform.” In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2004, pp. 623–629. DOI: `10.1145/985692.985771`.

[82] Gaetano Graf. *Teleoperator-arbeitsplatz*. Deutsches Patent- und Markenamt: DE 10 2021 117 450 A1. 2023.

[83] Gaetano Graf. *Verfahren und Unterstützungseinrichtung zum Unterstützen eines Fernsteuerbetriebs eines Kraftfahrzeugs*. DE 10 2021 106 390 A1. 2022.

[84] Gaetano Graf, Yomna Abdelrahman, Hao Xu, Yasmeen Abdrabou, Dmitrij Schitz, Heinrich Hußmann, Florian Alt. “The Predictive Corridor: A Virtual Augmented Driving Assistance System for Teleoperated Autonomous Vehicles.” In: *ICAT-EGVE 2020 - International Conference on Artificial Reality and Telexistence and*

Eurographics Symposium on Virtual Environments. Ed. by Ferran Argelaguet, Ryan McMahan, Maki Sugimoto. The Eurographics Association, 2020. ISBN: 978-3-03868-111-3. DOI: `10.2312/egve.20201260`.

[85] Gaetano Graf, Heinrich Hussmann. "User Requirements for Remote Teleoperation-Based Interfaces." In: AutomotiveUI '20 (2020), pp. 85–88. DOI: `10.1145/3409251.3411730`.

[86] Gaetano Graf, Henri Palleis, Heinrich Hussmann. "A Design Space for Advanced Visual Interfaces for Teleoperated Autonomous Vehicles." In: *Proceedings of the International Conference on Advanced Visual Interfaces*. AVI '20. New York, NY, USA: Association for Computing Machinery, 2020. ISBN: 9781450375351. DOI: `10.1145/3399715.3399942`.

[87] Gaetano Graf, Hao Xu, Dmitrij Schitz, Xiao Xu. "Improving the Prediction Accuracy of Predictive Displays for Teleoperated Autonomous Vehicles." In: *6th International Conference on Control, Automation and Robotics*. IEEE, 2020, pp. 440–445. DOI: `10.1109/ICCAR49639.2020.9108011`.

[88] Gerhard Graf. *Teleoperator-arbeitsplatz*. Deutsches Patent- und Markenamt: DE 10 2021 123 234 A1. 2023.

[89] Gerhard Graf. *Verfahren und Vorrichtung zum Steuern eines Fahrzeugs*. Deutsches Patent- und Markenamt: DE 10 2021 126 423 A1. 2023.

[90] Gerhard Graf. *Vorrichtung und Verfahren zum manuellen Fernsteuern eines Fahrzeugs*. Deutsches Patent- und Markenamt: DE 10 2021 127 063 A1. 2023.

[91] James Hardy et al. "The radiation of heat from the human body: III. The human skin as a black-body radiator." In: *The Journal of clinical investigation* 13.4 (1934), pp. 615–620. DOI: `10.1172/JCI100609`.

[92] Sandra Hart, Lowell Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183. DOI: `10.1016/S0166-4115(08)62386-9`.

[93] Dwight Hennessy, Sara Howard, Erin Carr. "Driver space preference: differences across age, gender and traffic conditions." In: *Traffic Psychology: an International Perspective. Nova Science Publishers, New York* (2011), pp. 233–249.

[94] Jutta Hild, Edmund Klaus, Jan-Hendrik Hammer, Manuel Martin, Michael Voit, Elisabeth Peinsipp-Byma, Jürgen Beyerer. “A Pilot Study on Gaze-Based Control of a Virtual Camera Using 360°-Video Data.” In: *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer. 2018, pp. 419–428. DOI: 10.1007/978-3-319-91122-9_34.

[95] Orin Hoffman, Peter Keefe, Eric Smith, John Wang, Andrew Labrecque, Brett Ponsler, Susan Macchia, Brian Madge. *Remotely operating a mobile robot*. US Patent 9,283,674. 2016.

[96] Amin Hosseini. “Conception of Advanced Driver Assistance Systems for Precise and Safe Control of Teleoperated Road Vehicles in Urban Environments.” PhD thesis. Technische Universität München, Germany, 2018.

[97] Stephen Hughes, Michael Lewis. “Task-driven camera operations for robotic exploration.” In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 35.4 (2005), pp. 513–522. DOI: 10.1109/TSMCA.2005.850602.

[98] Edwin Hutchins. *Cognition in the wild*. eng. 8. pr. A Bradford book. Cambridge, Mass.: MIT Press, 2006. ISBN: 0262581469.

[99] Karl Iagnemma. *Supervisory control of vehicles*. US Patent 10,303,166. 2019.

[100] Robert Jacob. “Eye tracking in advanced interface design.” In: *Virtual Environments and Advanced Interface Design*. Oxford University Press, Inc., 1995. DOI: 10.1093/oso/9780195075557.003.0015.

[101] Myounghoon Jeon. “Emotions in driving.” In: *Emotions and Affect in Human Factors and Human-Computer Interaction*. Elsevier, 2017, pp. 437–474. DOI: 10.1016/B978-0-12-801851-4.00001-X.

[102] Nak-Tak Jeong, Keonhee Baek, Su-Bin Choi, Seonguk Choi, Ho-Yong Lee, Siwoo Kim, Myung-Won Suh. “A study on the HMI assessment of a joy stick driving system using driver workload measurements.” In: *Journal of Mechanical Science and Technology* 32.6 (2018), pp. 2781–2788. DOI: 10.1007/s12206-018-0535-8.

[103] Shinji Kajiwara. “Evaluation of driver’s mental workload by facial temperature and electrodermal activity under simulated driving conditions.” In: *International journal of Automotive technology* 15.1 (2014), pp. 65–70. DOI: 10.1007/s12239-014-0007-9.

[104] Arjun Kumar Kanduri, Geb Thomas, Nathalie Cabrol, Edmond Grin, Robert Anderson. “The (in) accuracy of novice rover operators’ perception of obstacle height from monoscopic images.” In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 35.4 (2005), pp. 505–512. DOI: `10.1109/TSMCA.2005.850601`.

[105] Michaela Kauer, Michael Schreiber, Ralph Bruder. “How to conduct a car? A design example for maneuver based driver-vehicle interaction.” In: *2010 IEEE Intelligent Vehicles Symposium*. IEEE. 2010, pp. 1214–1221. DOI: `10.1109/IVS.2010.5548099`.

[106] Jennifer Kay. “STRIPE: Remote driving using limited image data.” In: *Conference Companion on Human Factors in Computing Systems*. 1995, pp. 59–60. DOI: `10.1145/223355.223426`.

[107] Hande Kaymaz Keskinpala, Julie Adams. “Objective data analysis for a PDA-based human robotic interface.” In: *2004 IEEE International Conference on Systems, Man and Cybernetics*. Vol. 3. IEEE. 2004, pp. 2809–2814. DOI: `10.1109/ICSMC.2004.1400758`.

[108] Baeyoung Kim, Choong Min Jeong, Myungwon Suh. “Study of a simulator database and an experiment for the evaluation of highway driving safety.” In: *Journal of mechanical science and technology* 25.2 (2011), pp. 537–542. DOI: `10.1007/s12206-010-1228-0`.

[109] Jae-Seok Kim, Jee-Hwan Ryu. “Shared teleoperation of a vehicle with a virtual driving interface.” In: *2013 13th International Conference on Control, Automation and Systems*. ICCAS 2013. IEEE. 2013, pp. 851–857. DOI: `10.1109/ICCAS.2013.6704032`.

[110] Jens Kohlmorgen, Guido Dornhege, Mikio Braun, Benjamin Blankertz, Klaus-Robert Müller, Gabriel Curio, Konrad Hagemann, Andreas Bruns, Michael Schrauf, Wilhelm Kincses, et al. “Improving human performance in a real operating environment through real-time mental workload detection.” In: *Toward Brain-Computer Interfacing* (2007), pp. 409–422.

[111] Dimitrios Kontarinis, Robert Howe. “Tactile display of vibratory information in teleoperation and virtual environments.” In: *Presence: Teleoperators & Virtual Environments* 4.4 (1995), pp. 387–402. DOI: `10.1162/pres.1995.4.4.387`.

[112] Józef Kotus, Bartosz Kunka, Andrzej Czyzewski, Piotr Szczuko, Piotr Dalka, Rafa Rybacki. “Gaze-tracking and acoustic vector sensors technologies for PTZ camera steering and acoustic event detection.” In: *2010 Workshops on Database and Expert Systems Applications*. IEEE. 2010, pp. 276–280. DOI: `10.1109/DEXA.2010.62`.

[113] Christopher Kuhn, Gaetano Graf, Zhenxin Zhang. *Bildbasiertes Verfahren zum Vereinfachen einer fahrzeugexternen Kontrollübernahme über ein Kraftfahrzeug, Assistenzeinrichtung und Kraftfahrzeug*. Deutsches Patent- und Markenamt: DE 10 2021 101 363 A1. 2022.

[114] Christopher Kuhn, Markus Hofbauer, Sungkyu Lee, Goran Petrovic, Eckehard Steinbach. “Introspective Failure Prediction for Semantic Image Segmentation.” In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2020, pp. 1–6. DOI: `10.1109/ITSC45102.2020.9294308`.

[115] Christopher Kuhn, Markus Hofbauer, Goran Petrovic, Eckehard Steinbach. “Introspective black box failure prediction for autonomous driving.” In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2020, pp. 1907–1913. DOI: `10.1109/IV47402.2020.9304844`.

[116] Martin Kurze. “Rendering drawings for interactive haptic perception.” In: *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. 1997, pp. 423–430. DOI: `10.1145/258549.258826`.

