Multimodal Presentation to Support Human for Resolving Semi-Autonomous Agent Failures Using Mixed Reality and Transfer of Control

by

Umar Farooq

Master Thesis

Faculty of Natural Sciences and Technology I
Department of Computer Science
Saarland University

Supervisor
Prof. Dr. Antonio Krüger

Advisor
Yannick Körber

Reviewers
Prof. Dr. Antonio Krüger
Dr. Arshad Khan

April 24, 2019

UNIVERSITÄT DES SAARLANDES

Declaration of Authorship

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Statement in Lieu of an Oath

I hereby confirm that I have written this thesis on my own and that I have not used any other media or materials than the ones referred to in this thesis.

Einverständniserklärung

Ich bin damit einverstanden, dass meine (bestandene) Arbeit in beiden Versionen in die Bibliothek der Informatik aufgenommen und damit veröffentlicht wird.

Declaration of Consent

I agree to make both versions of my thesis (with a passing grade) accessible to the public by having them added to the library of the Computer Science Department.

Datum/Date: ________________________________

Unterschrift/Signature: ________________________________

SAARLAND UNIVERSITY
Department of Computer Science

Abstract

Multimodal Presentation to Support Human for Resolving Semi-Autonomous Agent Failures Using Mixed Reality and Transfer of Control

by Umar Farooq

Modern technology has brought many wonders to this century and continues to do so albeit the growth curve still lacks the leap expected by scientists. Today, we are living in the hyperdimensional era of machines and software, from home automation to autonomous cars and automated industries we have seen the revolution in almost every sector. One such revolution is termed as "Industrie4.0" focusing towards the automation and digitalisation of manufacturing. Although automation has successfully become an integral part of industrial manufacturing, there are still some challenges towards successful implementation of automation such as autonomous agent's failures causing an interruption in the production. One solution to this problem can be human agents helping the autonomous agents to recover from failures. This work provides a framework focused on the generation of generalised multimodal interactive presentation for supporting a human to resolve semi-autonomous agent failures by implying the technique of transferring control. Mixed Reality(MR) is used as a tool to provide a visual aid to the human agent along with its spatial and interactive capabilities. The user study based evaluation of the framework provided encouraging results to consider it as a bridge between human and machine to resolve autonomy failures and a significant step towards the successful implementation of automation in the industry.

Content

Declaration of Authorship **iii**

Abstract **v**

1 Background **1**

1.1 Autonomous and Semi-Autonomous agents . . . 1

1.1.1 Autonomous Agents . . . 1

1.1.2 Semi-Autonomous Agents . . . 2

1.1.3 Dependability of Semi-Autonomous Agents . . . 2

1.1.4 Failures of Semi-Autonomous Agents . . . 4

1.2 Extended Reality(XR) . . . 5

1.2.1 Virtual Reality(VR) . . . 6

1.2.2 Augmented Reality(AR) . . . 6

1.2.3 Augmented Virtuality(AV) . . . 7

1.2.4 Mixed Reality(MR) . . . 7

1.2.5 Building Blocks for Mixed Reality . . . 8

1.2.6 Mixed Reality (MR) Devices . . . 10

1.3 Human Computer Interaction (HCI) for Mixed Reality (MR) . . . 12

2 Related Work **15**

2.1 Instructional Interface in AR/MR . . . 15

2.2 Navigation using AR/MR . . . 17

2.3 A Human Centeric approach to autonomous robot failures . . . 18

2.4 Asking for help using Inverse Semantics . . . 19

2.5 Resolving automated system failures in bin-picking tasks using assistance from remote human operator . . . 21

2.6 Autonomous driving: Investigating the feasibility of car-driver handover assistant . . . 22

2.7 Takeover times in highly automated driving . . . 22

2.8 Comparison of Related Works . . . 23

List of Figures 25

Bibliography 27

Study Questionnaires 31

Chapter 1

Background

This work is comprised of different fields of computer science and needs an understanding of different foundation concepts. This chapter provides necessary concepts that will further help in the understanding of this work.

1.1 Autonomous and Semi-Autonomous agents

Both autonomous and semi-autonomous agents are smart agents who can make decisions based on their interactions with physical or virtual environments. Many machine learning and artificial intelligence techniques make those agents capable of learning and taking decisions for different situations.

1.1.1 Autonomous Agents

Autonomous agents are those smart agents who perform tasks for their owners without accepting any instructions or commands from ownership entities. The Autonomous agents interact effectively and independently with their environment. For effective execution of tasks, an agent should have information about the surrounding environment.

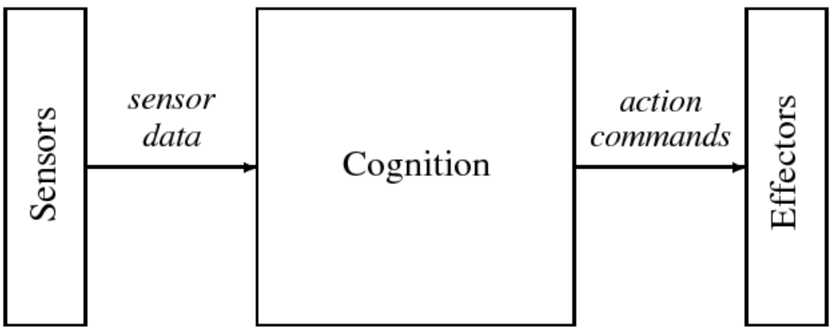


Figure 1.1: Simple architecture of autonomous agent [1]

The Figure 1.1 shows simple architecture of an autonomous agent [1]. The autonomous agent takes environment information through sensors. The cognition module takes sensory information and decides about required actions for argent. Effectors take actions on receiving commands from the cognitive component.

1.1.2 Semi-Autonomous Agents

As from the name of semi-autonomous agents, it is clear that they have partial autonomy. In simple words, we can say that these agents cannot work on their own all the time. They need external intervention at some point in time. Currently, most autonomous systems have some probability of failing. The real world environment is very dynamic with numerous situations that are impossible to handle by smart agents on their own.

For example, circuitry failure of an autonomous agent needs human repairment to continue to work again. There is always a possibility of situations when an autonomous agent faces failures and needs human help for recovery. True autonomy is difficult to achieve. We can say that currently, most autonomous agents are in fact, semi-autonomous agents.

1.1.3 Dependability of Semi-Autonomous Agents

According to Dubrova [2] and Laprie [3] Dependability is the capability of a system to provide the intended degree of services to its users. Measures or attributes, impairment and means are three essential characteristics of dependability.

Dependability Measures or Attributes

Reliability, availability, and safety are the three main attributes. Security, performability, testability, and maintainability are other possible attributes. Based on the application, one or more of these characteristics might be required for appropriate assessment a system behaviour [2].

Reliability: It is a measure of the continuous delivery of correct service. High reliability is required in situations when a system is expected to operate without interruptions. Given that a system was working correctly at time 0, then the probability that system will work without failure at time t during the interval [0, t] is known as the reliability of that system [2].

Availability: It is a degree of service accomplishment regarding the alternation of interruption and accomplishment [3]. Relatively few systems are created to run continuously with no interruption and with no maintenance of any type. For such systems, the desired attribute to maximise is the tiny proportion of time that the process is within the operational status, conveyed by availability. [2].

Safety: It can be viewed as an extension of reliability, specifically reliability regarding problems which can produce safety hazards. There is no distinction in failures from the reliability point of view. For safety considerations, failures are partitioned into the fail-unsafe and fail-safe ones. Formally, if a system was working correctly at time 0, then at time t the probability of a system to perform its work correctly or discontinues operation in a fail-safe state during the interval [0, t] is referred as the safety of a system [2].

Dependability Impairments

Dependability impairments are typically defined in the terminology of faults, errors, and failures. A typical characteristic of these terms is giving a sense of something went wrong. Failure, error and fault are referring terms to undesirable behaviours of autonomous agents. There are several existing definitions for these terms in literature.

