NAZARBAYEV
UNIVERSITY

Industrial Safety using Augmented Reality and Artificial Intelligence

by

Tolegen Akhmetov

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Robotics Engineering

Date of Completion
December, 2023

Industrial Safety using Augmented Reality and Artificial Intelligence

by
Tolegen Akhmetov

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Robotics Engineering

School of Engineering and Digital Sciences
School of Sciences and Humanities
Nazarbayev University

December, 2023

Supervised by
Prof. Huseyin Atakan Varol
Prof. Tohid Alizadeh
Prof. Erkan Kaplanoglu

Declaration

I, Tolegen Akhmetov, declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the author's original work. The thesis has not been previously submitted to this or any other university for a degree and does not incorporate any material already submitted for a degree.

Signature:

Date:

BLANK

Abstract

Industrialization brought benefits to the development of societies, albeit at the cost of the safety of industrial workers. Industrial operators were often severely injured or lost their lives during the working process. The causes can be cuts or lacerations resulting from moving machine parts, burns or scalds resulting from touch, or mishandling of thermal, electrical, and chemical objects. Fatigue, distraction, or inattention can exacerbate the risk of industrial accidents. The accidents can cause service downtime of manufacturing machinery, leading to lower productivity and significant financial losses. Therefore, regulations and safety measures were formulated and overseen by the government and local authorities.

Safety measures include effective training of workers, an inspection of the workplace, safety rules, safeguarding, and safety warning systems. For instance, safeguarding prevents contact with hazardous moving parts by isolating or stopping them, whereas a safety warning system detects accident risks and issues an alert warning. Warning systems were mostly mounted detection sensors and alerting systems. Mobile alerting devices can be gadgets such as phones, tablets, smartwatches, or smart glasses. Smart goggles can be utilized for industrial safety to protect, detect, and warn about potential risks.

Adopting new technologies such as augmented reality and artificial intelligence can enhance the safety of workers in the industry. Augmented reality systems developed for head-mounted displays can extend workers' perception of the environment. Artificial intelligence utilizing state-of-the-art sensors can improve industrial safety by making workers aware of potential hazards in the environment. For instance, thermal or infrared sensors can detect hot objects in the workplace. Built-in infrared sensors in smart glasses can detect the state of attention of users. Using smart glasses, potential hazards can be conveyed to industrial workers using various modalities, such as audial, visual, or tactile.

We have successfully developed advanced safety systems for industrial workers. Our innovative approach incorporates cutting-edge technologies such as eye tracking, spatial mapping, and thermal imaging. By utilizing eye tracking, we are able to identify instances of user inattention, while spatial mapping allows us to analyze the user's behavior and surroundings. Furthermore, the integration of thermal imaging enables us to detect hot objects within the user's field of view. The first system we developed is a warning system that harnesses the power of augmented reality and artificial intelligence. This system effectively issues alerts and presents holographic warnings to combat instances of inattention or distraction. By utilizing visual cues and immersive technology, we aim to proactively prevent accidents and promote worker safety.

The second safety system we designed involves the integration of a third-party thermal imaging system into smart glasses. Through this integration, our safety system overlays false-color holograms onto hot objects, enabling workers to easily identify and avoid potential hazards. To evaluate the effectiveness of our systems, we conducted comprehensive experiments with human participants. These experiments involved both qualitative and quantitative measurements, and we further conducted semi-structured interviews with the participants to gather their insights.

The results and subsequent discussions from our experiments have provided valuable insights for the future implementation of safety systems. Through this research, we envision the continued advancement and refinement of safety technologies to further enhance worker safety in industrial settings.

Acknowledgments

I would like to express my deepest gratitude to my wife, **Ayman**; my son, **Dimash**; my daughter, **Ayzere**; and my parents, **Kadyrbai** and **Altynai**, as well as my siblings, **Talgat** and **Aydana**. Their unwavering love and support have been the foundation of my journey. Your constant encouragement and understanding have provided me with the strength to pursue this doctoral degree. I am forever grateful for your presence in my life.

My friends and colleagues at the **Advanced Robotics and Mechatronics Systems (ARMS)** Laboratory and the **Institute of Smart Systems and Artificial Intelligence (ISSAI)** of Nazarbayev University, I sincerely thank you for being by my side throughout this challenging yet rewarding endeavor. Your companionship, laughter, and shared experiences have made this journey all the more memorable. Together, we have overcome countless obstacles, and I am truly fortunate to have you as lifelong companions. I would like to specifically mention and express my deepest appreciation to my colleague, **Zhanat Makhataeva**. Your dedication, insightful discussions, and collaborative spirit have made a significant impact on my research and personal growth.

I am indebted to my supervisor, **Prof. Huseyin Atakan Varol**, for his invaluable guidance, patience, and unwavering belief in my abilities. His expertise, dedication, and willingness to share knowledge have been instrumental in shaping my research. Despite the technical difficulties encountered along the way, his unwavering support has provided me with the strength to persevere and achieve significant results. Without his mentorship, this thesis would not have been possible.

I would also like to extend my heartfelt gratitude to my co-supervisors, **Tohid Alizadeh** and **Erkan Kaplanoglu**, for their invaluable contributions to my research endeavors. Their insights, guidance, and constructive feedback have greatly enriched the quality of my work. Their combined expertise and willingness to invest time in my progress have been pivotal in enhancing the depth and scope of my research.

Furthermore, I extend my gratitude to the **Nazarbayev University** for providing me with a conducive academic environment and access to exceptional resources. The opportunities for growth and learning that this institution has offered me are immeasurable. I am grateful for the guidance and support provided by the faculty members and staff who have played a significant role in shaping my academic journey.

During the course of my research, I faced numerous challenges, particularly in obtaining the desired results. It was through the combination of perseverance, experimentation, and collaboration that I was able to overcome these hurdles. The process of writing papers proved to be arduous, requiring immense dedication and countless revisions. However, the unwavering support and assistance of my supervisor made the journey more manageable. His constructive feedback and expert guidance helped refine my work and broaden my understanding.

As I near the completion of my Ph.D., I cannot help but reflect upon the multitude of experiences and memories that I have gained during this transformative period of my life. The late-night experiments, the exhilaration of breakthroughs, the camaraderie with fellow researchers, and the intellectual stimulation within the academic community will be cherished forever. I will undoubtedly miss the vibrant atmosphere and the constant pursuit of knowledge that has shaped my growth.

In closing, I offer my heartfelt appreciation to everyone who has contributed to my journey. Your support, belief in my abilities, and encouragement have been the driving force behind my success. This thesis stands as a testament to our collective efforts and serves as a reminder of the remarkable experiences I have had as a Ph.D. student.

Thank you all.

Contents

List of Tables

List of Figures

BLANK

Preface

I am pleased to present this research work that delves into the crucial realm of industrial safety and the role of emerging technologies in mitigating risks and protecting workers. Industrialization has brought various benefits to societies worldwide, but it has also come at the cost of the safety and well-being of industrial workers. Incidents involving severe injuries and even fatalities during the working process have underscored the urgent need for effective safety measures and regulations.