[117] Joseph La Viola, Ernst Kruijff, Ryan McMahan, Doug Bowman, Ivan Poupyrev. *3D user interfaces: theory and practice*. Second edition. Addison-Wesley usability and HCI series. OCLC: ocn935986831. Boston: Addison-Wesley, 2017. ISBN: 978-0-13-403432-4.

[118] Donald Lampton, Daniel McDonald, Michael Singer, James Bliss. “Distance estimation in virtual environments.” In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 39. SAGE Publications Sage CA: Los Angeles, CA. 1995, pp. 1268–1272. DOI: `10.1177/154193129503902006`.

[119] Corde Lane, Craig Carignan, Brook Sullivan, David Akin, Teresa Hunt, Rob Cohen. “Effects of time delay on telerobotic control of neutral buoyancy vehicles.” In: *Proceedings 2002 IEEE International Conference on Robotics and Automation*. Vol. 3. IEEE. 2002, pp. 2874–2879. DOI: `10.1109/ROBOT.2002.1013668`.

[120] Hemin Omer Latif. “Mobile robot teleoperation through eye-gaze.” PhD thesis. Nottingham Trent University, England, 2010.

[121] Seungyon Lee, Shumin Zhai. "The Performance of Touch Screen Soft Buttons." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2009, pp. 309–318. DOI: `10.1145/1518701.1518750`.

[122] Mark Lipsey, David Wilson. *Practical meta-analysis.* Applied social research methods series. Thousand Oaks, Calif: Sage Publications, 2001. ISBN: 0761921672.

[123] Krzysztof Lis, Krishna Nakka, Pascal Fua, Mathieu Salzmann. "Detecting the unexpected via image resynthesis." In: *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 2019, pp. 2152–2161. DOI: `10.1109/ICCV.2019.00224`.

[124] Ruilin Liu, Daehan Kwak, Srinivas Devarakonda, Kostas Bekris, Liviu Iftode. "Investigating remote driving over the LTE network." In: *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 2017, pp. 264–269. DOI: `10.1145/3122986.3123008`.

[125] Salvatore Livatino. "VR enhanced teleoperation." In: *17th International Conference on Artificial Reality and Telexistence.* ICAT 2007. IEEE. 2007, pp. 290–291. DOI: `10.1109/ICAT.2007.51`.

[126] Amanda Lee Kelly Lockwood, Ravi Gogna, Gary Linscott, Timothy Caldwell, Marin Kobilarov, Paul Orecchio, Dan Xie, Ashutosh Gajanan Rege, Jesse Sol Levinson. *Interactions between vehicle and teleoperations system.* US Patent 10,386,836. 2019.

[127] Monika Lohani, Brennan R Payne, David Strayer. "A review of psychophysiological measures to assess cognitive states in real-world driving." In: *Frontiers in human neuroscience* 13 (2019), p. 57. DOI: `10.3389/fnhum.2019.00057`.

[128] Maura Lohrenz, Stephanie Edwards, Stephanie Myrick, Marlin Gendron, Michael Trenchard. "Demonstration of a moving-map system for improved precise lane navigation of amphibious vehicles and landing craft." In: *Oceans 2003. Celebrating the Past... Teaming Toward the Future.* Vol. 3. IEEE. 2003, pp. 1247–1254. DOI: `10.1109/OCEANS.2003.178032`.

[129] David Lopez, Jared Frank, Vikram Kapila. "Comparing interface elements on a tablet for intuitive teleoperation of a mobile manipulator." In: *ASME 2014 12th Biennial Conference on Engineering Systems Design and Analysis.* American Society of Mechanical Engineers Digital Collection. 2014. DOI: `10.1115/ESDA2014-20126`.

[130] Jason Luck, Patricia McDermott, Laurel Allender, Alia Fisher. "Advantages of co-location for effective human to human communication of information provided by an unmanned vehicle." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 50. 2006, pp. 550–554. DOI: `10.1177/154193120605000401`.

[131] Jason Luck, Patricia McDermott, Laurel Allender, Deborah Russell. "An investigation of real world control of robotic assets under communication latency." In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 2006, pp. 202–209. DOI: `10.1145/1121241.1121277`.

[132] Shay Magzimof, Paul Samuel Spencer. *Safety of autonomous vehicles using a virtual augmented support environment*. US Patent App. 16/191,868. 2019.

[133] Martin Mallwitz, Niels Will, Johannes Teiwes, Elsa Andrea Kirchner. "The capio active upper body exoskeleton and its application for teleoperation." In: *Proceedings of the 13th Symposium on Advanced Space Technologies in Robotics and Automation*. 2015.

[134] Julie Marble, David Bruemmer, Douglas Few. "Lessons learned from usability tests with a collaborative cognitive workspace for human-robot teams." In: *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*. Vol. 1. IEEE. 2003, pp. 448–453. DOI: `10.1109/ICSMC.2003.1243856`.

[135] Peter Marsh, Peter Collett. "The car as a weapon." In: *Et cetera* 44.2 (1987), p. 146.

[136] Andrés Martín-Barrio, Juan Jesús Roldán, Silvia Terrile, Jaime del Cerro, Antonio Barrientos. "Application of immersive technologies and natural language to hyper-redundant robot teleoperation." In: *Virtual Reality* (2019), pp. 1–15. DOI: `10.1007/s10055-019-00414-9`.

[137] Steven Martin, Nick Hillier. "Characterisation of the Novint Falcon haptic device for application as a robot manipulator." In: *Australasian Conference on Robotics and Automation (ACRA)*. Citeseer. 2009, pp. 291–292. DOI: `http://hdl.handle.net/102.100.100/110236?index=1`.

[138] Thomas Massie, Kenneth Salisbury. "The phantom haptic interface: A device for probing virtual objects." In: *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems*. Vol. 55. 1994, pp. 295–300.

[139] Michael Massimino, Thomas Sheridan. "Teleoperator performance with varying force and visual feedback." In: *Human factors* 36.1 (1994), pp. 145–157. DOI: `10.1177/001872089403600109`.

[140] Tobias Matts, Anton Sterner. "Vision-based driver assistance systems for teleoperation of onroad vehicles - compensating for impaired visual perception capabilities due to degraded video quality." MA thesis. Linköping University, Sweden, 2020.

[141] Douglas McGovern. *Experiences in teleoperation of land vehicles*. Tech. rep. National Aeronautics and Space Administration, 1989.

[142] Harsha Medicherla, Ali Sekmen. "Human-robot interaction via voice-controllable intelligent user interface." In: *Robotica* 25.5 (2007), p. 521. DOI: `10.1017/S0263574707003414`.

[143] Natasha Merat, Hamish Jamson, Frank Lai, Michael Daly, Oliver Carsten. "Transition to manual: Driver behaviour when resuming control from a highly automated vehicle." In: *Transportation research part F: traffic psychology and behaviour* 27 (2014), pp. 274–282. DOI: `10.1016/J.TRF.2014.09.005`.

[144] David Meyer, David Kieras. "A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena." In: *Psychological review* 104.4 (1997), p. 749. DOI: `10.1037/0033-295X.104.4.749`.

[145] David Meyer, David Kieras, Erick Lauber, Eric Schumacher, Jennifer Glass, Eileen Zurbriggen, Leon Gmeindl, Dana Apfelblat. "Adaptive executive control: Flexible multiple-task performance without pervasive immutable response-selection bottlenecks." In: *Acta Psychologica* 90.1-3 (1995), pp. 163–190. DOI: `10.1016/0001-6918(95)00026-Q`.

[146] Matjaž Mihelj, Janez Podobnik. *Haptics for virtual reality and teleoperation*. Intelligent systems, control and automation– Science and engineering. OCLC: ocn835989898. Dordrecht ; New York: Springer, 2012. ISBN: 978-94-007-5717-2.

[147] Andrzej Mitas, Artur Ryguła. "Biomedically aided car driver safety system." In: *Information Technologies in Biomedicine*. Springer, 2010, pp. 229–236. DOI: `10.1007/978-3-642-13105-9_24`.

[148] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas Altman, Prisma Group. "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." In: *PLoS med* 6.7 (2009), e1000097. DOI: `10.1136/bmj.b2535`.

[149] Ehsan Moradi-Pari, Amin Tahmasbi-Sarvestani, Yaser P Fallah. “A hybrid systems approach to modeling real-time situation-awareness component of networked crash avoidance systems.” In: *IEEE Systems Journal* 10.1 (2014), pp. 169–178. DOI: `10.1109/JSYST.2014.2312172`.

[150] Koji Murai, Yuji Hayashi, Tadatsugi Okazaki, Laurie Stone, Nobuo Mitomo. “Evaluation of ship navigator’s mental workload using nasal temperature and heart rate variability.” In: *2008 IEEE International Conference on Systems, Man and Cybernetics*. IEEE. 2008, pp. 1528–1533. DOI: `10.1109/ICSMC.2008.4811503`.

[151] Robin Murphy. “Human-robot interaction in rescue robotics.” In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34.2 (2004), pp. 138–153. DOI: `10.1109/TSMCC.2004.826267`.

[152] Robin Murphy, Jennifer Burke. “Up from the rubble: Lessons learned about HRI from search and rescue.” In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 49. SAGE Publications Sage CA: Los Angeles, CA. 2005, pp. 437–441. DOI: `10.1177/154193120504900347`.

[153] Robin Murphy, Karen Dreger, Sean Newsome, Jesse Rodocker, Eric Steimle, Tetsuya Kimura, Kenichi Makabe, Fumitoshi Matsuno, Satoshi Tadokoro, Kazuyuki Kon. “Use of remotely operated marine vehicles at Minamisanriku and Rikuzentakata Japan for disaster recovery.” In: *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*. IEEE. 2011, pp. 19–25. DOI: `10.1109/SSRR.2011.6106798`.