Fault: Carlson *et. al.* [4] explained fault as anything which could cause the system to enter an error state. According to Dubrova [2] "A fault is a physical defect, imperfection, or flaw that occurs in some hardware or software component"

Error: Error is a state of system that leads to failure [4]. Dubrova [2] explained error as a result of a fault that deviates the computation from correctness or accuracy. Errors are typically connected with erroneous values in the system state.

Failure: According to Dubrova [2] "A failure is a non-performance of some action which is due or expected. A system is said to have a failure if the service it delivers to the user deviates from compliance with the system specification for a specified period of time." also according to Brooks [5] Failure is "a degraded state of ability which causes the behaviour or service being performed by the system to deviate from the ideal, normal, or correct functionality".

Dependability Means

Dependability means are the approaches and strategies allowing the development of a dependable system. It includes fault-forecasting, fault-removal, fault-tolerance, and fault-avoidance.

Fault Tolerance: The focus of fault tolerance is to develop systems which can operate adequately while faults are present. Usually, there is some sort of redundancy to attain fault tolerance. Redundancy can either mask the faults, or detect them, with consequent location, containment, and recovery [2].

Fault Forecasting: Fault forecasting is a set of methods aiming to estimate the number of faults that are present in the system, potential future occurrences and effects of faults. Fault forecasting is accomplished by executing an analysis of the system regarding fault activation or occurrences. The evaluation may be qualitative or quantitative [2].

Fault Avoidance: It is the collection of techniques which try to stop the introduction of fault in a system. Different hardware and software quality control techniques are used to prevent faults [2].

Fault Removal: It is the removal of faults present in the system by using different techniques. Fault removal is performed in the development and operation phases. In the development phase, it includes verification to check if the system is working correctly. In case, it doesn't then diagnosis and correction follows the verification. During operation of the system preventive and corrective maintenance is performed to remove faults [2].

1.1.4 Failures of Semi-Autonomous Agents

As we know faults can cause failures and can have an adverse effect on the dependability of autonomous systems. One should know the nature of fault to deal with them and make the system more dependable and autonomous [6]. Because of the possibility of many problems that can arise in autonomous robots, it becomes nearly impossible to identify and address all the problems [5]. As of limited available knowledge about common types of faults and problems in robots, it is difficult to specify methodologies to tackle robot failures [6].

G. Steinbauer [6] performed a survey with the teams who took part in RoboCup-2013 with a vision of providing the basis for the broader evaluation and classification of faults in robots. They collected data about general context data about used robots, faults mitigation strategies and possible faults in different robots. The survey showed that most of the teams were using custom robot platforms. Sensors were affected more by connectors. Manipulators are more sensitive parts of robots and are profoundly affected by faults. The fault in sensors and robot platform has almost the same frequency of occurrence, but sensor faults impact more negatively. The connector problems are more frequent with impact on multiple things. Connectors also affected control system faults. All surveyed groups have a good awareness of faults and dependability. They all used some techniques like watchdogs and monitors to tackle failures. Steinbauer also finds that more research required in the direction of validation and evaluation of algorithms, also toward configuration management. State machines based algorithms were more prone to faults.

1.2 Extended Reality(XR)

Extended reality is a broad term that covers all reality-based technologies or terms. Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR) all are included by XR. To understand all these technologies and terms reality-virtuality continuum is shown in Figure 1.2. This continuum ranged from the wholly virtual environment to complete the real environment. So all different variations of real and virtual objects can lie somewhere on the continuum.

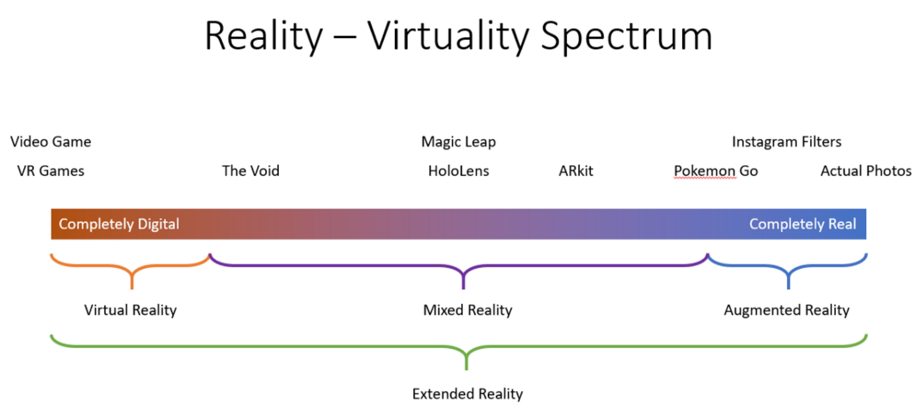


FIGURE 1.2: Reality-Virtuality Continuum [7]

1.2.1 Virtual Reality(VR)

Virtual Reality is more of a science fiction motivated technology. The goal of VR is to provide a feeling of another world to a human without the human's physical presence in that world. The ultimate goal of VR is true immersion. The human user in VR truly feels himself in another world without travelling physically to that world.

FIGURE 1.3: User playing video game in Virtual Reality(VR) [8]

Figure 1.3 shows a human user playing a video game in a virtual environment while wearing a VR head-mounted device. With the stereo sounds through headphones and graphics in VR glasses provide a genuine feeling of being present in the video game.

1.2.2 Augmented Reality(AR)

Augmented reality is a technology that deals with real-world environments. In this technology digital information is presented in the real world. Instead of going virtually to another world human gets digital information in the real world. The literal meaning of augment is "increase or expand" so augmented reality means expanding or increasing the reality. The famous example of AR includes "Pokemon Go" game where players try to catch Pokemon at different locations in the real-world environment.

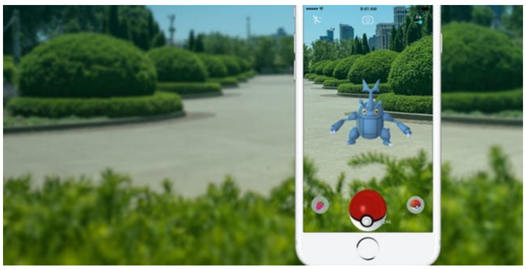


FIGURE 1.4: Example of Pokemon-Go, AR Game

1.2.3 Augmented Virtuality(AV)

Augmented Virtuality (AV) enhances the virtual experience. It is the opposite of AR. The examples of leap motion or Kinect devices are best suitable to explain the term AV. Users make different gestures in front of Kinect or leap motion devices and interact with the virtual environment.

1.2.4 Mixed Reality(MR)

Mixed Reality(MR) is a merger of Augmented Reality (AR) and Augmented Virtuality(AV). The rank of MR is somewhere between AR and VR. Introducing MR also introduce confusions. According to Jack [7], "ARkit" from Apple lies in the domain of mixed reality but have AR in its name. The classification shown by reality-virtuality continuum shows exclusiveness between AR and VR that is not the case in actual. Figure explains reality-virtuality continuum using a Venn diagram. If we need special headsets for visualisation we call it VR, reading real world with the help of computer vision is AR, if we do both things, getting data from the environment and shows through some headset then we call this combine AR and VR and termed as Mixed Reality (MR).

Still, confusion is present between AR and MR even on the corporate level. Microsoft markets Hololens device as mixed reality device while another company Meta2 markets similar device as AR device. There seems a lack of agreement on the use of these terms. Based on the aforementioned definitions and study we consider these devices as MR devices.

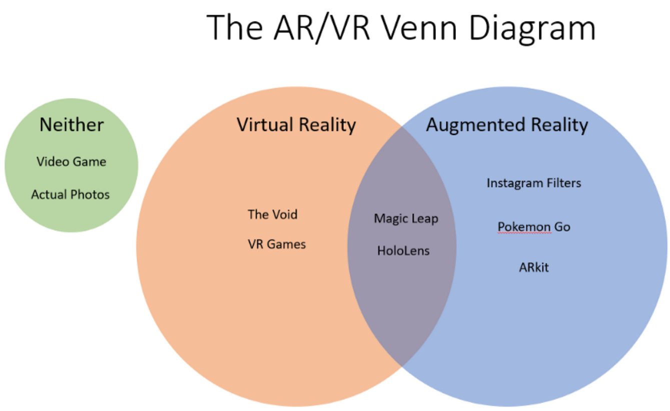


FIGURE 1.5: Venn Diagram of Reality-Virtuality continuum [7]

1.2.5 Building Blocks for Mixed Reality

To understand the working of Mixed Reality(MR) devices its good to have knowledge of common terms and vocabulary related to MR. Below are a few terms explained briefly.