Throughout history, industrial accidents have resulted from a range of causes, including cuts, burns, and scalds arising from contact with moving machine parts or mishandling of thermal, electrical, or chemical objects. The risks are further exacerbated by factors such as fatigue, distraction, or inattention. Such accidents not only jeopardize the safety and lives of workers but also lead to service downtime of manufacturing machinery, resulting in decreased productivity and significant financial losses for industries. Consequently, governments and local authorities have instituted regulations and safety measures to ensure the well-being of industrial workers.

These safety measures encompass a comprehensive approach, including effective worker training, workplace inspections, implementation of safety rules, safeguarding mechanisms, and the utilization of safety warning systems. For example, safeguarding practices aim to isolate or stop hazardous moving parts to prevent contact-related accidents. Furthermore, safety warning systems play a vital role in detecting potential risks and promptly issuing alerts to mitigate accidents. Traditionally, warning systems have relied on mounted detection sensors and alerting mechanisms. However, the adoption of mobile alerting devices, such as smartphones, tablets, smartwatches, or smart glasses, has emerged as a cost-efficient and inclusive solution.

Smart glasses, with their array of advanced features, have become indispensable tools for industrial safety. By integrating augmented reality and artificial intelligence technologies, these glasses have the potential to revolutionize worker safety in the industry. Augmented reality systems developed for head-mounted displays extend workers' perceptual capabilities, enhancing their situational awareness in industrial environments. Additionally, state-of-the-art sensors integrated with artificial intelligence algorithms enable real-time hazard detection and warning, significantly minimizing the risks workers face. For instance, thermal or infrared sensors integrated into smart glasses can effectively identify hot objects in the workplace, providing timely alerts to prevent accidents.

The research presented in this work explores the integration of emerging technologies into smart glasses for industrial safety applications. By leveraging augmented reality and artificial intelligence, the aim is to empower industrial workers with enhanced perceptual abilities and proactive risk detection. Various modalities, including auditory, visual, and tactile cues, can be employed through smart glasses to convey potential hazards, ensuring workers are well-informed and capable of responding effectively.

I would like to express my sincere gratitude to my supervisor, co-supervisors, and committee members for their guidance, expertise, and invaluable support throughout this research endeavor. Their suggestions and feedback have played an important role in shaping the direction of this work. Additionally, I would like to thank my family and friends who have been a constant source of encouragement and understanding, support during the difficulties and successes of this challenging journey.

As I conclude this preface, I cannot hide my feeling of a sense of nostalgia and gratitude for the experiences and knowledge gained throughout my Ph.D. studies. This life-changing journey has given me key opportunities for personal and professional growth, and I will always adore the memories and relationships formed along the way.

I hope that this research work contributes to the ongoing efforts to enhance industrial safety and serves as a stepping stone for further exploration in this critical field.

NAZARBAYEV UNIVERSITY **Tolegen Akhmetov**
Astana, December 2023

Acronyms

ADAS Advanced drive assistance system. 59

AFRL United States Air Force Research Laboratory. 7

AI Artificial Intelligence. 1

AM Attention manager. 29

AP Average precision. 37

API Application Programming Interface. 11

AR Augmented Reality. 1

CHEX Change History Explicit. 29

CITI Collaborative Institutional Training Initiative. 4

CNC Computer numerical control. 33

CNN Convolutional Neural Network. 16

CPS Cyber-physical systems. 27

CV Computer vision. 16

DL Deep Learning. 15

FPV First-person view. 19

FSR Force-sensitive resistors. 42

GAN Generative Adversarial Networks. 16

GDP Gross Domestic Product. 3

GPU Graphics processing unit. 13

HFOV Horizontal field of view. 62

HMD Head-mounted display. 7

HRTF Head-related transfer function. 40

HTI Human Interface Technology. 7

HUD Head-up display. 29

IMU Inertial measurement units. 11

IoU Intersection over Union. 37

IPD Inter-pupillary distance. 39

IR Infrared. 11

IRA Interruption recovery assistant. 29

IREC Institutional Research Ethics Committee. 4

LED Light-emitting diode. 17

MARS Mixed and AR Studio. 11

ML Machine Learning. 13

MRTK Mixed Reality Toolkit. 11

NASA TLX NASA Task Load Index. 54

NLP Natural language processing. 16

NU Nazarbayev University. 4

OST Optical see-through. 13

PPE Personal protective equipment. 25

RPN Region Proposal Network. 16

RQ Research Question. 4

RSI Repetitive strain injury. 27

s-SURF Shrunken Speeded-Up Robust Features. 17

SIFT scale-invariant feature transform. 67

SLAM Simultaneous Localization and Mapping. 17

SPI Serial peripheral interface. 62

SURF Speeded-up robust features. 67

SUS System usability scale. 55

ToF Time-of-flight. 11

UAV Unmanned aerial vehicle. 29

WEA Wireless Emergency Alerts. 30

Chapter 1

Introduction

This thesis focuses on understanding practical safety issues in industrial processes and devising solutions to prevent them. The subsection 1.1 of the Introduction discusses the motivation of the study. The subsection mainly describes the types of industrial accidents that cause harm to industrial workers and presents the utilization of augmented reality (AR) and artificial intelligence (AI) as a solution to those mishaps. Following this, subsection 1.2 defines the research objective and questions. Finally, subsection 1.4 describes the overall structure of the thesis.

1.1 Motivation

As per the findings of the International Labour Organization, there are more than 340 million incidents of work-related accidents occurring globally on an annual basis [2]. More than 2.3 million people die from injuries and diseases in the workplace every year. In the European region, approximately 15% of these mishaps occur in the manufacturing industry [3] (see Fig. 1.1). The manufacturing industry has seen one of the highest rates of non-fatal incidents in the United States as well [4]. Types of industrial incidents include:

- Machinery incidents: Machinery with moving parts can cause injuries or deaths in cases of use without proper training or machine failure [5].

- Contact with harmful objects: Industrial workers can get injured or die from contact with thermal, electrical, or chemical objects.

- Transportation incidents: These take place during the transportation of people or goods with various modes such as vehicles, trains, aviation, and maritime transportation.

- Falls, slips, or trips: These accidents occur when workers lose footing or balance and fall. These kinds of accidents may result in serious injuries and damage to equipment.

- Fires and explosions: Malfunctioning equipment can cause industrial fires and explosions.