[154] Yasufumi Nagai, Shigeru Tsuchiya, Shinichi Kimura. “A study on audio feedback system for the operation of Engineering Test Satellite VII- Effects on telemetry monitor.” In: *International Symposium on Space Technology and Science, 22 nd, Morioka, Japan*. 2000, pp. 1226–1231.

[155] Wassim Najm, John Smith, Mikio Yanagisawa. *Pre-crash scenario typology for crash avoidance research*. Tech. rep. United States. National Highway Traffic Safety Administration, 2007.

[156] Frederik Naujoks, Christoph Mai, Alexandra Neukum. “The effect of urgency of take-over requests during highly automated driving under distraction conditions.” In: *Advances in human aspects of transportation* 7.Part I (2014), p. 431.

[157] Carl Nehme, Stacey Scott, Mary Cummings, Carina Yumi Furusho. “Generating requirements for futuristic hetrogenous unmanned systems.” In: vol. 50. 2006, pp. 235–239. DOI: `10.1177/154193120605000306`.

[158] Ulric Neisser. *Cognition and reality: principles and implications of cognitive psychology.* San Francisco: W. H. Freeman, 1976. ISBN: 978-0-7167-0478-2.

[159] Stefan Neumeier, Philipp Wintersberger, Anna-Katharina Frison, Armin Becher, Christian Facchi, Andreas Riener. "Teleoperation: The Holy Grail to Solve Problems of Automated Driving? Sure, but Latency Matters." In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 2019, pp. 186–197. DOI: `10.1145/3342197.3344534`.

[160] Thanh Nguyen, Chee Peng Lim, Ngoc Duy Nguyen, Lee Gordon-Brown, Saeid Nahavandi. "A Review of Situation Awareness Assessment Approaches in Aviation Environments." In: *IEEE Systems Journal* 13.3 (2019), pp. 3590–3603. DOI: `10.1109/JSYST.2019.2918283`.

[161] Michael Nielsen Curtis Wand Goodrich. "Comparing the usefulness of video and map information in navigation tasks." In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction.* 2006, pp. 95–101. DOI: `10.1145/1121241.1121259`.

[162] Günter Niemeyer, Carsten Preusche, Stefano Stramigioli, Dongjun Lee. "Telerobotics." In: *Springer handbook of robotics.* Springer, 2016, pp. 1085–1108. DOI: `10.1007/978-3-319-32552-1_43`.

[163] Graham Nixon. "Effective human-machine interfaces for aerial telemanipulation." MA thesis. Delft University of Technology, Netherlands, 2015.

[164] Klaus Oberauer, Reinhold Kliegl. "Simultaneous cognitive operations in working memory after dual-task practice." In: *Journal of Experimental Psychology: Human Perception and Performance* 30.4 (2004), p. 689. DOI: `10.1037/0096-1523.30.4.689`.

[165] Shunsuke Okishiba, Rui Fukui, Mitsuru Takagi, Hitoshi Azumi, Shin'ichi Warisawa, Ryoichi Togashi, Hiroyuki Kitaoka, Takeshi Ooi. "Tablet interface for direct vision teleoperation of an excavator for urban construction work." In: *Automation in Construction* 102 (2019), pp. 17–26. DOI: `10.1016/j.autcon.2019.02.003`.

[166] Oscar Olmos, Christopher Wickens, Andrew Chudy. "Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts and the application of cognitive engineering." In: *The International Journal of Aviation Psychology* 10.3 (2000), pp. 247–271. DOI: `10.1207/S15327108IJAP1003_03`.

[167] Daniel Olsson. “Techniques for Selecting Spatially Variable Video Encoder Quantization for Remote Operation.” MA thesis. Linköping University, Sweden, 2020.

[168] Calvin Or, Vincent Duffy. “Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement.” In: *Occupational Ergonomics* 7.2 (2007), pp. 83–94. DOI: `10.3233/OER-2007-7202`.

[169] Antti Oulasvirta, Kasper Hornbæk. “HCI research as problem-solving.” In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* 2016, pp. 4956–4967. DOI: `10.1145/2858036.2858283`.

[170] George Panagopoulos, Ioannis Pavlidis. “Forecasting Markers of Habitual Driving Behaviors Associated With Crash Risk.” In: *IEEE Transactions on Intelligent Transportation Systems* 21.2 (2019), pp. 841–851. DOI: `10.1109/TITS.2019.2910157`.

[171] Tracy Pastore. *Improved Operator Awareness of Teleoperated Land Vehicle Attitude.* Tech. rep. Naval Command Control, Ocean Surveillance Center RDT, and E Div San Diego CA, 1994.

[172] Ioannis Pavlidis, Ashik Khatri, Pradeep Buddharaju, Mike Manser, Robert Wunderlich, Ergun Akleman, Panagiotis Tsiamyrtzis. “Biofeedback arrests sympathetic and behavioral effects in distracted driving.” In: *IEEE Transactions on Affective Computing* (2018). DOI: `10.1109/TAFFC.2018.2883950`.

[173] William Payre, Julien Cestac, Patricia Delhomme. “Fully automated driving: Impact of trust and practice on manual control recovery.” In: *Human factors* 58.2 (2016), pp. 229–241. DOI: `10.1177/0018720815612319`.

[174] Luis Peñín, Rafael Aracil, Manuel Ferre, E Pinto, Miguel Hernando, Antonio Barrientos. “Telerobotic system for live power lines maintenance: ROBTET.” In: *Proceedings. 1998 IEEE International Conference on Robotics and Automation.* Vol. 3. IEEE. 1998, pp. 2110–2115. DOI: `10.1109/ROBOT.1998.680632`.

[175] Peter Perla, Michael Markowitz, Albert Nofi, Christopher Weuve, Julia Loughran. *Gaming and shared situation awareness.* Tech. rep. Center For Naval Analyses Alexandria VA, 2000.

[176] Anthony Perry, Dawn Baldwin. “Further evidence of associations of type A personality scores and driving-related attitudes and behaviors.” In: *Perceptual and motor skills* 91.1 (2000), pp. 147–154. DOI: `10.2466/pms.2000.91.1.147`.

[177] Helena Pongrac. "Gestaltung und Evaluation von virtuellen und Telepräsenzsystemen an Hand von Aufgabenleistung und Präsenzempfinden." PhD thesis. Universität der Bundeswehr München, Germany, 2011.

[178] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Ng. "ROS: an open-source Robot Operating System." In: *ICRA workshop on open source software*. Vol. 3. Kobe, Japan. 2009, p. 5.

[179] Gary Reid, Thomas Nygren. "The subjective workload assessment technique: A scaling procedure for measuring mental workload." In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 185–218. DOI: `10.1016/S0166-4115(08)62387-0`.

[180] Paul Richard, Georges Birebent, Philippe Coiffet, Grigore Burdea, Daniel Gomez, Noshir Langrana. "Effect of frame rate and force feedback on virtual object manipulation." In: *Presence: Teleoperators & Virtual Environments* 5.1 (1996), pp. 95–108. DOI: `10.1162/pres.1996.5.1.95`.

[181] Bob Ricks, Curtis Nielsen, Michael Goodrich. "Ecological displays for robot interaction: A new perspective." In: *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*. Vol. 3. IEEE. 2004, pp. 2855–2860. DOI: `10.1109/IROS.2004.1389842`.

[182] Jennifer Riley, Mica Endsley. "The hunt for situation awareness: Human-robot interaction in search and rescue." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 48. 2004, pp. 693–697. DOI: `10.1177/154193120404800389`.

[183] John Riva, Keshena Malik, Stephen Burnie, Andrea Endicott, Jason Busse. "What is your research question? An introduction to the PICOT format for clinicians." In: *The Journal of the Canadian Chiropractic Association* 56.3 (2012), p. 167.

[184] Jennifer L Rochlis, Dava J Newman. "A tactile display for international space station (ISS) extravehicular activity (EVA)." In: *Aviation, space, and environmental medicine* 71.6 (2000), pp. 571–578.

[185] Emilie M Roth, Mark L Hanson, Chris Hopkins, Vince Mancuso, Greg L Zacharias. "Human in the loop evaluation of a mixed-initiative system for planning and control of multiple UAV teams." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 48. SAGE Publications Sage CA: Los Angeles, CA. 2004, pp. 280–284. DOI: `10.1177/154193120404800301`.

[186] Gabriella Rydenfors. “Telepresence and remote communication through virtual reality.” MA thesis. Linköping University, Sweden, 2017.

[187] Eduardo Salas, Carolyn Prince, David Baker, Lisa Shrestha. “Situation awareness in team performance: Implications for measurement and training.” In: *Human factors* 37.1 (1995), pp. 123–136. DOI: `10.1518/001872095779049525`.

[188] Paul Salmon, Guy Walker, Darshna Ladva, Neville Stanton, Daniel Jenkins, Laura Rafferty. “Measuring situation awareness in command and control: comparison of methods study.” In: *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore!* 2007, pp. 27–34. DOI: `10.1145/1362550.1362559`.

[189] Paul Salmon, Guy Walker, Neville Stanton. “Pilot error versus sociotechnical systems failure: a distributed situation awareness analysis of Air France 447.” In: *Theoretical Issues in Ergonomics Science* 17.1 (2016), pp. 64–79. DOI: `10.1080/1463922X.2015.1106618`.