Gaze

Gaze is the basic building block for interaction in MR. It provides information about the position where a human user is looking through the device. In most of MR devices, instead of tracking human eyes, the location and direction of a user's eye are obtained by using gaze the head position and orientation. Gaze can be described as a ray in the forward direction from the centre of both eyes. It is also the basic input source for interaction. For example, hiding certain objects by looking at them for some time. It is also used with other input sources to achieve certain functionalities. Use of cursor at Gaze position can provide real-time feedback to the user.

Gestures

Gestures is another building block for MR interactions. Usually, the user brings hand in front of the MR device and perform pre-defined gestures. MR devices detect those gestures using their sensors. The gesture is a great way for giving input to the device as it provides a better sense of interaction and control. Considering the example of placing an object at the user's eye gaze, the user can tap when an object is reached at the desired location by the following Gaze.

Voice Commands

Voice commands provide another interaction interface. For mixed reality devices like Hololens, it is one of the three essential input types. It gives the capability of interacting with elements without using gesture and hence allows the user to do interactions while hands are free. A user places the gaze at an interactive element and speak the command. It can be considered a natural way of interaction.

Spatial Mapping

Spatial mapping is the map of the real-time environment. It can be both visible or invisible in applications based on requirements. The spatial map provides exact locations to place holograms in the real-time environment so that placement looks more natural and precise. To perform spatial mapping sensors are assembled within the head-mounted device. Small triangular shapes are joined based on sensory information to make the whole map of the environment. Figure 1.6 shows a spatial map of the real environment. The map is visible as wire-frame mesh that reconstructs the surface of the room.

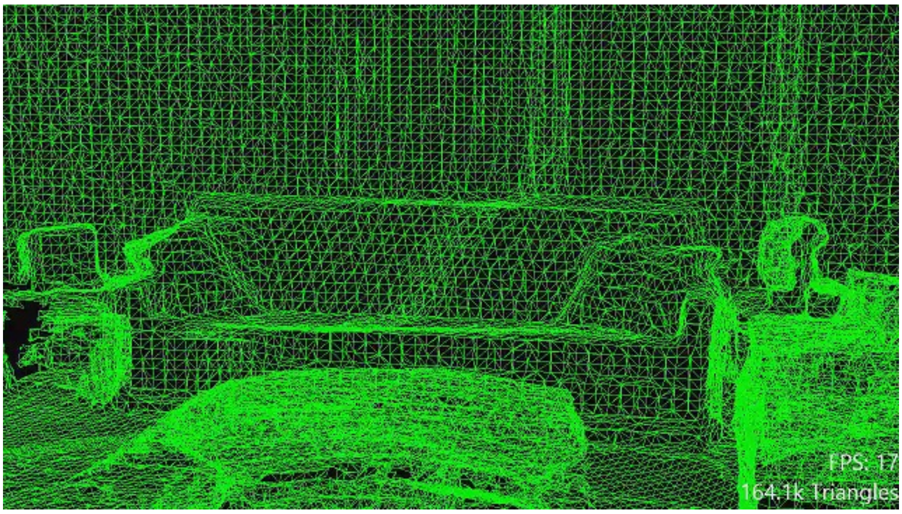


FIGURE 1.6: An example of a spatial mapping mesh covering a room

Spatial Sound

The sound is one of the ways to perceive the things, not in front of our eyes. Spatial sound is 3D sound simulated based on environmental values. It uses direction and distance of the object from where the sound is coming. The circular dimension of head-mounted MR devices simulates the sound at angle values facing the object. For example, if a human is looking at a hologram in front of the eyes and a virtual assistant is on the back side of a human user. The audio message from a virtual assistant is simulated in a way that human perceives the sound is coming from the backside of view.

Spatial Anchors

Spatial anchors are important for elements position in real world. Spatial anchors are assigned to the object that needs to be placed at a fixed position in the world. For example, to place an object at a specific position permanently in the application, manipulate the object and place at the desired position. After placing the object, a spatial anchor should be assigned. If no spatial anchor is assigned and the application is restarted, the object will reset to initial position instead of maintaining the assigned position.

1.2.6 Mixed Reality (MR) Devices

Mixed Reality is rapidly growing and various kinds of devices are actually introduced in the market. Following is the brief explanation of some famous Mixed Reality (MR) devices available in the market.

1-Magic Leap

Magic Leap [9] very famous startup launched their first mixed reality device in July 2018. Its a wearable head-mounted device.

Features: Magic leap combines the natural light waves and a soft layer of artificial light to provide an excellent mixed reality experience. It has contextual awareness and the spatial sound of the surroundings. Magic Leap makes virtual objects to know about the surrounding walls, tables, chairs, etc. It uses a custom built operating system luminous. LuminOS is optimised for environment recognition. Magic Leap does not have impressive Field of View (FoV). Its FoV is actually up to 50 degrees.

Specifications and Hardware: Magic Leap has two processing units, a 64-bit NVIDIA CPU and a GPU and 8GB of RAM. It has a small portable device which has all processing units. The user can connect the device with the waist belt. Magic Leap has a 128GB of storage capacity. It has lithium ion battery that can sustain up to 3 hours. Magic Leap has Bluetooth, USB-C and WIFI connectivity modes.

Inputs: It has hand tracking for gesture detection, Eye tracking and voice recognition.

2-Meta

Meta [10] is also a device to interact with virtual objects in the real world. Meta markets itself as Augmented Reality device. In September 2014 Meta 1 was released, and Meta 2 was released in early 2016.

Features: Unlike other famous MR devices, Meta is tethered head-mounted device. So it is to work in a fixed location. Wearing a device and moving around is not possible with Meta. As Meta 2 uses the processing power of traditional PC, sensory data is passed to PC via tethered wire and processed there. The bigger size of Meta results in a larger field of view (FoV). Meta2 has FoV up to 90 degrees.

Specifications and Hardware: Meta uses the processing power of a host PC. It does not have any built-in processing units.

Figure 1.7: From Left; Magic Leap, Meta, and Microsoft Hololens [11]

3-Microsoft Hololens

Hololens [12] is a wearable head mounted device developed by Microsoft and widely used as compared to other devices. Hololens doesn't need any tethering. Hololens is adjustable for different head sizes. The user can wear Hololens by adjusting wheel on the backside of the headband.

Features: Hololens has the capability of spatial scanning to make itself aware of surroundings. It also supports spatial sounds. Holograms are shown as a part of the real world. The FoV for Hololens is not very impressive and approximates to 35 degrees. It does not have any fans and gets cooled passively. It has rechargeable battery powered with 2-3 hours of active time and around two weeks standby time.

Specifications and Hardware: The front side of Hololens has sensor related hardware including 3D camera and 12MP of RGB photo/video camera. It has two processors, first is Intel 32 bit central processing unity (CPU) and the second custom made holographic processing unity (HPU). It has 64GB of storage and 2GB of RAM.

Inputs: Hololens support multiple input methodologies for interacting with 3D interfaces. It supports Gesture Detection, Gaze Tracking and Voice Inputs, sometimes referred to as "GGV" inputs. It also has a clicker button that pairs with Hololens using Bluetooth connectivity.

1.3 Human Computer Interaction (HCI) for Mixed Reality (MR)

The interface design is an important aspect to achieve seamless interaction in MR/AR environments. D. Andreas *et. al.* [13] tries to apply HCI rules of human-centric designs for AR interfaces. They think generalisation is the biggest problem in defining specific design guidelines for AR. The reasons are a large number of AR system ranging from mobile devices to head-mounted devices and AR device-specific interaction techniques or input/output devices. According to Andreas *et. al.* [13] AR designers have to produce solutions for individual problems, but still, specific guidelines can be applied for human-centric designs. According to the following human-centric design, principals can be applied in the AR interface design.