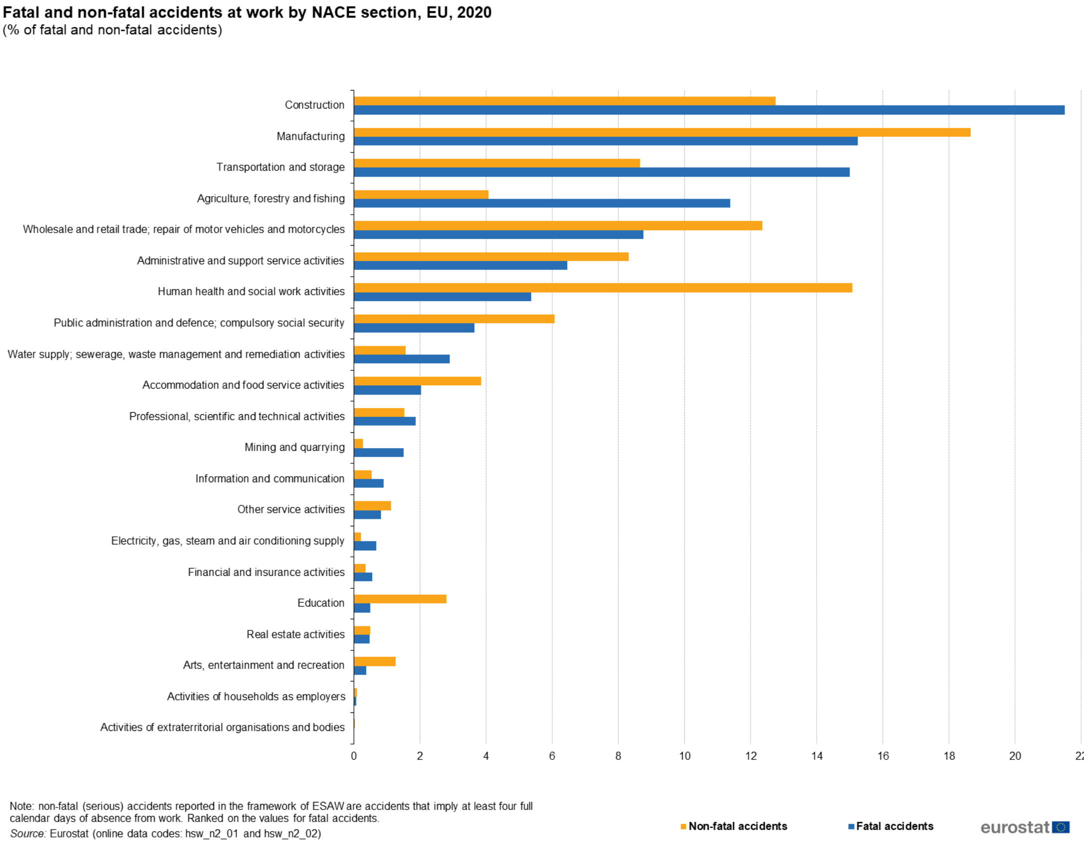


Figure 1.1: Statistics of fatal and non-fatal accidents in the industrial sectors of Europe. **Source:** Eurostat

Industrial mishaps may cause different injuries. For instance, contact with moving machinery can cause cuts or lacerations. Between 2005 and 2011, over one-fifth of fatal accidents in Poland occurred during machine operation [6]. Another type of incident is an accidental touch or mishandling of thermal, electrical, or chemical objects that can cause burns or scalds. Specifically, burns result from contact with hot and dry objects or fire, and scalds result from contact with wet hot objects or steam. From 2015 to 2018, approximately 23,000 people were injured by burns or scalds in workplaces in the United Kingdom [7]. According to the United States Bureau of Labor Statistics [8], 11,840 thermal and 3,540 chemical burns were sustained in the industry in 2020. The incidence of electrical burns continues to increase in many developing countries due to the growth of the industrial sector of their economies [9].

Industrial safety measures can prevent industrial accidents, financial losses, injuries, and deaths [10]. Effective training of workers, wearing personal protective equipment, and inspection of a workplace and equipment should not be neglected for the safety of workers. To enhance safety in the industry, we decided to investigate the potential of state-of-the-art technologies such as AR and AI. AI has already had successful implementations of fixed-position cameras in the industry. For instance, an AI-driven monitoring system was implemented to enhance safety in human-robot

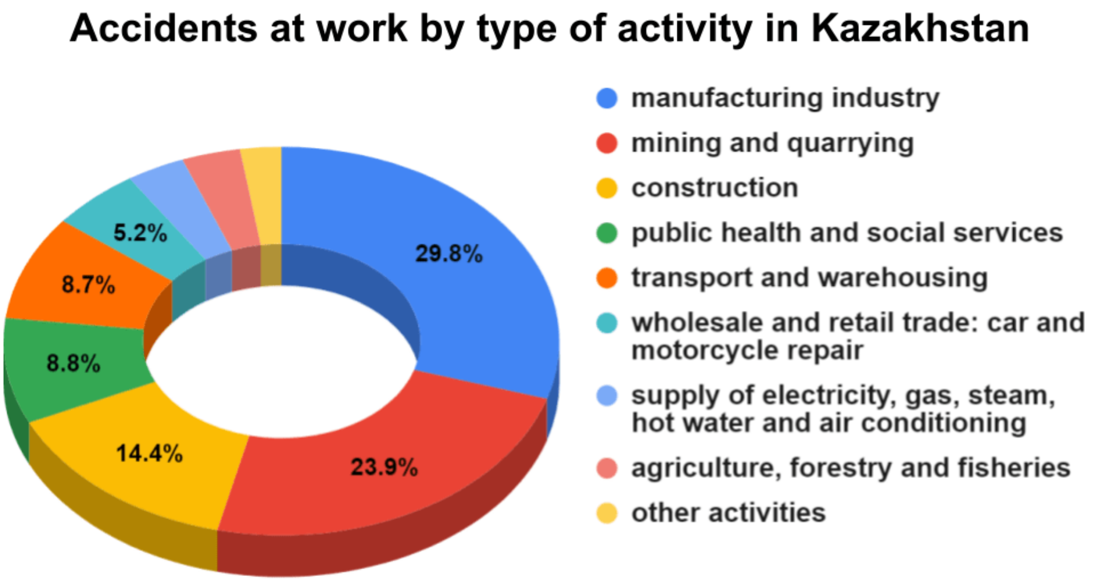


Figure 1.2: Statistics of accidents at work by type in Kazakhstan. **Data source: Bureau of National Statistics of Kazakhstan [1]**

collaboration [11]. The vision system utilized two Kinect V2 cameras to detect the risk of occlusion. However, conveying processed information about the industrial environment using first-person view devices can decrease costs and personalize the experience further.

Kazakhstan is a country with rich natural resources and significant industrial sectors such as oil and gas, mining, metallurgy, chemicals, and manufacturing. The monetary value of all goods and services produced within the borders over a specified period of time, typically a year, is called Gross Domestic Product (GDP) of the country. The services sector contributed the largest to Kazakhstan's GDP, accounting for approximately 53.92% of the economy. In comparison, the agricultural sector contributed around 5.03%, while the industry sector accounted for 35.27%. In contrast, a developed country like the United Kingdom had a lower share of agriculture contributing 0.67% and industry accounting for 17.49%. This highlights that Kazakhstan has a relatively higher proportion of its economy focused on the industrial sector, making industrial safety an even more important problem for the country.