[190] Enrique Sanchis-Sánchez, Carlos Vergara-Hernández, Rosa Cibrián, Rosario Salvador, Enrique Sanchis, Pilar Codoñer-Franch. “Infrared thermal imaging in the diagnosis of musculoskeletal injuries: a systematic review and meta-analysis.” In: *American Journal of Roentgenology* 203.4 (2014), pp. 875–882. DOI: `10.2214/AJR.13.11716`.

[191] Nadine Sarter, David Woods. “Situation awareness: A critical but ill-defined phenomenon.” In: *The International Journal of Aviation Psychology* 1.1 (1991), pp. 45–57. DOI: `10.1207/S15327108IJAP0101_4`.

[192] Shlomo Sawilowsky. “New effect size rules of thumb.” In: *Journal of Modern Applied Statistical Methods* 8.2 (2009), p. 26. DOI: `10.22237/JMASM/1257035100`.

[193] Felix Schill, Robert Mahony, Peter Corke, Luke Cole. “Virtual force feedback teleoperation of the insectbot using optic flow.” In: *Proceedings of the 2008 Australasian Conference on Robotics and Automation.* The Australian Robotics and Automation Association Inc. 2008, pp. 1–9.

[194] Dmitrij Schitz, Gaetano Graf, Dominik Rieth, Harald Aschemann. “Corridor-Based Shared Autonomy for Teleoperated Driving.” In: *IFAC-PapersOnLine* 53.2 (2020). 21st IFAC World Congress, pp. 15368–15373. ISSN: 2405-8963. DOI: `10.1016/j.ifacol.2020.12.2351`.

[195] Dmitrij Schitz, Gaetano Graf, Dominik Rieth, Harald Aschemann. “Interactive Corridor-Based Path Planning for Teleoperated Driving.” In: *2021 7th International Conference on Mechatronics and Robotics Engineering (ICMRE)* (2021), pp. 174–179. DOI: `10.1109/ICMRE51691.2021.9384848`.

[196] Dmitrij Schitz, Gaetano Graf, Dominik Rieth, Harald Aschemann. “Model-Predictive Cruise Control for Direct Teleoperated Driving Tasks.” In: *2021 European Control Conference (ECC)* (2021), pp. 1808–1813. DOI: `10.23919/ecc54610.2021.9655187`.

[197] MacKenzie Scott, Ware Colin. “Lag as a determinant of human performance in interactive systems.” In: *Proceedings of the INTERACT’93 and CHI’93 conference on Human factors in computing systems.* 1993, pp. 488–493. DOI: `10.1145/169059.169431`.

[198] Ronald Scott, Emilie Roth, Stephen Deutsch, Erika Malchiodi, Thomas Kazmierczak, Robert Eggleston, Samuel Kuper, Randall Whitaker. “Work-centered support systems: A human-centered approach to intelligent system design.” In: *IEEE Intelligent Systems* 20.2 (2005), pp. 73–81. DOI: `10.1109/MIS.2005.37`.

[199] David Scribner, James Gombash. *The effect of stereoscopic and wide field of view conditions on teleoperator performance.* Tech. rep. Army Research Lab Aberdeen Proving Ground MD Human Research and Engineering Directorate, 1998.

[200] Dvijesh Shastri, Manos Papadakis, Panagiotis Tsiamyrtzis, Barbara Bass, Ioannis Pavlidis. “Perinasal imaging of physiological stress and its affective potential.” In: *IEEE Transactions on Affective Computing* 3.3 (2012). Conference Name: IEEE Transactions on Affective Computing, pp. 366–378. ISSN: 1949-3045. DOI: `10.1109/T-AFFC.2012.13`.

[201] Xiaotong Shen, Zhuang Jie Chong, Scott Pendleton, Guo Ming James Fu, Baoxing Qin, Emilio Frazzoli, Marcelo Ang. “Teleoperation of on-road vehicles via immersive telepresence using off-the-shelf components.” In: *Intelligent Autonomous Systems 13.* Springer, 2016, pp. 1419–1433. DOI: `10.1007/978-3-319-08338-4_102`.

[202] Thomas Sheridan. “Space teleoperation through time delay: Review and prognosis.” In: *IEEE Transactions on robotics and Automation* 9.5 (1993), pp. 592–606. DOI: `10.1109/70.258052`.

[203] Thomas Sheridan. *Supervisory control of remote manipulators, vehicles and dynamic processes: Experiments in command and display aiding*. Tech. rep. Massachusetts Inst Of Tech Cambridge Man-Machine Systems Lab, 1983.

[204] Thomas Sheridan, William Ferrell. “Remote Manipulative Control with Transmission Delay.” In: *IEEE Transactions on Human Factors in Electronics* HFE-4.1 (1963), pp. 25–29. DOI: `10.1109/THFE.1963.231283`.

[205] Russell Shilling, Barbara Shinn-Cunningham. “Virtual auditory displays.” In: *Handbook of virtual environment technology* (2002), pp. 65–92.

[206] Leonid Slutski. “Control Methods in the Systems "Operator-Remote Manipulator".” In: *Remote Manipulation Systems*. Springer, 1998, pp. 59–120. DOI: `10.1007/978-94-011-5153-5_4`.

[207] Kip Smith, Peter Hancock. “Situation awareness is adaptive, externally directed consciousness.” In: *Human factors* 37.1 (1995), pp. 137–148. DOI: `10.1518/001872095779049444`.

[208] Christopher Smyth. “Indirect vision driving with fixed flat panel displays for near unity, wide, and extended fields of camera view.” In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 44. SAGE Publications Sage CA. Los Angeles, CA, 2000, pp. 541–544. DOI: `10.1177/154193120004403624`.

[209] Donna Spencer, Jesse James Garrett. *Card sorting: designing usable categories*. OCLC: ocn325125041. Brooklyn, N.Y: Rosenfeld Media, 2009. ISBN: 978-1-933820-02-6.

[210] Charles D. Spielberger. “State-Trait Anxiety Inventory.” In: *The Corsini Encyclopedia of Psychology*. John Wiley and Sons, Ltd, 2010. ISBN: 9780470479216. DOI: `10.1002/9780470479216.CORPSY0943`.

[211] Daniele Sportillo, Alexis Paljic, Mehdi Boukhris, Philippe Fuchs, Luciano Ojeda, Vincent Roussarie. “An immersive Virtual Reality system for semi-autonomous driving simulation: a comparison between realistic and 6-DoF controller-based interaction.” In: *Proceedings of the 9th International Conference on Computer and Automation Engineering*. 2017, pp. 6–10. DOI: `10.1145/3057039.3057079`.

[212] Neville Stanton, Peter Chambers, John Piggott. “Situational awareness and safety.” In: *Safety science* 39.3 (2001), pp. 189–204. DOI: `10.1016/S0925-7535(01)00010-8`.

[213] Neville Stanton, Paul Salmon, Guy Walker, Eduardo Salas, Peter Hancock. "State-of-science: situation awareness in individuals, teams and systems." In: *Ergonomics* 60.4 (2017), pp. 449–466. DOI: `10.1080/00140139.2017.1278796`.

[214] Neville Stanton, Rebecca Stewart, Don Harris, Robert Houghton, Chris Baber, Richard McMaster, Paul Salmon, Geoff Hoyle, Guy Walker, Mark Young. "Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology." In: *Ergonomics* 49.12-13 (2006), pp. 1288–1311. DOI: `10.1080/00140130600612762`.

[215] Jork Stapel, Freddy Antony Mullakkal-Babu, Riender Happee. "Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving." In: *Transportation research part F: traffic psychology and behaviour* 60 (2019), pp. 590–605. DOI: `10.1016/j.trf.2018.11.006`.

[216] Aaron Steinfeld. "Interface lessons for fully and semi-autonomous mobile robots." In: *IEEE International Conference on Robotics and Automation.* Vol. 3. ICRA'04. IEEE. 2004, pp. 2752–2757. DOI: `10.1109/ROBOT.2004.1307477`.

[217] Christoph Stiller. "Fahrerassistenzsysteme-Von realisierten Funktionen zum vernetzt wahrnehmenden, selbstorganisierenden Verkehr." In: *Fahrerassistenzsysteme mit maschineller Wahrnehmung.* Springer, 2005, pp. 1–20. DOI: `10.1007/3-540-27137-6_1`.

[218] Sonja Stockert, Natalie Tara Richardson, Markus Lienkamp. "Driving in an increasingly automated world–approaches to improve the driver-automation interaction." In: *Procedia Manufacturing* 3 (2015), pp. 2889–2896. DOI: `10.1016/J.PROMFG.2015.07.797`.

[219] Martin Strobl. "SPIDER Das innovative Software Framework der BMW Fahrsimulation. SPIDER The innovative software framework of the BMW driving simulation." In: *VDI-Berichte* 1745 (2003).

[220] Taylor. "Situational Awareness." In: ed. by Eduardo Salas. Fisrt edition. Routledge, 1990. Chap. Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design. ISBN: 978-1-315-08792-4. DOI: `10.4324/9781315087924`.

[221] Lisa Thomas, Christopher Wickens. *Effects of display frames of reference on spatial judgments and change detection.* Tech. rep. Illinois Univ at Urbana-Champaign Savoyaviation Research Lab, 2000.

[222] Lisa Thomas, Christopher Wickens. “Visual displays and cognitive tunneling: Frames of reference effects on spatial judgments and change detection.” In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 45. Los Angeles, CA, 2001, pp. 336–340. DOI: `10.1177/154193120104500415`.

[223] Jennifer E Thropp, Jessie Y Chen. *The effects of slow frame rates on human performance*. Tech. rep. Army Research Lab Aberdeen Proving Ground MD Human Research and Engineering Directorate, 2006. DOI: `10.21236/ada449963`.