- **Affordance:** Wisely chosen interaction metaphors can help a user to interact with virtual objects nicely. Interaction devices registered in 3D space are also good choices for interaction.
- **Cognitive Overhead Reduction:** learning and training can be reduced with cognitive overhead. To reduce cognitive overhead user must interact with the application.
- **Less Physical Work:** In case of HMDs user may feel uncomfortable after prolonged use, so it is important to consider use time while designing.
- **Learnability:** Intuitive interaction techniques, consistency and self-descriptiveness of elements can enhance learnability.
- **Satisfaction of User:** Interface should be able to engage user along with solving the main task. The interface usability depends on subjective user perceptions as well as objective measurements.
- **Flexible to Use:** Considering users abilities and preferences different modalities can be integrated into an application by using multiple interaction techniques and I/O devices.
- **Responsiveness and Feedback:** User must get immediate feedback for performed actions if no responsiveness or feedback in a particular time user will not be able to develop the persistent cause-and-effect model.
- **Error Tolerance:** Identifying and resolving errors can increase the robustness of application.

Consideration of explained generalised design principals for AR systems by D. Andreas *et. al.* [13] can help design process to result in an application with good usability.

S. Ganapathy [14] provided design guidelines for mobile augmented reality (MAR) by considering three main segments as usage scenario, modalities for interaction, and types of products. She explores use scenarios such as shopping, navigation associated with a personal guide and tourism, personal translator and gaming with social network situation. Voice/audio, touch as well as tactile feedbacks are considered for interaction modalities whereas digital camera, smartphones, and ultra-mobile unit considered for device form factors. S. Ganguly [14] provided primary guidelines for creating MAR user interfaces. Based on the guidelines, they emphasize to make sure that the textual info is actually clear with easily readable fonts and also have high contrast between the text as well as the background. Design in such a way that text has good visibility in different backgrounds. Keep info organised by grouping. Vital information should use placement in a means it remains clear. Sensitive info like emergency actions by a user must be recognisable immediately. She additionally stressed on the usage of unique icons for easy information accessibility along with different filters based on visibility and distance.

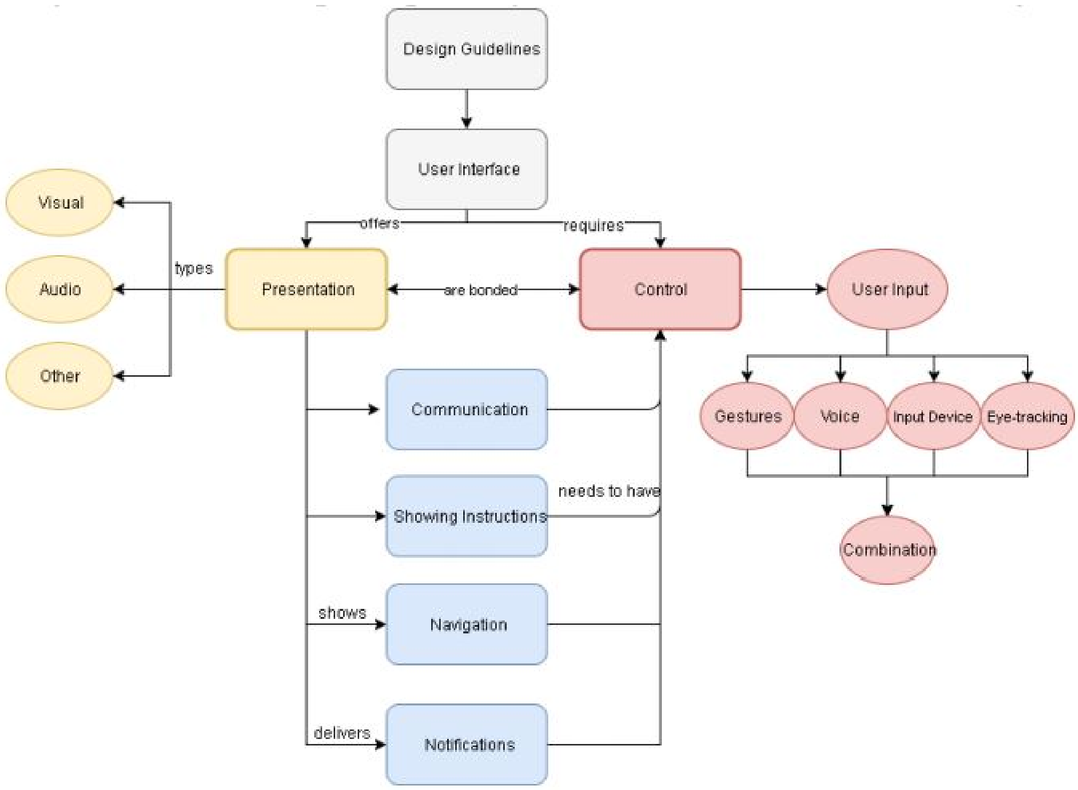


FIGURE 1.8: Concept Map of AR interface design key elements

Jakub [15] worked to provide detailed design guidelines for Augmented Reality aiming to provide the optimal way for showing data and creating user interfaces focusing industrial applications and keeping limitations of AR devices he provided necessary guidelines.

According to Jakub it is better to use combination of 2D and 3D interfaces based on requirements. Keeping in view the industrial applications he focused on repairing work with the use case of repairing an aircraft engine. Jakub provided nice concept map of essential elements in AR interface designing. The figure shows the concept map by Jakub [15] and is very much related to our work in this thesis. The user interface has two main components, presentation and control; both are bonded to each other. Presentation modes can have visual, audio whereas control is about inputs from the user that includes gestures, speech commands, dedicated input devices or eye-tracking. He further provided guidelines about showing instruction, providing navigation, notification and communication.

Chapter 2

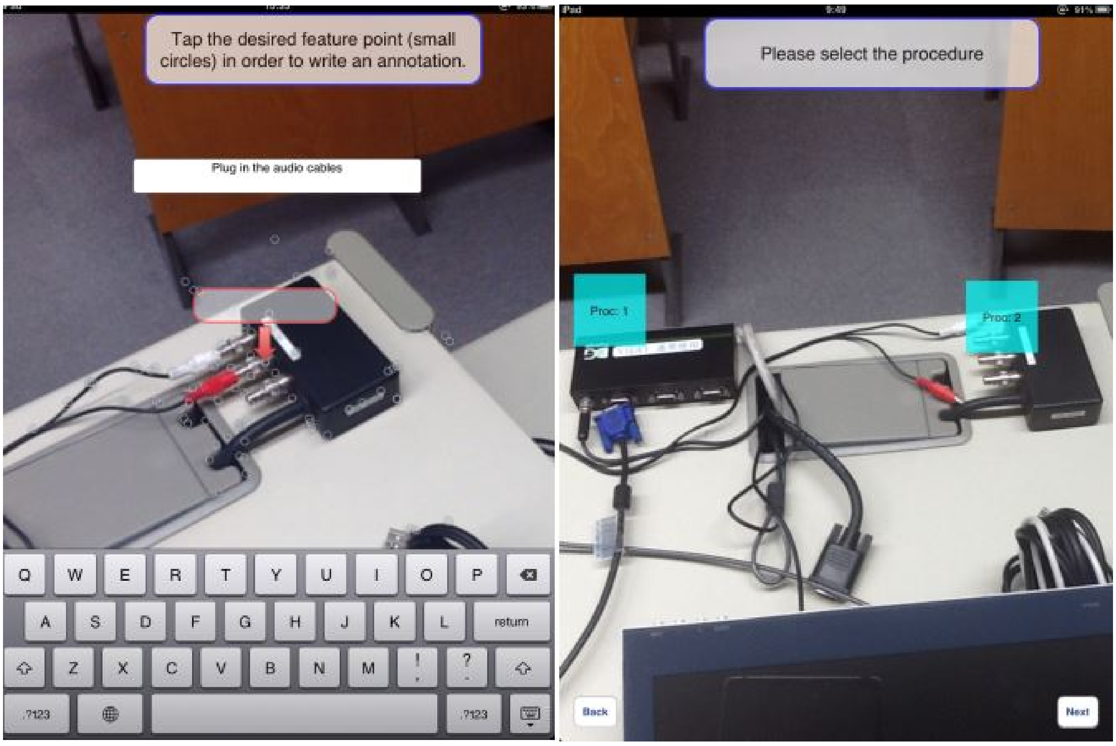
Related Work

This section briefly describes some related works and their approaches to solve an autonomous agent's failures. Although there is not much work present that uses the transfer of control approach for handling autonomy failures however below are some related approaches.