Heavy industries in Kazakhstan cause safety hazards that lead to injuries, illnesses, or deaths. In 2022, there were 2.449 industrial accidents in Kazakhstan, which is 14.8% more than in 2021 according to the Bureau of National Statics of Kazakhstan [1]. Most of the injuries occurred in the manufacturing industry among the others (29.8%) (see Fig.1.2). The number of workers who got thermal burn injuries was 108. To prevent accidents in the industry, technologies such as robotics, the Internet of Things (IoT), drones, AR, and AI can be integrated into safety procedures in Kazakhstan. These tools might facilitate worker training, improve situational awareness, reduce human errors that might lead to injuries, and preemptively address safety risks.

1.2 Research Objective and Questions

The main objective of this thesis is to investigate the effectiveness of AR on industrial safety to prevent incidents and augment human perception to sense potential hazards. In our research, we utilized two approaches to enhance safety in the industry: eye tracking and thermal imaging. In the first approach, we developed an AR-based warning system to track the attention of a worker and warn in case of inattention. The system tracks the eye movements of the user while using the system. The system triggered warnings with different types of warnings such as auditory, visual, and audiovisual.

In the second approach, we integrated a third-party thermal imaging module into AR glasses for detecting hazardous hot objects in the field of view of the user. This safety system overlays false-color holograms onto hot objects.

Using these two approaches, we sought to answer a set of research questions (RQs) as summarized in Table 1.1. Firstly, we were interested in whether an AR-based warning system tracking eye movements could prevent distractions in an industrial environment (RQ1). We tested the warning system with different modalities on the personal perception of people in industrial tasks (RQ2). Safety systems with different approaches, such as safety mats and sawstop safety systems were implemented in the industry. We had to test whether AR-based warning systems outperform alternative systems productively used in the industry (RQ3). Another question is whether AR-based safety systems with thermal imaging can be utilized to increase safety in the industry (RQ4).

Table 1.1: Overview of the research questions addressed in this thesis.

No.	Research Question
RQ1	Can an AR warning system with eye tracking reduce human operator inattention?
RQ2	What is the subjective perception of the user experience for different warning modalities in industrial tasks?
RQ3	Do AR warning systems outperform business-as-usual systems such as sensorized safety mats in increasing the attention of operators?
RQ4	Can AR safety systems with thermal imaging increase safety in the industry?

1.3 Research Ethics

Before conducting experiments with participants, this study protocol went through ethical approval of the Institutional Research Ethics Committee (IREC) of Nazarbayev University (NU). I received training for conducting experiments with human participants from the Collaborative Institutional Training Initiative (CITI) under the requirements of the NU IREC committee. The confidentiality of participants was maintained, and all the subjects' names were paired with key codes. All the steps of our study were written in the application form and conducted within the scope of ethical norms.

1.4 Outline

The rest of the thesis is organized as follows:

Chapter 2 reviews literature and discusses the history of AR and AI technologies.

Chapter 3 demonstrates the technologies used in implementing the warning system for enhancing industrial safety.

Chapter 4 discusses experimental setup and participants.

Chapter 5 describes the objective and subjective assessments used in the study. Furthermore, the chapter reports and analyzes the quantitative results and qualitative results and discusses the results of the semi-structured interview and future perspectives in warning systems.

Chapter 6 presents the safety system utilized for detecting hot objects and conveying them in the visor of smart glasses. The chapter discusses the technologies leveraged for the system.

Chapter 7 provides with experimental setup and results conducted with the AR-based thermal perception system.

Chapter 8 concludes the thesis and presents future perspectives for industrial safety studies using AR.

Appendix A provides custom questionnaires utilized in the study.

Appendix B contains detailed results of objective and subjective measurements in the experiments.

Appendix C contains the code for implementing AR-based safety systems.

BLANK

Chapter 2

Background Research

2.1 A Brief History of AR

The history of AR can be traced back to the invention of the 'Sword of Damocles' by Ivan E. Sutherland in 1965 [12] (see Fig. 2.1). The head-mounted display (HMD) suspended from the ceiling enabled users to view 3D images while staying strapped to the device. However, the origins of both AR and VR can be traced even further back to Charles Wheatstone's invention of a stereoscope in 1838, which allowed people to view separate images for each eye as a combined three-dimensional image [13]. Unlike VR, which immerses users in a fully three-dimensional virtual world, AR combines the virtual world with the real physical world in 3D space [14].

The term 'Augmented Reality' was coined by Caudell and Mizell in 1990 [15]. Working as engineers for the aerospace company Boeing, they were tasked to create a practical system that would instruct wiring on large plywood boards. The authors proposed to superimpose the positions of wires on HMDs. Instead of looking back and forth between different instructions of Boeing's old instruction system, the device displayed instructions in a single visor. This saved time for workers and made easier the assembly process.

In 1992, Virtual Fixtures [16] platform was developed by Rosenberg at the United States Air Force Research Laboratory (AFRL). The platform allowed military staff to control and manipulate virtual holograms. The users donned sensors on full upper-body parts to interact with virtual 3D content. The system integrated haptic, auditory, and visual feedback. One of the first open-source software libraries for AR, ARToolkit was created by Kato et al. [17] in 1999 and was supported further by Human Interface Technology (HTI) Laboratory at the University of Washington. The toolkit allowed developers to collaborate peer-to-peer and advance the AR field. Wikitude [18], a location-based AR application for mobile phones was released in 2008. The user's position was estimated using GPS or Wi-Fi and orientation was extracted using an accelerometer. Pokemon Go [19] is one of the well-known location-based AR applications. Google Glass was released by Google in 2013 [20]. The pair of glasses allowed users to visualize information in the visor and capture photos and videos, and record audio. However, this system was expensive and had privacy issues. Specifically, Google Glass, with its audio and video recording capabilities, has raised concerns regarding potential unauthorized recording in public spaces, private conversations, and

Figure 2.1: Sword of Damocles, one of the early systems for AR by Ivan E. Sutherland.

sensitive areas. The inclusion of facial recognition technology has also raised alarms about the possibility of mass surveillance without individuals' consent. Additionally, the storage and utilization of the collected data have raised apprehensions regarding privacy and its potential misuse. Therefore, it was not accepted widely by users. Regardless, AR has evolved in the last three decades tremendously and it has now the potential to extend human perception and transform our interactions with the environment.