[224] Alvin Tilley. *The Measure of man and woman: human factors in design*. Ed. by Roberto de Alba. New York: Whitney Library of Design, 1993. ISBN: 978-0-8230-3031-6.

[225] Alvaro Uribe, Silas Alves, João Rosário, Humberto Ferasoli Filho, Byron Pérez-Gutiérrez. “Mobile robotic teleoperation using gesture-based human interfaces.” In: *IX Latin American Robotics Symposium and IEEE Colombian Conference on Automatic Control, 2011 IEEE*. IEEE. 2011, pp. 1–6. DOI: `10.1109/LARC.2011.6086812`.

[226] Alessandro Valli. “The design of natural interaction.” In: *Multimedia Tools and Applications* 38.3 (2008), pp. 295–305. DOI: `10.1007/s11042-007-0190-z`.

[227] Jan Van Erp, Pieter Padmos. “Image parameters for driving with indirect viewing systems.” In: *Ergonomics* 46.15 (2003), pp. 1471–1499. DOI: `10.1080/0014013032000121624`.

[228] Boris Velichkovsky, Mikhail Rumyantsev, Mikhail Morozov. “New solution to the midas touch problem: Identification of visual commands via extraction of focal fixations.” In: *procedia computer science* 39 (2014), pp. 75–82.

[229] Michael Vidulich, Christopher Wickens. “Causes of dissociation between subjective workload measures and performance: Caveats for the use of subjective assessments.” en. In: *Applied Ergonomics* 17.4 (1986), pp. 291–296. ISSN: 0003-6870. DOI: `10.1016/0003-6870(86)90132-8`.

[230] Wolfgang Viechtbauer. *Metafor: meta-analysis package for R*. 2010.

[231] Tobias Vogelpohl, Mark Vollrath, Matthias Kühn, Thomas Hummel, Tina Gehlert. *Übergabe von hochautomatisiertem Fahren zu manueller Steuerung*. Gesamtverband der Deutschen Versicherungswirtschaft, 2016.

[232] Athanasios Vourvopoulos, Fotis Liarokapis. “Brain-controlled NXT Robot: Teleoperating a robot through brain electrical activity.” In: *2011 Third International Conference on Games and Virtual Worlds for Serious Applications*. IEEE. 2011, pp. 140–143. DOI: `10.1109/VS-GAMES.2011.27`.

[233] Steve Eric Vozar, Dawn Tilbury. “Augmented reality user interface for mobile robots with manipulator arms: Development, testing, and qualitative analysis.” In: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Vol. 45011. American Society of Mechanical Engineers. 2012, pp. 1341–1350. DOI: `10.1115/DETC2012-70347`.

[234] Steven Eric Vozar. “A Framework for Improving the Speed and Performance of Teleoperated Mobile Manipulators.” PhD thesis. The University of Michigan, Ann Arbor, Michigan, 2013.

[235] Guy Walker, Neville Stanton, Paul Salmon, Daniel Jenkins, Laura Rafferty. “Translating concepts of complexity to the field of ergonomics.” In: *Ergonomics* 53.10 (2010), pp. 1175–1186. DOI: `10.1080/00140139.2010.513453`.

[236] Lee-ming Wang, Vincent. Duffy, Yingzi Du. “A Composite Measure for the Evaluation of Mental Workload.” en. In: *Digital Human Modeling*. Vol. 4561. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 460–466. ISBN: 978-3-540-73318-8 978-3-540-73321-8. DOI: `10.1007/978-3-540-73321-8_53`.

[237] Wenbi Wang. “Human navigation performance using 6 DOF dynamic viewpoint tethering in virtual environments.” PhD thesis. University of Toronto, Canada, 2004.

[238] Rodney Wellens. “Group situation awareness and distributed decision making: From military to civilian applications.” In: *Individual and group decision making: Current issues* (1993), pp. 267–291.

[239] Steffen Werner. “Cognitive reference systems and their role in designing spatial information displays.” In: *Künstliche Intelligenz* 16 (2002), pp. 10–13.

[240] Avinash Wesley, Ioannis Pavlidis, Shastri Dvijesh. “A novel method to monitor driver’s distractions.” en. In: *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA ’10*. 2010, p. 4273. DOI: `10.1145/1753846.1754138`.

[241] John Wilson. “Fundamentals of systems ergonomics.” In: *Work* 41.Supplement 1 (2012), pp. 3861–3868. DOI: `10.1016/j.apergo.2013.03.021`.

[242] Alan Winfield. "Future directions in tele-operated robotics." In: *Telerobotic applications. London: Professional Engineering Pub* (2000), pp. 147–163.

[243] Hermann Winner, Stephan Hakuli, Gabriele Wolf. *Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort: mit 550 Abbildungen und 45 Tabellen.* Springer-Verlag, 2009. DOI: `10.1007/978-3-658-05734-3`.

[244] Bob Witmer, Paul Kline. "Judging perceived and traversed distance in virtual environments." In: *Presence* 7.2 (1998), pp. 144–167. DOI: `10.1162/105474698565640`.

[245] David Woods. "Visual momentum: a concept to improve the cognitive coupling of person and computer." In: *International Journal of Man-Machine Studies* 21.3 (1984), pp. 229–244. DOI: `10.1016/S0020-7373(84)80043-7`.

[246] David Woods, James Tittle, Magnus Feil, Axel Roesler. "Envisioning human-robot coordination in future operations." In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34.2 (2004), pp. 210–218. DOI: `10.1109/TSMCC.2004.826272`.

[247] Yijing Wu, Qijie Zhao, Junye Jiang. "Driver's Head Behavior Detection Using Binocular Infrared Camera." In: *2018 IEEE International Conference on Information and Automation (ICIA)*. IEEE. 2018, pp. 565–569. DOI: `10.1109/ICInfA.2018.8812434`.

[248] Felix Wulf, Maria Rimini-Döring, Marc Arnon, Frank Gauterin. "Recommendations supporting situation awareness in partially automated driver assistance systems." In: *IEEE Transactions on Intelligent Transportation Systems* 16.4 (2014), pp. 2290–2296. DOI: `10.1109/TITS.2014.2376572`.

[249] Yingda Xia, Yi Zhang, Fengze Liu, Wei Shen, Alan L Yuille. "Synthesize then compare: Detecting failures and anomalies for semantic segmentation." In: *European Conference on Computer Vision*. Springer. 2020, pp. 145–161. DOI: `10.1007/978-3-030-58452-8_9`.

[250] Xuehan Xiong, Fernando De la Torre. "Supervised Descent Method and Its Applications to Face Alignment." en. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2013, pp. 532–539. ISBN: 978-0-7695-4989-7. DOI: `10.1109/CVPR.2013.75`.

[251] Holly Yanco, Jill Drury. “"Where am I?" Acquiring situation awareness using a remote robot platform.” In: *2004 IEEE International Conference on Systems, Man and Cybernetics.* Vol. 3. IEEE. 2004, pp. 2835–2840. DOI: `10.1109/ICSMC.2004.1400762`.

[252] Tsun-Yi Yang, Yi-Ting Chen, Yen-Yu Lin, Yung-Yu Chuang. “Fsa-net: Learning fine-grained structure aggregation for head pose estimation from a single image.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2019, pp. 1087–1096. DOI: `10.1109/CVPR.2019.00118`.

[253] Mark Young, Karel Brookhuis, Christopher Wickens, Peter Hancock. “State of science: mental workload in ergonomics.” In: *Ergonomics* 58.1 (2015), pp. 1–17. ISSN: 0014-0139. DOI: `10.1080/00140139.2014.956151`.

[254] Mingxin Yu, Xiangzhou Wang, Yingzi Lin, Xiaoping Bai. “Gaze tracking system for teleoperation.” In: *The 26th Chinese Control and Decision Conference.* CCDC’14. IEEE. 2014, pp. 4617–4622. DOI: `10.1109/CCDC.2014.6852997`.

[255] Meng Zhang, Klas Ihme, Uwe Drewitz. “Discriminating drivers’ emotions through the dimension of power: Evidence from facial infrared thermography and peripheral physiological measurements.” In: *Transportation Research Part F: Traffic Psychology and Behaviour* 63 (2019), pp. 135–143. DOI: `10.1016/j.trf.2019.04.003`.

[256] Qiushi Zhou, Joshua Newn, Namrata Srivastava, Tilman Dingler, Jorge Gonçalves, Eduardo Velloso. “Cognitive Aid: Task Assistance Based On Mental Workload Estimation.” In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). DOI: `10.1145/3290607.3313010`.

[257] Dingyun Zhu, Tom Gedeon, Ken Taylor. “Keyboard before head tracking depresses user success in remote camera control.” In: *IFIP Conference on Human-Computer Interaction.* Springer. 2009, pp. 319–331. DOI: `10.1007/978-3-642-03658-3_37`.

[258] Dingyun Zhu, Tom Gedeon, Ken Taylor. “Natural interaction enhanced remote camera control for teleoperation.” In: *CHI’10 Extended Abstracts on Human Factors in Computing Systems.* 2010, pp. 3229–3234. DOI: `10.1145/1753846.1753963`.

[259] Ferdinand Rudolf Hendrikus Zijlstra. “Efficiency in work behaviour: A design approach for modern tools.” PhD thesis. Delft University of Technology, Netherlands, 1993.

All URLs cited were proved in March 2024.

Glossary

ANOVA The Analysis of Variance (ANOVA) is a statistical technique used to test whether there is a significant difference between the means of two or more groups.

API An Application Programming Interface (API) is a set of protocols, routines, and tools for building software and applications. It specifies how software components should interact with each other.