2.1 Instructional Interface in AR/MR

Jakub [15] provided guidelines for showing repair work instructions to the user in AR. According to him, there should be various ways of showing instructions like holographic objects, texts, sounds. Allow the user to hide and show the manual. Make manual diverse, means more details for a worker with less experience and vice versa. Provide feedback to the user if he makes a mistake. If a worker needs to move an object, then emphasise that object by highlighting. Jakub also suggests showing the user the required tools before they are needed. Provide Correct indication for moving parts. He suggested simple animations for this purpose. Complex animations can create confusions. Consider extreme working conditions, like speech commands, may have a recognition problem in noisy areas. He also suggests enabling remote communication for getting an expert opinion while repairing.

J. Pilvi *et. al.* [16] worked on designing the user interface for state-of-the-art pragmatic simultaneous localization and mapping (SLAM) based handheld augmented reality (HAR). They performed user study for their prototype system. The purpose of their system is to provide instructions for different work-related tasks like the use of equipment and different devices. Their system has two modes, viewing mode and authoring mode. In authoring mode, different parts are annotated, and practical instructions are added along with tracking markers, and in viewing mode, instructions can be seen for annotated parts. The main idea is that once an author annotates the task with virtual instructions then it is useable for users in viewing mode to perform the task by following textual instruction and guidance. Figure 2.1a shows authoring mode of the system developed by Polvi [16] where author enters input for virtual instruction whereas figure 2.1b shows viewing mode where the user selects one of two tasks and then further instructions will appear annotated during authoring mode.



(A) Authoring mode (B) Viewing Mode

FIGURE 2.1: Prototype System by [16](a) Authoring mode where author inputs virtual instruction (b) Viewing mode where user is asked to select procure of two of the tasks [17]

Text Readability in AR/MR:

M. Fiorentino *et. al.* [18] performed a user study to find a suitable scheme of text and background for showing instruction in see-through HMDs. They used "Serif Helvetica" font as it is used in most readability studies. Four styles text only, text with two points wide line, text on rectangular billboard and text with outline and billboard were examined. They used four colours black, green, red and white in different combinations. According to their results, they found both presentation mode and background effects the text readability. For effective text, visualisation uses black text and white billboard or just white text only. Use coloured billboard with black text when colours have a specific meaning. Best performance found by billboards but at the cost of scene occlusion.

Fiorentino also suggested a high contrast scheme when reading time is important and the use of colours to reduce errors.

R. Razayev *et. al.* [19] investigated the text position and presentation while sitting and walking in smart glasses by comparing Rapid Serial Visual Presentation (RSVP) and line-by-line scrolling. They supported their investigation with quantitative aim and qualitative feedback, and subjective data. They discovered that presenting text in the top right corner for smart glasses results in lower text comprehension and greater workload compared to the middle and the bottom centre positions for both walking and sitting. RSVP leads to greater comprehension while sitting. Reading with scrolling leads to greater comprehension while walking. One can use these design suggestions for applying reading based applications for smart glasses.

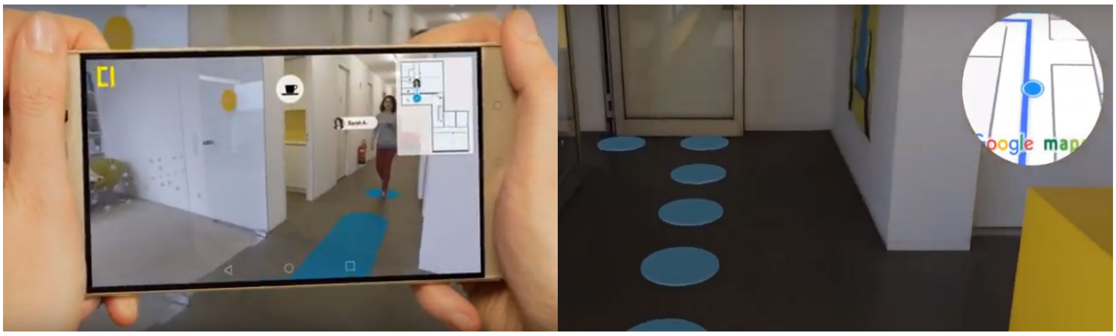
2.2 Navigation using AR/MR

AR can be used for navigation for different transport types like cars, flights or pedestrians. There are few guidelines for navigation by Jakub [15]. Design navigation interface based on transport type. For example, showing speed is important if travelling through the car and not of much important while walking. Make navigation understandable for the user by providing clear-cut instruction to navigate from one place to another. Visual or combination of visual and audio navigation is used in most cases as compared to audio-based navigation. Highlighting real objects can make navigation more understandable. Figure 2.2 shows an AR-based highlighted runway through a glass to pilot before landing.

Figure 2.2: Highlighted runway before landing [20]

Indoor Navigation

Indoor navigation is an important aspect related to this thesis work. Cologne-Intelligence [21] developed INPOSITION and INPLACES applications for indoor navigation using ARKit. INPOSITION helps user navigating to the known location of another user. INPLACES shows navigation path when a location is selected on the map provided within the application interface. Figure 2.3a shows an application user finding the location of another user whereas Figure 2.3b shows the navigation interface where the user is navigating to the desired location.



(A) INPOSITION Interface (B) INPLACES Interface

FIGURE 2.3: Applications by [21](a) User navigating to the position of another user (b) User navigating to the desired location

The tracking of HMD like Hololens is not good enough and stable to perform smooth navigation in larger workplaces. STIMULAB [22] created a Hololens application for navigation in the office area. They first scan the area and place the network of objects using unity3D objects. They also label those objects. Once setup completed, the user speaks a command to navigate to some destination then animated particles show the path towards the destination.

2.3 A Human Centeric approach to autonomous robot failures

Daniel J. Brook [5] worked on human-centric approach for autonomous robot failures. This work focuses on robot failures, human reactions to robot failures, types of recoveries and human help for failure recovery. This work uses a smartphone-based interface for human-robot interactions. The system detects robot failures and sends the notification to nearby human. The visual presentation includes icons and text messages. Several user surveys were performed to finalize the icons. There are two types of notifications; push-style notification where the human user gets the pop-up notification on the smartphone,

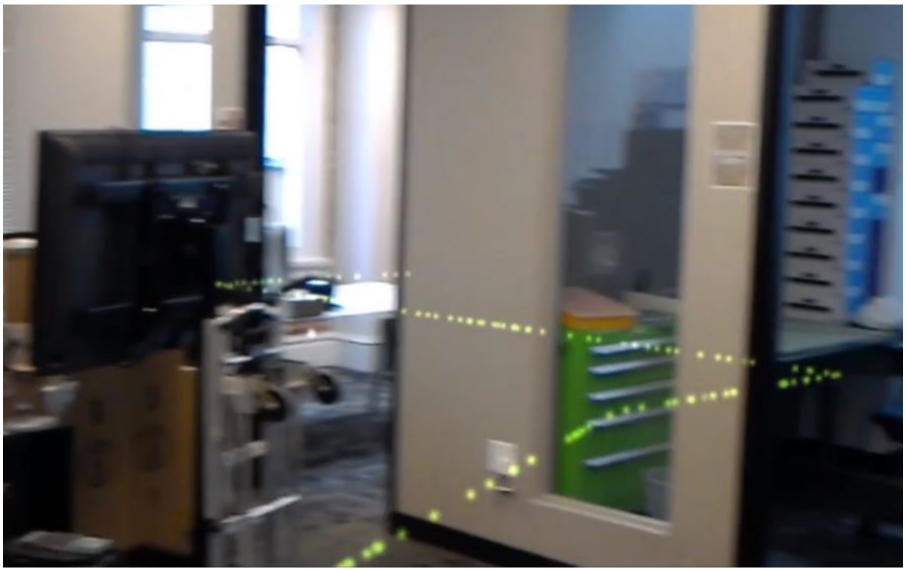


FIGURE 2.4: Animated particles showing highlighted path [22]

the second type is the pull-style notification where the human user opens the application and navigate in application to check robot's status. The system was evaluated by user studies. The human user was performing secondary tasks like playing mobile games and receives a notification from the vacuum cleaning robot.