2.2 AR Display Technologies

As we reviewed the history of AR, the first HMDs were fixed to custom-designed rooms and were far from reaching the consumer market. The HMD technology progressed significantly over the decades. The adoption of the first HMDs in the 1990s was limited due to their cost, size, latency, battery life, computation power, field of view, and resolution. Technological advances in sensors, optics, and computing devices stimulated AR research. In particular, the miniaturization of the components helped to reduce the size and weight of HMDs. Multispectral and high-resolution cameras integrated into HMDs enabled the precise tracking of space and human behavior and the placing of holograms accurately. Holographic interactions and visualizations of users became more realistic and convincing. Besides advancements in wearable AR, the issues such as high price, short battery life, narrow field of view, and low resolution of visors exist in modern smart goggles. Smartphones and tablets are not as pricey as smart glasses, however, they cannot deliver an immersive experience. Projection-based AR is immersive and doesn't require any device to be worn or held, but it is not mobile like smart glasses and smartphones. The aforementioned limitations of the technology should be resolved to make the devices more user-friendly and deployable.

2.3 Devices and Tools for AR

As previously mentioned, AR can be experienced through various devices such as smart glasses, handheld devices, and projection-based systems (see Fig. 2.2). These technologies have led to the emergence of wearable, mobile, and spatial types of AR.

(a) Projection-based AR

(b) Wearable AR

(c) Mobile AR

Figure 2.2: Major display technologies for AR: a) Projection-based AR, b) Wearable AR, and c) Mobile AR

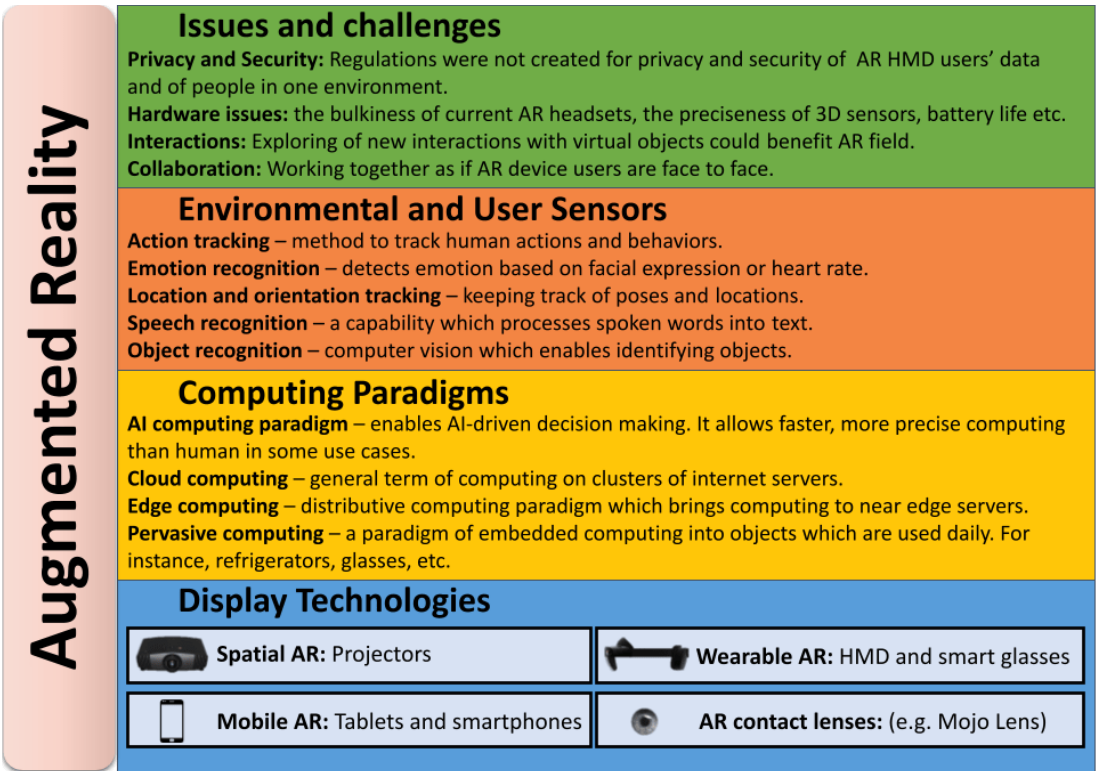


Figure 2.3: Different sensors, computing paradigms, software and display technologies for the state-of-the-art AR systems.

Currently, some of the most advanced AR-capable HMDs include Apple Vision Pro, Meta Quest Pro, and Microsoft Hololens 2. In addition, earlier HMDs like Microsoft Hololens 1, Magic Leap, Google Glass, Epson MOVERIO, and Vuzix Blade smart glasses have been utilized in AR research successfully.

Recent news in the technology industry highlights successful applications of AR in aerospace, transportation, telecommunication, and manufacturing. For example, NASA [21] has developed AR integrated with software for real-time drone tracking. They employed Microsoft HoloLens to visualize the spatial relationships between multiple flight paths and objects during flights. Additionally, Airbus collaborated with Microsoft to create an AR-based solution for various purposes in the aerospace industry, including training novice workers, remote collaboration, product design, and manufacturing. German company RE'FLEKT also developed an AR-powered system for repairing and maintaining complex systems in the manufacturing, telecommunication, and transportation sectors [22].

Modern AR applications utilize natural feature-based, model-based, and marker-based positioning. Marker-based tracking [23] utilizes markers in the image for positioning whereas natural features such as posters or road symbol signs are taken as references in natural feature-based positioning [24]. Model-based tracking [25] can recognize the position and orientation of the model referring to the sequence of images.

Various software tools were introduced in recent years for simplifying the design and use of AR, including the game development platforms such as Unity and Unreal Engine, Vuforia engine for marker-based, natural-feature and model-based tracking, Mixed Reality Toolkit (MRTK) which is integrable with both Unity and Unreal Engine, Stereokit library, and Mixed Reality capabilities of Robotic Operating System (ROS) (see Fig. 2.3). The game development engine Unity has received enhancements in the form of Mixed and AR Studio (MARS), which enable object and event tracking using 3D objects as markers. In a recent development, the Khronos Group has introduced OpenXR, a royalty-free standard for AR and VR devices and platforms. This standardization allows most browsers to support the WebXR Device Application Programming Interface (API) using OpenXR, enabling users to experience 3D AR content directly through their browsers.

For example, developers can utilize conventional JavaScript code and leverage libraries like Babylon.js, which relies on WebXR, to create immersive realities. Notably, major technology companies including Microsoft, Oculus, Samsung, Magic Leap, Apple, Unity Technologies, and Epic Games (the developer of the Unreal Engine) have updated their products to ensure compatibility with OpenXR. As a result, developers now have the capability to design cross-platform applications or websites and host online AR/VR environments. This advancement allows users to experience AR/VR content either through their browsers or through dedicated AR/VR headsets.

2.4 Sensors and Display Technologies for AR

Modern AR headsets have various sensors, visors, and cameras that enable to build of realistic immersive experiences. For instance, the spatial awareness capability of AR glasses utilizes built-in sensors and camera signals to build a 3D map of the real physical world. Frequently used camera technologies in AR are time-of-flight (ToF), depth, infrared (IR), and stereo cameras. Popular camera-based techniques utilized for augmenting virtual content into the real environment use fiducial markers (e.g., ArUco [26] [27] and Vuforia [28]). In order to track the movement of the device and define its orientation, inertial measurement units (IMUs) are used. As illustrated in the Figure 2.4, Microsoft HoloLens 2 is a noteworthy example of new-generation AR goggles that incorporate a wide range of sensors.