ATAS The Advanced Teleoperator Assistance System (ATAS) is a system that allows a human operator to control a robotic device or system remotely.

AV An Autonomous Vehicle (AV) is a vehicle capable of sensing its environment and navigating without human input.

BCI A Brain-Computer Interface (BCI) is a system that allows a person to communicate with a computer or other device using brain activity.

BMW Bayerische Motoren Werke (BMW) is a German multinational corporation that produces luxury vehicles and motorcycles.

CARLA The Car Learning to Act (CARLA) is an open-source platform for developing autonomous driving technology. It provides tools for simulating and testing autonomous vehicles in a realistic environment.

CI Confidence Intervals (CIs) are a statistical measure of the uncertainty surrounding a sample estimate of a population parameter.

DARPA The Defense Advanced Research Projects Agency (DARPA) is an agency of the United States Department of Defense responsible for developing new technologies for use in military applications.

EBS The Emergency Braking Stop (EBS) is a feature of some vehicles that automatically applies the brakes in an emergency to prevent a collision.

ECG An Electrocardiogram (ECG) is a test that measures the heart's electrical activity. It is often used to diagnose heart problems.

EEG An Electroencephalography (EEG) is a test that measures the brain's electrical activity. It is often used to diagnose brain disorders.

EMG An Electromyogram (EMG) is a test that measures the electrical activity of muscles. It is often used to diagnose muscle disorders.

EOG An Electrooculogram (EOG) is a test that measures the electrical activity of the eye muscles. It is often used to diagnose eye disorders.

FC The Free Corridor (FC) is a path free of obstacles and can be safely traversed by a vehicle.

FoV The Field of View (FoV) is the portion of the environment that is visible to an observer at a given time, typically defined by the lens of a camera or imaging system.

FR The Frame Rate (FR) define the number of images or frames displayed per second in a video or animation, typically measured in frames per second (fps).

FSA-Net The Fine-Grained Structure Aggregation Network (FSA-Net) is a machine-learning model that classifies images based on their fine-grained structure.

GDTA The Goal-Directed Task Analysis (GDTA) is a method of analyzing tasks to understand how they are performed and identify potential improvements.

GSR The Galvanic Skin Response (GSR) is a measure of the skin's electrical conductivity, which can detect changes in arousal or stress levels.

GUI A Graphical User Interface (GUI) is a user interface that allows users to interact with a computer or device through graphical icons and visual indicators rather than text-based commands.

HCI Human-Computer Interaction (HCI) studies how people interact with computers and other technology.

HMD A Head Mounted Display (HMD) is a device that displays images or video in front of a person's eyes, typically using a headset or glasses.

HMI A Human-Machine Interface (HMI) is a system that allows humans to interact with machines, typically through input and output devices such as keyboards, touchscreens, and displays.

HOG The Histograms of Oriented Gradients (HOD) is a feature descriptor used in computer vision algorithms for object detection and classification. It is based on the distribution of gradient orientations in an image.

LC The Lane-Change (LC) is a driving maneuver in which a vehicle changes lanes on a roadway.

LiDAR A Light Detecting and Ranging (LiDAR) is a remote sensing technique that uses lasers to measure the distance to an object or surface. It is often used for mapping and surveying applications.

NASA The National Aeronautics and Space Administration (NASA) is a U.S. government agency responsible for the nation's civilian space program and for conducting scientific research in space.

NASA-TLX The NASA Task Load Index (NASA-TLX) is a subjective workload assessment tool used to measure the mental demand, physical demand, temporal demand, performance, effort, and frustration of a task.

NETD The Noise Equivalent Temperature Difference (NETD) is a measure of the sensitivity of a thermal imaging camera, defined as the most negligible temperature difference that the camera can detect at a given signal-to-noise ratio.

ORAD The On-Road Automated Driving (ORAD) committee describes the use of automated driving technologies on public roads.

PBA The Predictive Brake Assistance (PBA) is a vehicle safety feature that uses sensors and algorithms to anticipate and prevent collisions by automatically applying the brakes when necessary.

PC The Predictive Corridor (PC) is a feature in autonomous vehicles that uses sensors and algorithms to anticipate and plan for future vehicle positions on the road.

PD The Predictive Display (PD) is a feature that uses sensors and algorithms to predict and display future road events, such as the actions of other vehicles or pedestrians.

PICOTS The Population, Intervention, Comparison group, Outcomes, Time frame, Setting (PICOT) is a framework used in research studies to describe the critical components of the study design. It stands for Population (the group of people being studied), Intervention (the treatment or intervention being tested), Comparison group (the group being compared to the intervention group), Outcomes (the measures being used to evaluate the effectiveness of the intervention), Time frame (the duration of the study), and Setting (the location where the study is being conducted).

PRISMA The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is a set of guidelines for reporting the results of systematic reviews and meta-analyses of healthcare interventions.

RoI A Region of Interest (RoI) is a specific area in an image or video frame that is of particular interest and should be analyzed in more detail.

ROS The Robot Operating System (ROS) is an open-source software framework that provides tools and libraries for building and operating robots.

RSME The Root Square Mean Error (RSME) is a measure of the accuracy of a model or prediction, calculated as the square root of the mean squared error between the predicted and actual values.

SA The Situation Awareness (SA) is the ability to understand and interpret the current situation to anticipate and respond to potential future events.

SAE The Society of Automotive Engineers (SAE) is a professional organization for engineers and technical experts in the automotive industry.

SAGAT The Situation Awareness Global Assessment Technique (SAGAT) is a tool for measuring situation awareness in complex systems, such as aviation or automotive.

SART The Situation Awareness Rating Technique (SART) is a method for evaluating an individual's situational awareness or ability to perceive and comprehend their environment.

SMD The Standardized Mean-Difference (SMD) is a statistical measure used to compare the means of two groups or samples.

SME A Subject Matter Expert (SME) is an individual with specialized knowledge or expertise in a particular field or topic.

SPIDER The Software Programming Interface for Distributed Real-time Driving Simulation (SPIDER) is a tool for creating and running realistic simulations of driving scenarios.

ToR A Takeover Request (ToR) is a request made by a driver or operator to transfer control of a vehicle or system to another person or entity.

ToT The Takeover Time (ToT) is the time a driver or operator takes to regain control of a vehicle or system after a takeover request.

UAV An Unmanned Aerial Vehicle (UAV) is an aircraft operated remotely or autonomously without a human pilot on board.

UGV An Unmanned Ground Vehicle (UGV) is a ground-based vehicle operated remotely or autonomously without a human operator on board.

UUV An Unmanned Underwater Vehicle (UUV) is an underwater vehicle operated remotely or autonomously without a human operator on board.

List of Figures

1.1 This illustration shows the environmental sensor of a On-Road Automated Driving (ORAD) Level 3 vehicle. Moreover, we see a possible teleoperation scenario, i.e., an unexpected non-moving obstacle lying on the vehicle path. In such a scenario, according to traffic regulations, the Autonomous Vehicle (AV) is not allowed to drive over the solid marking lane. However, the AV may wirelessly send a Takeover Request to the teleoperator to solve the issue. 15

1.2 The diagram visually shows an overview of the content interplay in my research. 23

2.1 The data transmission scheme for teleoperated AVs. The AV transfers radio-based environmental data to the teleoperator, who can remotely assess and manage the vehicle. Operators interpret the data and transfer operational and strategic commands to the vehicle. 28

2.2 An adaptation of the diagram of automation level by Sheridan [203]. Many different authors have adopted Sheridan's [203] model, e.g., [32, 68, 75], though with some adaptation, i.e., instead of having five layers of automation, there were only three, and in all, computer aiding mediation is present. It can be observed that the more abstract the task is, the higher degree of automation is required, i.e., onboard vehicle intelligence. 30
2.3 In this image, a *Phantom Auto* remote operator controls an autonomous mobile delivery robot via a dual joystick controller that asked for support. An image from the 2020 *Phantom Auto* Press Kit. 44
2.4 The illustration shows a teleoperator workstation envisioned by Lockwood et al. [126]. In particular, the teleoperator can assess and conduct the remote vehicle using touch-sensitive input devices. 45
2.5 This array of images shows the gesture-based driving mode proposed by Fong et al. [69]. To start, the operator raises his or her left hand to activate the robot and uses the right hand to specify the direction, e.g., to the right (a), to the left (b), or stop (c). 46
2.6 In this image, the equipment in Amai et al. [6]. Clockwise from left, the MindMouse headband and its hardware attached to a portable computer, the vehicle robot, radio data hardware, lastly, a battery belt. 49
2.7 In this image, the professional drifter Vaughn Gittin Jr. drives an AV remotely using a steering wheel and pedals as input devices and an Head Mounted Display as output devices. An image from the 2019 *Samsung* Press Kit. 54

3.1 The Situation Awareness (SA) model in dynamic decision-making processes proposed by Endsley [63]. Here the SA, formed by three layers, is mediated by systemic and operator factors. . . . 65

3.2 The Situation Awareness (SA) model by Smith and Hancock [207] is based upon Niesser's [158] perceptual cycle model. The model describes the cycle of perception and action, i.e., the human interaction with the world and thus the formation of SA. 66
3.3 The model that explains team Situation Awareness (SA). We see team members must possess the SA related to their roles and goals while also holding the SA related to other team members. 68

4.1 A summary illustration of the literature we collected following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [148]. Of 5204 unique records, 54 were retrieved for screening, 13 studies were analyzed in-depth, and four coded for the meta-analysis using the Population, Intervention, Comparison group, Outcomes, Time frame, Setting (PICOTS) framework [183]. Unfortunately, eight studies lacked the necessary statistics needed for the meta-analysis. 80
4.2 An overview of the Regions of Interest (RoI) we identified during the literature analysis. From left to right, (a) supraorbital, (b) forehead, (c) nose-tip, and (d) perinasal regions. 85