FIGURE 2.5: Icons resulted in Surveys: OK, HELP, OFF, SAFE, DANGEROUS [5]

After human opens the notification, the robot explains its problem and the human approach to resolve that problem. Human communicates with the robot and sends feedback after resolving the issue. This work has some limitations. It is using the robot's names for identification. In the case of multiple same co-located robots, it will be difficult for the user to identify the failing robot. Our work guides the user and shows the exact robot that needs human help. Overlaying visuals on real objects make identification and awareness easier. Brook [3] is using mobile radio signals for communication, hence cannot perform distant communication. We are using some networking library for network communication so that human do not need to be present near the robot.

2.4 Asking for help using Inverse Semantics

Knepper [23] used inverse semantics for helping robots when they fail. They use "Generalized Grounding Graph" framework in the algorithm to generate verbal help request.

FIGURE 2.6: Starting from left; Push-Style notification, Responding to robot as "YES", Responding as "LATER",Follow up Question [5]

The probability-based algorithm checks similarity in natural language and the robot's desired actions for natural language generation. The system detects robot failure by comparing an actual and expected state of the robot, then calculates all potential steps towards failure resolving. The solution in symbolic representation is further used to find the most suitable words from natural language. This work presents a use case of robot assembling IKEA furniture. The robot's task is attaching legs with the table top. In the case of leg misplaced for the robot, the system generates an audio help request in natural language. As a downside of this work human has to be present near the robot whereas in our work autonomous agent has the ability of distant communication. This work also lacks multimodality because of speech only interface. We are developing a multimodal presentation to make human understand the situation better.

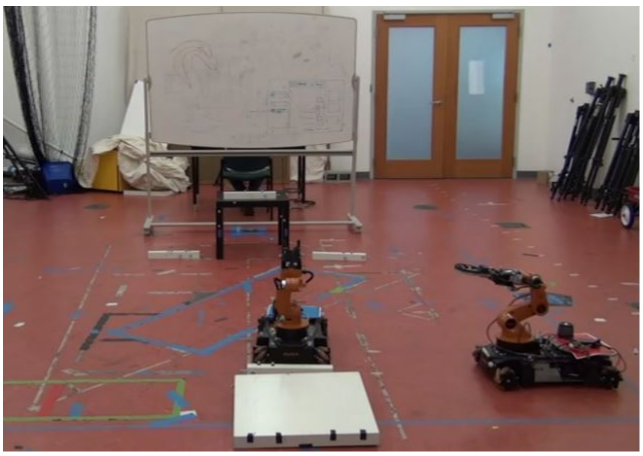


FIGURE 2.7: Configuration for evaluation study , User is behind the white board [23]

2.5 Resolving automated system failures in bin-picking tasks using assistance from remote human operator

Krishnanand [24] used human abilities to resolve automatic robotic perception failures. The work is about automation failure of small-scale robots. They develop the system for Baxter robot and human operator for bin picking task. Robot recognises 3D objects by measuring the similarity of saved 3D CAD and real-time point cloud data through ESCENO 3D camera. Iterative Closest Point (ICP) algorithm finds the similarity. Sometimes the object cannot be recognised because of ambiguity. The system takes a raw picture and sends to the human at a remote location. The human agent detects the object manually and informs the robot. This work is not generalised for different failure scenarios whereas our work is generalised about different failure scenarios. This work depends on human perception, and it does not handle for human perception failure situations. Human has the only raw image without annotations to understand the situation. We will show 3D graphics overlaying the real objects, highlighting the problem and providing better situational awareness.

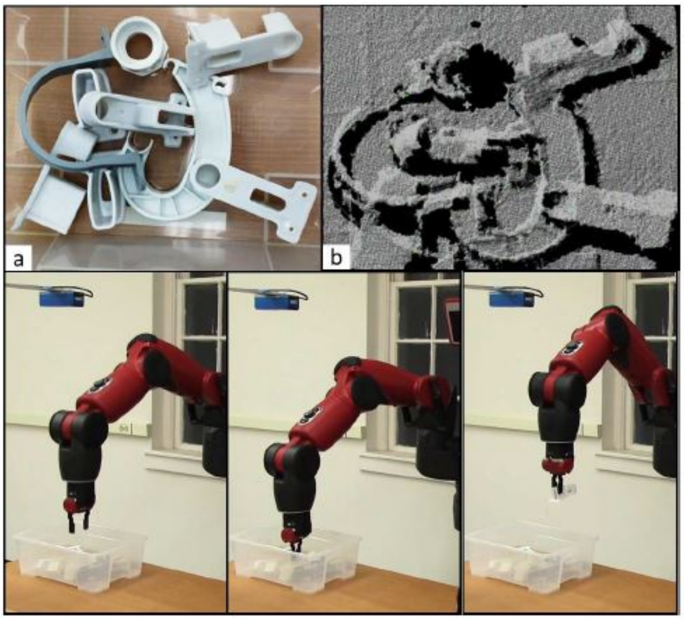


Figure 2.8: a- Mixed Bin Scene b- Failure of perception by automatic detection c- Robot pick up the target part based on info by remote human operator [24]

2.6 Autonomous driving: Investigating the feasibility of car-driver handover assistant

Marcel Welch [17] explores the transfer of control in autonomously driving cars when the car faces unresolvable situations. They developed a car driving simulation. Car identifies fog on the road and sends the control transfer request to the human driver. The Take Over Request (TOR) time is the time before reaching fog and selected as 4 and 6 seconds. The system notifies human along with instructions to take control at 4 and 6 seconds before failure. Results showed that the average response time of the human driver is three seconds. On the downside, they have explored just one use case with fog. This work is not generalised for different fail situations but, our work will deal with different failures. Also, there are limited predefined graphics for car driving scenario because of limited human actions for transfer of control during car driving. We are not using just pre-defined graphics; they are dynamic based on their position and generation.

(A) Alert dialogue with reason of transfer

(B) take over request with instruction

FIGURE 2.9: Showing Caution before fog in (a) Alert dialogue along with reason of control transfer (b) Request to take over control along with instructions [17]

2.7 Takeover times in highly automated driving

There is another research for finding the best time to transfer of control in automated vehicles by Mathias Kuehn [25] from German Insurance Corporation. This work explores the situations when car autonomy fails, and human should get control. They analysed five different scenarios for control transfer by user studies. SILAB software is used to create simulations keeping the capabilities and limitations of automatic cars on German roads. Audible warnings along with icons provide situation awareness. The results showed that 90percent drivers were able to take control and switch off automation in about 8 seconds. This work also lacks generalisation for different fail situations and less situation information because of pre-defined graphics. Our system is focusing on

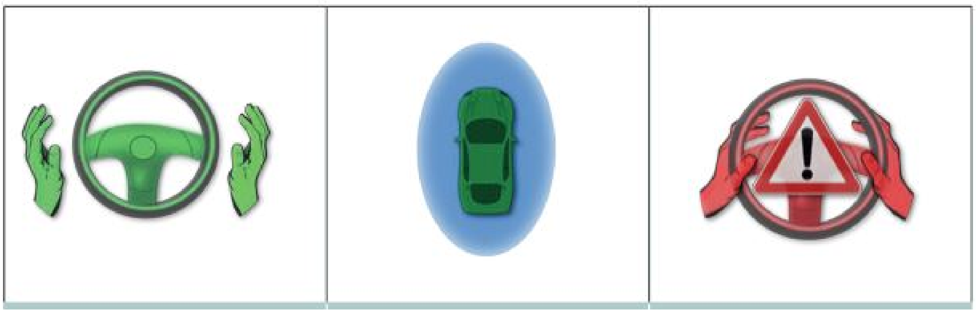


FIGURE 2.10: First symbol showing system can drive autonomously; second symbol shows system is monitoring; third symbol shows system identified unhandeled situation [25]

better and more intuitional situation awareness and problem-solving with the use of Augmented Reality.