Sensors play a crucial role in environmental perception and data collection in AR devices. These devices, such as HoloLens 2 and Vuzix glasses, employ various sensors including visible light environment tracking cameras, depth cameras, IR-reflectivity streams, accelerometers, gyroscopes, and magnetometers. The depth camera has two modes: short-throw mode for hand tracking and long-throw mode for spatial mapping. IR sensors are used to calculate the depth of the environment. Interactions with AR systems can be established through hand gestures, voice commands, eye movements,

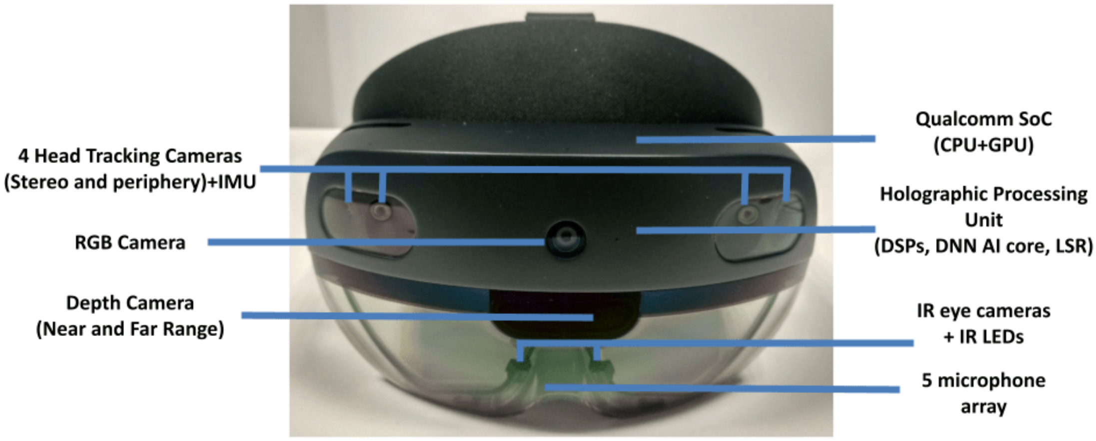


Figure 2.4: Microsoft HoloLens 2 smart glass with annotated sensors and computational elements.

or external controllers. While HoloLens 2 supports gesture and voice control, the first-generation HoloLens utilized an external controller called a clicker. Other AR glasses, like Smart Swim and Vuzix M-400, are designed for specific purposes and come with features like GPS, Bluetooth, and WiFi for tracking performance and connecting with other devices. These glasses also incorporate noise-canceling microphones for usability in noisy environments. Human behaviors such as eye movements and facial expressions can be utilized in AR glasses to detect human intention or attention. However, challenges such as weight, size, thermal properties, battery life, and display resolution still pose limitations in the field of wearable AR technology.

In addition to digital glasses, smartphones are also frequently utilized for AR. New generation of smartphones, tablets, and other mobile devices are equipped with advanced features such as spatial mapping and object recognition. For example, in [29], a smartphone was used for navigation, assisted by AR-enhanced automatic object recognition. Also, many modern smartphones and tablets are equipped with ToF depth cameras and LiDAR sensors, which are utilized for the creation of 3D-immersive environments. 3D scanners can be paired with mobile devices in order to develop an AR environment as well. Interestingly, mobile devices can be used to set up the WebAR environment, in which web browsers of mobile devices can be utilized to acquire and display AR features [30]. General trends indicate that mobile devices carry a significant potential for providing a cost-efficient AR experience in the 21st century.

In addition to wearable and mobile AR, there is also spatial AR environment, where projectors are utilized to superimpose or integrate virtual 3D objects onto a real physical environment. Within spatial AR, the concept "spatial user interface" (SUI) [31] lets users freely interact with both 3D virtual environment and traditional 2D workstation applications. Nonetheless, SAR has its unique constraints: 1) SAR can be projected onto surfaces and can’t be projected in mid-air; 2) Projector brightness limits the use in outdoor environments; 3) It is vital for SAR to have geometric surfaces that must be known a priori. The reader is referred to the book [32] for a detailed treatment of SAR. Such limitation for outdoor use is shared by the modern optical see-through (OST) HMDs [33] as well. Specifically, the contrast between the AR imagery of the HMDs and physical environment is lost in bright outdoor conditions.

2.5 Computing Paradigms for AR

Several computing paradigms have emerged as a result of widespread Internet connectivity and the shrinking of computing components (see Figure 2.3). The first paradigm, AI computing includes deterministic and non-deterministic algorithms that drive speech and object recognition technologies [34]. The second one, edge computing [35], is a distributed processing paradigm that brings the delivery of services, data, and intelligence closer to users and devices. Autonomous vehicles, intelligent street lighting, automated industrial machinery, smart houses, and mobile devices are well-known instances of edge computing. In the third paradigm, cloud computing [36], data is collected and processed in the "cloud," i.e., in a remote data center with extensive processing and storage capacities. The fourth paradigm, often known as ubiquitous or pervasive computing [37], integrates computation into commonplace things rather than performing computations on separate machines. Computers could appear as commonplace items like a refrigerator or a pair of glasses in ubiquitous computing.

AI paradigm has spawned new algorithms in recent years that expand the potential applications, including AR. AI is the science of creating intelligent agents that can take action based on information received from the environment [38]. In certain tasks including speech recognition, handwritten character identification, and spam classification, modern AI systems can compete with or even outperform humans. Machine Learning (ML), a subset of AI, analyzes data to identify patterns that are utilized for making certain decisions. ML explores the questions of how to create computer systems that improve by experience and what are the underlying principles that govern these learning systems [39].

ML techniques have been employed in numerous AR applications. By combining ML and AR, the researchers created the framework to help medical residents by improving their comprehension of the spatial relationships between tools, implants, and anatomical objects [40]. For object recognition, extracting their positions/orientations, and creating an augmented vision, the learning algorithm utilized depth and X-ray images. A visibility management technique for spatial AR, VisLP, was proposed by Ichihashi et al. [41], which put annotations and connection lines based on visibility estimation using ML. Advances in computer technology, particularly the advent of graphics processing units (GPUs), enabled the use of powerful deep learning (DL) algorithms for object recognition. Accurate and fast object recognition opens up a lot of possibilities for

AR. Sutanto et al. [42] presented a markerless AR method that uses a deep network for object recognition. Hoppenstedt et al. [43] developed an object recognition system using voice commands from AR glasses to label objects.