4.3 The Forest Plot shows the Standardized Mean-Difference (SMD) and the Confidence Intervals (CI) of the studies taken into consideration. The last line shows the size, with its CI, of the Random Effects model. The Forest Plot also includes yellow "diamonds" for each study. These are the predicted effect sizes computed from the meta-regression model. The column *Setting* indicates where the study was conducted (in-situ or ex-situ), the *RoIs* where the thermophysiological data was extracted (forehead, nose tip, and perinasal). In column *Conditions*, C.D. stands for Cognitive Distraction, U.E. for Unexpected Event, and S.D. for Sensorimotor Distraction. To understand the output analysis, the SMD value is interpreted similar to the effect size (*d*) defined as: *d* (.01) = very small, *d* (.2) = small, *d* (.5) = medium, *d* (.8) = large, *d* (1.2) = very large [192]. Moreover, the CI indicates the precision with which we can calculate the SMD, facilitating the comparison between the studies. 89

4.4 (a) The Funnel Plot of the Random-Effects analysis without explanatory variables, and (b) the Funnel Plot with explanatory variables (i.e., study setting, RoIs, and conditions). In (a), the data points are expected to lay in a confidence triangle 95% of the time. However, this is not the case; an excessive amount of points are outside the interval due to heterogeneity. In (b), the Funnel Plot of the meta-regression model is shown. Herewith, we see that the plot is symmetrical in the vertical axis around the meta-analytic compounded effect size estimate. Suggesting that the moderators (i.e., study setting, RoIs, and conditions) do explain the differences between the studies; however, there is some excess of variance (> 5%). 91

5.1 (a) The Human-UAV control strategy developed by Drury and Scott [54]. Here, the authors describe the requirements with the Levels of Automation of the remote vehicles. (b) The decision ladder. That is, to generate and collect user requirements, Nehme et al. [157] proposed a decision ladder that intends to replicate the decision-making processes of the human operator. 99
5.2 A screenshot from the driver perspective: the image shows the starting point from which the participants have begun the user study. 102
5.3 A collection of 80 users' requirements was collected in the first interview. Flagged in red are the requirements that the participants in the second interview have chosen. 106
5.4 In some cases, a requirement category (e.g., vehicle position) has been selected in its entirety. While in other cases, only one category requirement was selected (e.g., vehicle speed). 107
5.5 An adapted illustration of Tilley's [224] anthropocentric data for adult males at the workstation. In yellow, we highlighted the areas where we envision the displays to be, i.e., a primary display arranged vertically and a second horizontally, both in front of the operator. 109
5.6 In this illustration, the activities for the second interview session are shown. The first activity was the requirement sorting. Each participant was asked to choose from the same set of requirements the information they thought they needed to conduct an AV safely (Figure 5.3, and Figure 5.4), and decide which requirements they would like to display on a primary or secondary display device. The second activity was the rating task. Each participant rated the chosen requirement on a 5-point Likert scale (1 = "unimportant" and 5 = "important"). 110
5.7 The descriptive plot displays the reported count frequency and the corresponding confidence intervals (set to 95%) for every rate. (a) The descriptive plot for the primary and (b) for the secondary display devices in shown. 111

5.8 In this plot, on the y-axis, the participant rating and the requirements are on the x-axis. Each requirement interval (categorical predictors) is shown with a vertical line and displays a dot or a triangle for the median. Thereupon: (1) 360° Remote view, (2) Charge/Fuel level, (3) Communication state, (4) Vehicle assigned to an area, (5) Vehicle to infrastructure, (6) Objects/Obstacles, (7) Vehicle issues, (8) Projected other road user trajectory, (9) Potential for collision, (10) Other sensors status, (11) Vehicle speed, (12) Projected control actions, (13) Task objectives, (14) Terrain features, (15) Traffic rules, (16) Vehicle characteristics, (17) Overall vehicle damage, (18) Vehicle position, (19) Projected vehicle position considering communication latency, (20) Weather condition. On the top left, the display type indicates the requirements selected for the primary and secondary display devices. . 112

6.1 An adapted version of the Takeover Request (ToR) diagram of Gold et al. [78]. Having detected a system limit, the ToR starts from the AV that sends a ToR to a human operator and ends when the operator takes control of the vehicle. The time between the vehicle handover control and the ToR is called Takeover Time (ToT), and the time between the system limit and the ToR is called Time Budget. 120

6.2 In this image, a screenshot from the nuScenes scenario we showed to the participants [19]. In particular, the AV detected an unexpected non-moving obstacle lying on its path, and since it is not allowed to drive over the solid marking lane, it sends a ToR to the teleoperator. 121

6.3 This illustration shows an egocentric 360° view of the remote vehicle. In particular, besides the front cameras (A, B, C), three additional back cameras are added (D, E, F). Other data, overlayed onto the video images, is available to the teleoperator, as the highlighted objects/obstacles of interest (G). Moreover, in a dedicated space (H), vehicles to infrastructure data may also be available to the operators, e.g., upcoming road construction or traffic light information. Here also we see status information, such as vehicle speed, vehicle charge/fuel level, communication state, i.e., the latency from and to the vehicle, other sensors status, and traffic rules signs, e.g., traffic-restricted zone or speed limit signs. 122
6.4 This illustration shows a 360° view of the remote vehicle. In particular, on the left, we see LiDAR point-cloud, and on the right, map-based information. At the bottom left, teleoperators, if required, can call a more detailed view of the ToR (A), and at the bottom right, operators can start and end the ToR (B). More other information is available to the teleoperator, as the objects/obstacles of interest (D), vehicle characteristics and position in relation to other objects (C), target destination (G), projected trajectory of other road users (F), e.g., of D. 124
6.5 The illustrations depict the two types of scenarios we planned to deploy during the user study. In particular, the type I scenario displays an unexpected non-moving obstacle lying on the vehicle path. In contrast, type II scenario shows a road closures sign due to, e.g., an upcoming construction site. 127
6.6 (a) The descriptive boxplots of the participant's takeover Time (TOT). (b) The descriptive boxplots of the SAGAT questionnaire. 131
6.7 (a) The descriptive boxplots of the subjective workload assessed via NASA-TLX questionnaire, and (b) the linear regression that predicts the SA index based on ToT values. 132

7.1 An adapted overview of the introspective failure prediction approach for semantic segmentation proposed by Kuhn et al. [114]. 136
7.2 To detect anomalies in the remote environment, an input image (a) is used to deliver a semantic prediction (b). Using a dedicated neural network, the semantic prediction can be synthesized (c) and compared to the input image (a). The comparison is shown in (d), which presents the semantic prediction errors. The results of an image-based failure prediction are used to visualize the anomalies for the teleoperator. The critical and the misclassified objects on the road are highlighted, allowing the teleoperator to be immediately aware of the AV challenge. 139
7.3 In this image, the explanatory scenario that includes the two independent variables of the user study, i.e., the registered visualization (a) and its absence (b). 140
7.4 The project architecture based on CAR Learning to Act (CARLA) [47]. The system relies on a client-server concept that provides on the client-side a flexible Python Application Programming Interface (API), and on the server-side, the rendering allowing for the association of multiple clients. 141
7.5 The driving scenario we administrated during the user study. In particular, in this image, we see the remote environments augmented with a warning signal close to each hazardous area. 142
7.6 In these descriptive plots, (a) we placed the SA as assessed by the SAGAT questionnaire [58] in relation with time and visualizations. Similarly, we compared (b) the perceived cognitive load assessed via the NASA-TLX questionnaire [92]. 146

8.1 Two remote driver assistance systems for vehicle teleoperation. In both concepts, the teleoperator's prediction of the remote vehicle position is shown as an overlay on the delayed video images. In the Free Corridor (a), the teleoperator sees a path of the full braking [32]. The Predictive Display (b) forecasts the movements of the remote vehicle and the movements of other road users to predict the vehicle position. 153
8.2 The novel curvature model proposed in this chapter. In prior works, authors exclusively consider the remote vehicle inputs, i.e., the velocity and yaw rate, to predict the remote vehicle position. In contrast, we propose a novel curvature model that considers both the teleoperator inputs (in red) and the remote vehicle states to calculate the curvature value. 154
8.3 The Predictive Corridor combines into a single virtual augmented element the Predictive Display (in blue) as proposed by Graf et al. [87] and the Free Corridor (FC), in green, as proposed by Chen [32]. The prior is based on the clothoid method, the latter on the circular arc method. The total length of the FC is calculated by adding (1) video transmission delay, 2) operator's perception-reaction delay, 3) control delay, 4) brake activation, 5) brake distance. 155
8.4 The experimental setup. In this image, we see (1) thermal camera Optris Xi 400, (2) three 23,8-inch LCDs, (3) Tobii Eye Tracker 4C, (4) Logitech G29 steering wheel and pedals, and (5) experimenter computer. 156
8.5 User studies procedure line-up. The whole user study consisted of three blocks. In the first part, participants were briefed, and the temperature baseline was assessed. In the second part, we administered two tasks. Lastly, we debriefed the participants. . 160