2.8 Comparison of Related Works

Figure 2.11 shows a comparison of different related works. The presented works solve autonomy failures with human help. Almost every related work in tabular implements the notification system to notify the human agent on failure. Situation awareness is a significant element to resolve failures. To compare the quality of situation awareness in different works, we mainly consider three factors that include visualisation, audio and dynamic generation of content. [23] does not have the implementation of graphics or visualisation for explaining the situation to a human agent. [5], [23], [17] and [25], has visual information for the human agent but they all show graphics on display devices. In our work, we are laying graphics over the physical object in the real world. The motivation of bringing visualisation in the real environment is the quick and better understanding of a situation. [23], [17] implements audio help for situation understanding. We are also using audio cues and audible speech. The audios in our work are spatial audios instead of simple 2D audios. Just [23] from all presented related works is generating content dynamically by using inverse semantics while other presented related works do not have dynamic content generation. Our presentation generation is dynamic and generalised. It adapts according to the predefined commands.

	Notification System	Situation Awareness			Problem Solution
		Visualization	Audio	Dynamic Generation	
Human Centric approach to autonomous robot failures [20]	Yes, pull-style and push-style notifications to nearby humans	Yes, Icons Text	No	No	Human accepts request and resolve robot failure
Asking for help using Inverse Semantics [38]	Yes, Audio Message to co-located human worker	No	Yes, From Robot	Yes, using inverse semantic to generate speech	Human solves robot issue after verbal communication
Resolving automated system failures in bin-picking tasks using assistance from remote human operator [39]	Yes communication initiated with remote human	Yes, 3D Raw Image transfer to remote human	No	No	Human solves ambiguity of object selection using natural perception
Autonomous driving: Investigating the feasibility of car-driver handover assistant [32]	Yes	Yes, Icons/Text	Yes, cues only no speech	No	Human take over car control after take over request
Takeover times in highly automated driving [40]	Yes	Yes, Icons	No	No	Human takes car driving control
Transfer of control using Mixed Reality	Yes	Yes	Yes, Spatial 3D Audio	Yes	Human solves failure with real time presentation of multi-modal system

FIGURE 2.11: Comparison Table of different related works with the this thesis work

List of Figures

Figure 1.1 Simple architecture of autonomous agent [1] 1
Figure 1.2 Reality-Virtuality Continuum [7] 5
Figure 1.3 User playing video game in Virtual Reality(VR) [8] 6
Figure 1.4 Example of Pokemon-Go, AR Game 6
Figure 1.5 Venn Diagram of Reality-Virtuality continuum [7] 7
Figure 1.6 An example of a spatial mapping mesh covering a room 9
Figure 1.7 From Left; Magic Leap, Meta, and Microsoft Hololens [11] 11
Figure 1.8 Concept Map of AR interface design key elements 13

Figure 2.1 Prototype System by [16](a) Authoring mode where author inputs virtual instruction (b) Viewing mode where user is asked to select procure of two of the tasks [17] 16
Figure 2.2 Highlighted runway before landing [20] 17
Figure 2.3 Applications by [21](a) User navigating to the position of another user (b) User navigating to the desired location 18
Figure 2.4 Animated particles showing highlighted path [22] 19
Figure 2.5 Icons resulted in Surveys: OK, HELP, OFF, SAFE, DANGEROUS [5] 19
Figure 2.6 Starting from left; Push-Style notification, Responding to robot as "YES", Responding as "LATER",Follow up Question [5] 20
Figure 2.7 Configuration for evaluation study , User is behind the white board [23] 20
Figure 2.8 a- Mixed Bin Scene b- Failure of perception by automatic detection c-Robot pick up the target part based on info by remote human operator [24] 21
Figure 2.9 Showing Caution before fog in (a) Alert dialogue along with reason of control transfer (b) Request to take over control along with instructions [17] 22
Figure 2.10 First symbol showing system can drive autonomously; second symbol shows system is monitoring; third symbol shows system identified unhandeled situation [25] 23

Figure 2.11 Comparison Table of different related works with the this thesis work 24

Bibliography

[1] Paul Davidsson. Autonomous agents and the concept of concepts. 1996.

[2] Elena Dubrova. *Fault-Tolerant Design*, page 9. Springer Publishing Company, Incorporated, 2013. ISBN 1461421128, 9781461421122.

[3] J-C. Laprie. DEPENDABLE COMPUTING AND FAULT TOLERANCE : CONCEPTS AND TERMINOLOGY. In *Twenty-Fifth International Symposium on Fault-Tolerant Computing, 1995, ' Highlights from Twenty-Five Years'.*, Pasadena, CA, 1995. IEEE. ISBN 978-0-8186-7150-0. doi: 10.1109/FTCSH.1995.532603. URL `http://ieeexplore.ieee.org/document/532603/`.

[4] J. Carlson and R. R. Murphy. How ugvs physically fail in the field. *Trans. Rob.*, 21(3):423–437, June 2005. ISSN 1552-3098. doi: 10.1109/TRO.2004.838027. URL `http://dx.doi.org/10.1109/TRO.2004.838027`.

[5] Daniel J. Brooks. A human-centric approach to autonomous robot failures. Apr 2017.

[6] Gerald Steinbauer. A Survey about Faults of Robots Used in RoboCup. In Xiaoping Chen, Peter Stone, Luis Enrique Sucar, and Tijn van der Zant, editors, *RoboCup 2012: Robot Soccer World Cup XVI*, pages 344–355, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-39250-4.

[7] Jack Liu. The difference between ar, vr, mr, xr and how to tell them apart. URL `https://hackernoon.com/the-difference-between-ar-vr-mr-xr-and-how-to-tell-them-apart-45d76e7fd50`.

[8] Chloe Herring. A virtual reality makeover for game night? yes, at this party in the a+e district. URL `https://www.miami.com/things-to-do-in-miami/a-virtual-reality-makeover-for-game-night-yes-at-this-party-in-the-ae-district-1`

[9] Magic Leap. URL `https://www.magicleap.com/`.

[10] Meta Vision. URL `https://www.metavision.com/`.

[11] Constantine. Meta2. URL `https://vron.news/meta-2/`.

[12] Microsoft Hololens. URL `https://www.microsoft.com/en-CY/hololens`.

[13] D Andreas, nser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. Applying hci principles to ar systems design. 2007. URL https://ir.canterbury.ac.nz/handle/10092/2340.

[14] Subhashini Ganapathy. Design guidelines for mobile augmented reality : User experience. 2012. URL https://www.semanticscholar.org/paper/Design-Guidelines-for-Mobile-Augmented-Reality-%3A-Ganapathy/58fc08f19cb3a382d855b70db1b7907e15f3c818.

[15] Jakub Blokša. 2017.

[16] Jarkko Polvi, Juhyun Kim, Takafumi Taketomi, Goshiro Yamamoto, Jun Miyazaki, and Hirokazu Kato. User interface design of a slam-based handheld augmented reality work support system. 2014. URL https://www.semanticscholar.org/paper/User-Interface-Design-of-a-SLAM-based-Handheld-Work-Polvi-Kim/6f3329ba653310b328580e5a5f415e2c4fca19b6.

[17] Marcel Walch, Kristin Lange, Martin Baumann, and Michael Weber. Autonomous driving: Investigating the feasibility of car-driver handover assistance. pages 11–18, 2015. doi: 10.1145/2799250.2799268. URL http://doi.acm.org/10.1145/2799250.2799268.