AR uses edge computing for complex computational processes. Particularly, DL and other computationally demanding processes are typically handled on the server side, whereas AR goggles are typically utilized as clients. For HoloLens, Microsoft created the Research Mode add-on to allow for smooth server connectivity. This mode allows devices to send information to an independent edge computing device that was collected by internal sensors. Mobile computers ("backpacks") with AR/VR technologies have recently entered the market. In order to create an adaptive projection AR system, researchers utilized these specialized mobile computers for DL-based object detection [44].

According to the principles of ubiquitous computing, the processing units of AR systems reduce latency by avoiding data transmission between edge and cloud computing devices. Microsoft HoloLens versions 1 and 2 offer a preview of the next generation of smart eyewear with ubiquitous computing. Holographic processing units (HPUs) in these gadgets make them standalone holographic computers. However, the Hololens 1's HPU needs about a minute to process just one frame of data using the cutting-edge object recognition engine "YOLO" [45]. As a result, the built-in holographic processing unit of an AR system may be unable to handle heavy workloads, which could lower quality-of-service.

2.6 Privacy Issues in AR

AR technology is becoming prevalent in various domains. However, the integration of this technology raises privacy concerns such as the collection and storage of personal data, sensitive biometric data, location, and human interactions in private and public places. Considering all these aspects before AR gets more popular is imperative for the protection of individual privacy.

Data Collection and Storage: AR glasses with cameras and sensors can capture and store visual, audio, and environmental data. This data may be personal information, private images, and sensitive details in a photo. Robust data collection, storage, and retention protocols should be established to address this concern. These protocols should ensure that data is secure and that user consent is obtained for any data processing or sharing activities.

Biometric Data and Facial Recognition: AR glasses with built-in cameras raise concerns about biometric data's potential capture and analysis, such as facial features and expressions. This data can be used for facial recognition or tracking, raising privacy and surveillance issues. Protocols should be established to control biometric data collection, use, and storage, ensuring compliance with relevant privacy laws and regulations.

Location Tracking and Geospatial Data: AR glasses often rely on location-based services and geospatial data to provide contextually relevant information. However, this raises concerns about location tracking and potentially disclosing sensitive information about users' whereabouts. Privacy protocols should include mechanisms for transparent location tracking practices, explicit user consent, and secure geospatial data handling.

User Interaction and Personalization: Hand recognition, eye tracking, and voice commands widely used for interacting with holographic content may cause privacy issues. The biometric nature of this data makes it sensitive and raises privacy concerns. Protocols should be established to ensure that user interaction is secure and utilized for intended purposes only.

Researchers have actively addressed the privacy and security concerns associated with the use of AR technology. Gobel et al. [46] have specifically focused on privacy concerns related to eye-tracking in AR glasses. The presence of cameras and microphones collecting data for tracking human behavior and commands can potentially infringe upon the privacy of the general public. To mitigate this, the authors propose a solution that captures only the areas where the user gazes, thereby avoiding the collection of irrelevant data that may compromise the privacy of other individuals. Additionally, they suggest activating the microphone only when the user intends to speak, and this intention can be detected through eye movements.

In [47], the authors explored the use of multiple sensors in AR glasses to enhance the security of personal data. They recommend implementing popup security consents that appear before using applications that access sensitive personal data. These consents serve as a protective measure, ensuring that users are aware of and provide explicit permission for the collection and usage of their personal information.

Industrial implementations, such as Microsoft's HoloLens, have also recognized the importance of privacy. Microsoft has incorporated privacy controls and features into their AR glasses, allowing users to manage their privacy settings and control data collection and sharing [48].

The risks associated with privacy in AR glasses can be mitigated by implementing privacy protocols, adhering to relevant regulations, and adopting privacy-conscious design principles. The academy and industry must prioritize privacy in all steps of implementing and consuming AR products.

2.7 Computer Vision in AR

Deep Learning (DL) is the subset of ML that uses multiple neural network layers to extract features from the raw input data. DL is capable to process large-scale data but requires substantial computational power. DL excels in terms of accuracy and performance in areas such as computer vision (CV), natural language processing (NLP), and speech recognition. However, interpreting DL's internal working and decision processes is challenging for the human mind and DL models are usually regarded as black box. DL can be supervised, unsupervised or semi-supervised.

Supervised learning trains its neural networks using labeled data. As its name implies, the system works similarly to the supervising of a teacher. One of the well-known supervised learning algorithms is Convolutional Neural Networks (CNNs). CNNs have convolutional layers with filters and the features are convolved through these layers. CNNs are widely used for AR [42]. Hoppenstedt et al. [43] labeled data voice commands in Hololens and implemented object detection using CNNs. There are state-of-the-art object recognition models such as YOLO, tinyYOLO that are used in AR [49, 50].

Runz et al. [51] introduced an object detection method using the MaskFusion model which is based on Mask-RCNN for object detection. RGB and RGB-depth cameras were used for capturing the environment in real-time. Mask-RCNN extends a small overhead over Faster RCNN [52]. Faster-RCNN firstly finds regions of interest. Next, Region Proposal Network (RPN) and Fast R-CNN are merged into a single network by sharing their convolutional features. Faster R-CNN with Single Shot Detector (SSD) and MobileNet V2 with SSD is used to detect objects using Epson Moverio M350 in small series production by Zidek et al. [53]. Two models, Faster R-CNN V2 and MobileNet V2, were trained on the newly generated 2D sample data from virtual 3D models by transfer learning technique. Transfer learning is a technique in ML that utilizes knowledge learned from one task for boosting the performance in a related task. MobileNet V2 with SSD is a model optimized for devices with low computational capabilities such as smartphones. Mixed memory CNN (mmCNN) is proposed for limited memory devices by Li et al. [54]. Nonetheless, supervised learning is computationally costly and time-consuming compared to unsupervised learning. To solve the lack of labeled data, researchers employed unsupervised generative models [55]. Synthesized data from 3D models and generative adversarial networks (GANs), an unsupervised learning algorithm, was used to generate image training samples to increase the amount of data. Since unsupervised learning does not use any class labels to train, less human labor is needed for preparing the data in unsupervised learning.

The AR registration framework proposed by Gao et al. showed robust results on object recognition [56]. Zubizarreta et al. [57] introduced the ARgitu algorithm for recognizing 3D objects. This is a 3D non-Lambertian object recognition algorithm that uses a chamfer-matching approach in arbitrary environments. The authors trained algorithms using 3D CAD models. The iterative Levenberg-Marquardt algorithm is used to optimize the pose of the object, and the Tukey loss function is used to remove outliers. Shrunken Speeded-Up Robust Features (s-SURF) are used on mobile devices to match objects. SURF is an algorithm that extracts features to recognize targets similar to SIFT. During the feature extraction process data is accompanied by the scale of the object whereas SIFT does not use scales [58].