8.6 In this illustration, we see the two tasks administrated, namely the Emergency Braking Stop (EBS) and the Lane-Change (LC) maneuver. The PC, on the top right, shows to the teleoperator the predicted position (1) of the remote AV and the path of the full braking (2). The width of the PC (3) corresponds to that of the AV. 161
8.7 Image (a) shows the forehead and nose temperature (difference to the baseline) over the Lane-Change (LC) maneuver and the Emergency Braking Stop (EBS) task. Both are experienced with and without Predictive Corridor (PC). Image (b) shows the subjective assessment of workload via NASA-TLX. 162
8.8 The image a) shows the gaze patterns of one participant during the Lane-Change (LC) maneuver when performing without the Predictive Corridor (PC). Furthermore, the b) image shows the gaze pattern of the same participant during the LC using the PC. In the image c), the gaze patterns during the Emergency Braking Stop (EBS) task without PC, and in the image d), the gaze plot during the EBS task with the PC. To provide clarity and contextual understanding, it is important to explicitly mention that the images depict a typical snapshot of a random participant's gaze patterns during the maneuver. This clarification ensures that readers understand that the images serve as representative examples of the gaze patterns observed during the maneuver. . 163
8.9 The image (a) displays the descriptive boxplots of the cones collision number for the Lane-Change (LC) maneuver. In image (b), the distance to the stop line on the Emergency Braking Stop (EBS) task can be seen. 164
8.10 In image (a), a linear regression predicts the number of cones collisions based on the number of trials during the Lane-Change (LC) maneuver. In image (b), a linear regression predicts the distance to the stop line based on the number of trials during the Emergency Braking Stop (EBS) task. 166

8.11 In this illustration, we can see the graph of the lateral control conduction with two changes of lanes. In particular, the first lane change occurred when the participant had to steer the vehicle to the left. The second lane change occurred when the participant returned to their original lane by steering the vehicle to the right. 167

9.1 The images show the gaze-based remote camera control designs by Zhu et al. [257] and Zhu et al. [258]. In the image (a), the pan and tilt of the camera are mapped along with the head movement. In image (b), the double arrow stands for the rapid movement the user has to perform to turn on/off the camera to move in the desired direction. Image (c) shows that the teleoperator's gaze can move the image to the center of the display. 172

9.2 This figure illustrates a comparison of two control interfaces for additional video feedback in teleoperation, i.e., continuous vs. discrete mode. Figure (a) visualises the continuous control mode: here the video feedback is dynamically linked to the operator's head movement relative to the centre of the screen. As the user turns their head to the left or right, away from the central area, the camera view intuitively pans to follow the direction of the operator's gaze, providing a continuous view of the corresponding side camera, like a smoothly guided tour of the environment. A label indicating the camera name and its spatial orientation relative to the vehicle (e.g. 'rear left') is provided at the bottom centre of each video feed for orientation. Figure (b) highlights the discrete control mode: this interface divides the display into specific active zones. When the operator's gaze falls on one of these zones, such as the top left for the rear side camera, the designated camera feed is superimposed on the current front camera view. 176

9.3 The project architecture based on CARLA [47]. The system relies on a client-server concept that provides a flexible Python Application Programming Interface (API) on the client-side, and on the server-side, the rendering allows for the association of multiple clients. 178

9.4 The experimental setup. In particular, in this image, we see (1) thermal camera Optris Xi 400, (2) three 23,8-inch LCDs, (3) Logitech G29 steering wheel and pedals, and (4) experimenter computer. 179

9.5 The driving scenario we administrated during the user study. In particular, in this image, the ego-vehicle encounters a broken standing vehicle ahead. 181

9.6 Overview of the user study flow. The user study was conducted in a single session and comprised a series of consecutive activities. Initially, participants received an orientation briefing and engaged in a practice session to familiarize themselves with the prototypes. Following this, participants underwent baseline temperature assessments for physiological workload measurement. Subsequently, the main evaluation consisted of two driving scenarios, one for each camera control interface (discrete and continuous). These scenarios were administered in a counterbalanced order using a Latin Square design to control for order effects and participants shifted between driving and non-driving tasks within the same session. After completing the tasks, a debriefing session concluded the study. 183

9.7 The image shows the boxplots summarising the pre-task assessments of Situation Awareness and workload. In particular, Figure (a) shows the distribution of scores from the SAGAT questionnaire, which was administered before participants began the driving task to assess their baseline situational awareness. Figure (b) shows the subjective workload ratings provided by participants using the NASA-TLX questionnaire, which was also administered prior to the driving task to assess participants' perceived workload in anticipation of the activity. Figure (c) shows the results of an objective workload assessment using thermal imaging to analyse physiological markers of stress, such as changes in facial temperature patterns, measured before participants engaged in the driving task. 184
9.8 This plot shows, for both control interfaces (continuous and discrete), the driving errors made by the participants during the five trials. The red intervals show the standard errors with a dot for the participants' driving error mean. 186
9.9 Image (a) shows the descriptive boxplots of the subjective workload was evaluated via NASA-TLX questionnaire during the driving task. In image (b), the descriptive boxplots of the objective workload was evaluated via thermal imaging during the driving task. 187
9.10 Image (a) and (b) shows the linear regression that predicts the cognitive index when performing the driving task. In image (a), the outcome is based on the participant's forehead/nasal temperature difference when asked, before driving, to examine the traffic scene. In image (b), the outcome is based on the participant's NASA-TLX mean when asked to examine the traffic scene before driving. 189

10.1 A visual summary of the research findings. 196
10.2 The chart flow diagram of the system we developed. In bold are the topics handled in this research. 197

List of Tables

1.1 The table provides an overview of the six levels of driving automation [37]. 14
1.2 The table provides a structured summary of the research questions addressed [169]. 17
1.3 The table provides a structured summary of the main contributions, highlighting the collaborator, as well as the research papers and the patents derived from it. 20

2.1 The table outlines the literature concerning the principal factors that affect remote operations. In particular, a summary of the findings of video image quality, Field of View (FoV), camera viewpoint, orientation, depth perception, and the time delay is provided. 38
2.2 The table resumes the review of the current teleoperation HMI concepts. We divided the literature findings in interfaces for teleoperation *inputs* and interfaces for teleoperation *feedback*. In particular, we include controls-based, touch-based, gesture-based, speech-based, gaze-based, and physiological-based input modalities. For feedback modalities, we include visual devices, auditory devices, and haptic/tactile devices. 52

2.3 The table summarizes the qualitative requirements from the literature research on vehicle teleoperation embracing hardware and software solutions. 59

3.1 The table provides an overview of the reviewed Situational Awareness (SA) models. In particular, we report the SA frameworks developed to explain individual and distributed decision-making processes. 70
3.2 The table summarizes the different methods for measuring Situation Awareness. 72

4.1 The table shows an overview of the thermal cameras used in the reviewed literature. Sensitivity is specified as the temperature difference where a change is indistinguishable from noise, the Noise Equivalent Temperature Difference (NETD). 84
4.2 A detailed overview of the participants recruited in the reviewed literature. Note: some research papers have separated the participants into groups of age: *a)* two groups of ages: 18-27 and 60+, *b)* two groups of ages: 18-27 and 55+, gender-balanced, *c)* two groups of ages: 18-35 and 36-64. 87
4.3 The table shows the estimated coefficients (via Wald test) for the Random-Effects analysis, along with their significance. Here, next to the intercept, one coefficient for Setting (in situ), two for ROIs (Nose Tip and Perinasal), and three for Conditions: Drive + Cognitive Distraction (C.D.), Unexpected Event (U.E.), and Sensorimotor Distraction (S.D.). Note: SE stands for Standard Error. 90

5.1 The contingency table shows the frequency distribution of two variables, rate and type. Here type indicates the requirements selected for the primary (Type 1) and secondary (Type 2) display devices. 113

6.1 The descriptive statistics of the variable tested during the first user study, in particular, those of the SAGAT queries [58], Takeover Time (ToT), and NASA-TLX questionnaire [92] are reported. . . 130

7.1 Descriptive statistics of the subjective assessment of workload assessed via NASA-TLX and Situation Awareness (SA) via the SAGAT questionnaire. Note: SA stands for the SAGAT questionnaire, NT for the NASA-TLX, R for registered, N for none, 7 and 10 is the duration. 145

8.1 The descriptive statistics of the participant's driving performance with and without the Lane-Change (LC). In particular, we recorded the number of cones collisions for the Lane-Change (LC) maneuver, whereas the distance to the stop line for the Emergency Braking Stop (EBS) task. 165

9.1 Descriptive statistics of the subjective and objective assessment of workload assessed via NASA-TLX and thermal imaging. . . . 188

Erklärung der Urheberschaft

Hiermit versichere ich, dass ich die vorliegende Dissertation selbständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich aus der Literatur oder aus anderen Quellen übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

München, Wed. 12 Feb. 2025, G. Graf

Ort, Datum, Unterschrift

FAR OUT
INTERFACE AND INTERACTION CONCEPTS FOR TELEOPERATED AUTONOMOUS VEHICLES

As vehicle automation technology progresses and autonomous ride-hailing fleets begin test operations, practical challenges for offering such services emerge. Situations like infrastructure failure, interpreting traffic controls (e.g., police hand signals), or navigating obstacles may require human assistance. Teleoperation has been proposed as a solution but faces challenges such as high latency and low situation awareness. This dissertation evaluates interface and interaction concepts to enhance teleoperators' spatial perception, state awareness, and decision-making.

Designing a safe and ergonomic teleoperator workplace began with collecting first-hand operator requirements to address the question: What information enhances situation awareness when remotely controlling an autonomous vehicle? These insights were translated into functional design concepts.

We designed and evaluated user interfaces for take-over requests to facilitate vehicle control and developed software enabling teleoperators to quickly identify the cause of these requests. Additionally, we introduced a camera control system that predicts and displays the most relevant feed based on the operator's head position and gaze. Finally, we created the Predictive Corridor, an assistance system offering decision support by showing the vehicle's predicted position and path during network losses.