[18] Michele Fiorentino, Saverio Debernardis, Antonio Uva, and Giuseppe Monno. Augmented reality text style readability with see-through head-mounted displays in industrial context. *Presence: Teleoperators and Virtual Environments*, 22:171–190, 08 2013. doi: 10.1162/PRES_a_00146. URL https://www.researchgate.net/publication/262319591_Augmented_Reality_Text_Style_Readability_with_See-Through_Head-Mounted_Displays_in_Industrial_Context.

[19] Rufat Rzayev, Paweł W. Wozniak, Tilman Dingler, and Niels Henze. Reading on smart glasses: The effect of text position, presentation type and walking. pages 1–9, 04 2018. doi: 10.1145/3173574.3173619. URL https://www.researchgate.net/publication/324665702_Reading_on_Smart_Glasses_The_Effect_of_Text_Position_Presentation_Type_and_Walking.

[20] Ghost Protocol. Aero glass - future of aviation piloting (hd). URL https://www.youtube.com/watch?v=TMRNMLF1fP0.

[21] TM Robotics. A philosophy of automation. URL http://tmrobotics.com/case-studies/a-philosophy-of-automation/.

[22] STIMULANT. Hololens prototypes from the stimulab. URL https://stimulant.com/hololens-stimulab/.

[23] Stefanie Tellex, Ross Knepper, Adrian Li, Daniela Rus, and Nicholas Roy. Asking for help using inverse semantics. July 2014. doi: 10.15607/RSS.2014.X.024.

[24] K. N. Kaipa, S. S. Thevendria-Karthic, S. Shriyam, A. M. Kabir, J. D. Langsfeld, and S. K. Gupta. Resolving automated perception system failures in bin-picking tasks using assistance from remote human operators. pages 1453–1458, Aug 2015. ISSN 2161-8070. doi: 10.1109/CoASE.2015.7294304.

[25] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer. Human-robot interaction in handing-over tasks. pages 107–112, Aug 2008. ISSN 1944-9445. doi: 10.1109/ROMAN.2008.4600651.

Appendix A

Study Questionnaires

Familiarity with Devices: **Pseudonym: HU101**

Please choose the correct numeric value based on your agreement with each statement.
Numeric values scale is as follows:

1 = Strongly Agree **2 = Agree** **3 = Slightly Agree**
4 = Neutral
5 = Slightly Disagree **6 = Disagree** **7 = Strongly Disagree**

Statements	Strongly Agree	Agree	Slightly Agree	Neutral	Slightly Disagree	Disagree	Strongly Disagree
I like robots.	1	2	3	4	5	6	7
I have knowledge about robot/s.	1	2	3	4	5	6	7
I have practical experience of working with robots.	1	2	3	4	5	6	7
I know about head mounted devices (HMDs).	1	2	3	4	5	6	7
I have experience of using head mounted device.	1	2	3	4	5	6	7
I have used Microsoft Hololens before.	1	2	3	4	5	6	7
I know interaction methods for Microsoft Hololens.	1	2	3	4	5	6	7
I know about Extended Reality technologies like AR,VR, MR etc.	1	2	3	4	5	6	7
I know how to use tablets.	1	2	3	4	5	6	7
I am a proficient tablet user.	1	2	3	4	5	6	7

Please make your evaluation now.

For the assessment of the product, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

attractive	○	⊗	○	○	○	○	○	unattractive

This response would mean that you rate the application as more attractive than unattractive.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Please assess the product now by ticking one circle per line.

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

Survey about application on Microsoft Hololens: **Pseudonym: HU101**

Consider both tasks you performed using Microsoft Hololens and rate the statements given below. Please choose the correct numeric value based on your agreement with each statement. Numeric values scale is as follows:

1 = Strongly Agree **2 = Agree** **3 = Slightly Agree**
4 = Neutral
5 = Slightly Disagree **6 = Disagree** **7 = Strongly Disagree**

Statements	**Strongly Agree**	**Agree**	**Slightly Agree**	**Neutral**	**Slightly Disagree**	**Disagree**	**Strongly Disagree**
Overall, I like the application to perform tasks	1	2	3	4	5	6	7
I recommend this Microsoft Hololens application	1	2	3	4	5	6	7
I read the text shown in applications	1	2	3	4	5	6	7
I had no problems while reading the text	1	2	3	4	5	6	7
I like the audio feature for instructions narration	1	2	3	4	5	6	7
I was able to understand the audio speech for each instruction	1	2	3	4	5	6	7
The audio speech was not annoying for me	1	2	3	4	5	6	7
I like the guidance path on the floor	1	2	3	4	5	6	7
I knew the exact location to stop near the robot	1	2	3	4	5	6	7
The path on the real floor was helpful to move around in the workspace	1	2	3	4	5	6	7
I felt safe while performing the task	1	2	3	4	5	6	7
I like the safety wall near the robot	1	2	3	4	5	6	7
I exactly knew when it is safe to interact with the robot	1	2	3	4	5	6	7
I was aware of my safe zone near the robot	1	2	3	4	5	6	7
I knew all the time if I have control or not	1	2	3	4	5	6	7
The control indicator on the top right corner was helpful	1	2	3	4	5	6	7
I learned the important components of the robot	1	2	3	4	5	6	7
The highlighted box around each part was helpful to understand robot components	1	2	3	4	5	6	7
The blue arrow to guide eye gaze towards each part was helpful	1	2	3	4	5	6	7
I felt sound cue coming directly from the highlighted boxes of robot components	1	2	3	4	5	6	7
The sound cues were helpful to find the components of the robot	1	2	3	4	5	6	7
I understood the exact problem of robot	1	2	3	4	5	6	7
I was able to solve the problem of robot	1	2	3	4	5	6	7
The animated arrow helped me to connect the cable correctly in the charging socket	1	2	3	4	5	6	7

Survey about application on Android Tablet: **Pseudonym: HU101**

Consider both tasks you performed using Android Tablet and rate the statements given below. Please choose the correct numeric value based on your agreement with each statement.

Numeric values scale is as follows:

1 = Strongly Agree **2 = Agree** **3 = Slightly Agree**
4 = Neutral
5 = Slightly Disagree **6 = Disagree** **7 = Strongly Disagree**

Statements	Strongly Agree	Agree	Slightly Agree	Neutral	Slightly Disagree	Disagree	Strongly Disagree
Overall, I like the application to perform tasks	1	2	3	4	5	6	7
I recommend this Android Tablet application	1	2	3	4	5	6	7
I read the text shown in applications	1	2	3	4	5	6	7
I had no problems while reading the text	1	2	3	4	5	6	7
I like the audio feature for instructions narration	1	2	3	4	5	6	7
I was able to understand the audio speech for each instruction	1	2	3	4	5	6	7
The audio speech was not annoying for me	1	2	3	4	5	6	7
I like the map showing my location and destination	1	2	3	4	5	6	7
I knew the exact location to stop near the robot	1	2	3	4	5	6	7
The arrow line on map was helpful to move around in the workspace	1	2	3	4	5	6	7
I felt safe while performing the task	1	2	3	4	5	6	7
I exactly knew when it is safe to interact with the robot	1	2	3	4	5	6	7
I was aware of my safe zone near the robot	1	2	3	4	5	6	7
I knew all the time if I have control or not	1	2	3	4	5	6	7
The control indicator on the top right corner was helpful	1	2	3	4	5	6	7
I learned the important components of the robot	1	2	3	4	5	6	7
I understood all components of robot by looking at the labelled picture	1	2	3	4	5	6	7
I feel easy to find components on real robot	1	2	3	4	5	6	7
I understood the exact problem of robot	1	2	3	4	5	6	7
I was able to solve the problem of robot	1	2	3	4	5	6	7
I exactly knew where to put the cable in the charging socket	1	2	3	4	5	6	7
I knew how to put the cable in the charging socket	1	2	3	4	5	6	7

Final Thoughts **Pseudonym: HU10**

Q- What do you prefer Hololens or Tablet for the robot failure recovery process? Any Reasons?

Q- Any valuable suggestions or feedback?

Thank you very much for participating in this study ☺