2.8 Spatial Mapping in AR

Spatial mapping is a fundamental capability in modern AR glasses that allows the creation of a 3D map of the real-world environment. This feature enables accurate positioning of virtual objects and interactions with their physical surroundings. Several studies have explored and implemented spatial mapping in AR glasses, demonstrating its potential in various applications. Modern glasses such as Hololens and Meta Quest Pro utilize various sensors and algorithms to build spatial maps of the environment. Several environmental cameras, IMU sensors, and depth cameras can be utilized for spatial mapping. The simultaneous localization and mapping (SLAM) algorithm is widely used in modern smart glasses for building holographic maps of the physical environment [59]. Third-party depth sensors can be integrated to overcome the limitations in the spatial mapping of AR glasses [60]. In [61], a DL-based mobile AR system was implemented for the spatial mapping between holographic and physical objects utilizing snapshot-based RGB-D data. The system positioned the real object in a holographic spatial map without prior knowledge and segmented the object using the Mask R-CNN algorithm.

2.9 Eye Tracking in AR

Eye tracking, also named as gaze tracking, involves the measurement of eye orientation and position relative to the head. Several eye-tracking devices are available in the market: eye trackers integrated into smartglasses, separate eye trackers that can be donned to the head, or eye trackers fixed to a setup such as an interior panel of a car. Eye tracking technologies are used in psychology [62], marketing [63], and human-computer interaction [64]. For instance, social psychologists use eye-tracking to obtain information about human intention, arousal, and subjective weighting of motivation [62]. In visual marketing, eye tracking is used to study advertising, health and nutrition warnings, branding, and choice and shelf search behaviors [63]. Researchers also mentioned that eye tracking has the potential to become a central interaction type in gaming and computing [64].

Eye tracking provides vital information about the user's visual attention, intent, and focus. In the Microsoft HoloLens 2, two IR cameras are used together with IR light-emitting diodes (LEDs) for gaze tracking. In the software, data about the origin

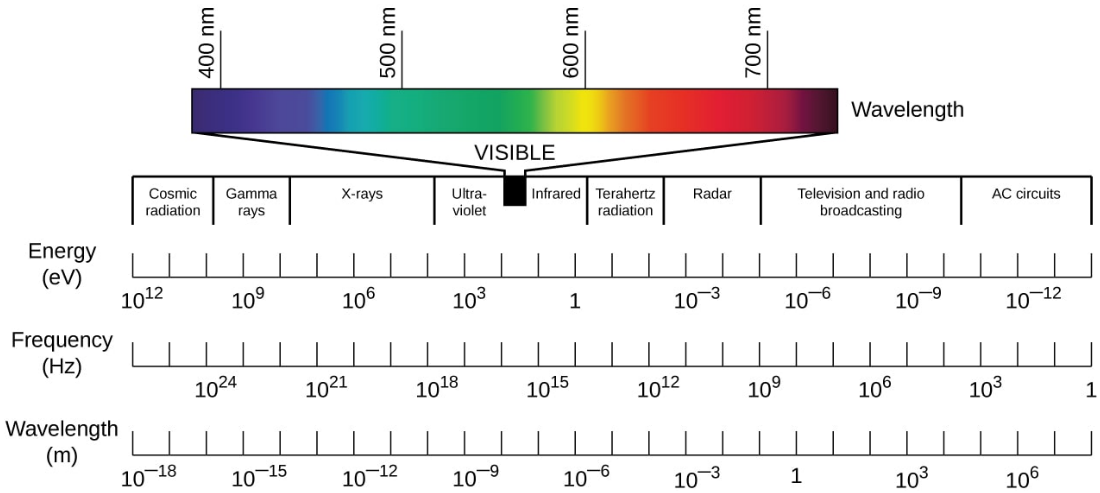


Figure 2.5: Electromagnetic spectrum showing energy, frequency, and wavelength regions by CFCF licensed under CC-BY-4.0

and direction of gaze can be extracted using the MRTK. The ARETT toolkit was developed for the Microsoft HoloLens 2 to enable seamless posthoc analysis of gaze data [65]. OpenXR by Khronos is a universal development platform for AR and VR devices. The gaze origin and orientation of the AR goggles can be extracted using an OpenXR plugin [66]. Third-party eye trackers can be integrated into AR devices for understanding the user's visual activities [67]. Pupil Labs and Tobii eye trackers are the most popular of them. Tobii eye trackers were used for comparing single-modal and multimodal interactions in AR.

2.10 Multispectral Vision in AR

The electromagnetic spectrum encompasses a wide range of wavelengths, including human visible light, IR, ultraviolet, microwaves, and radio waves. The human eye can see a small portion of the light which ranges from 400 to 700 nanometers (see Fig. 2.5). Different wavelengths within the visible spectrum are associated with different colors such as violet, blue, yellow, orange, and red. There are regions of the spectrum not visible to the human eye. For instance, IR has longer wavelengths than human visible light whereas shorter wavelengths than human visible light are in ultraviolet radiation. AR technology can help expand human perception using sensors that can perceive beyond what the human eye can see. This kind of simultaneous capturing and analyzing information from different regions of the electromagnetic spectrum is called multispectral vision [68].

Augmenting human perception using non-human sensing modalities can enhance human situational awareness and improve industrial safety. As an example of this trend, Erickson et al. [68] implemented an egocentric multispectral vision system that utilized AR goggles (Microsoft Hololens 1), ultraviolet, and IR cameras. Images obtained in the non-visible wavelengths were visualized on the AR goggles, which artificially extended the user's visible spectrum. In another study [69], the researchers extended the capabilities of AR goggles by adding two IR cameras. The authors simulated a temperature rise (hot) and fall (cold) in experiments with human participants. The simulation studies were conducted in a special isolation booth with a heater and a cooler placed on the table. The simulations affected the participants' temperature estimates of their bodies and environment. The Vismerge system [70] was implemented to fuse IR and RGB camera views in the user's field of view. The authors utilized the time-warping technique [71] to solve temporal latency issues and synchronize the video streams. The Oculus Rift HMD was used to convey visualizations to the participants.

IR light falls outside the range of human visible light, but it can be detected and captured by thermal cameras. A thermal camera is a non-contact device that captures IR radiation (heat) as a visible image representing the temperature distribution in space. There is a variety of thermal cameras on the market with different form factors (e.g., handheld (pistol-grip), monocular, phone-connected, and thermal imaging smartphones).

2.11 Industrial AR Applications

In the manufacturing sector, innovations directly impact productivity and profit. Technologies such as AR are indispensable in many use cases. AR can be used for maintenance, equipment identification, inspection, product design, real-time data visualization, training, quality control, remote assistance, and safety. AR can assist industrial workers in performing complex maintenance or assembly tasks. The instructions and detailed information can be ingested for operators about each device. After completing each step, the holographic instructions can navigate to the position of the next device with animations. In this way, AR can decrease errors in complex processes.

Experts from different parts of the world can collaborate using AR. The suggestions with different modes and first-person view (FPV) videos of AR devices help to create an interactive and shared space experience [72]. For instance, remote technical experts from Germany consult United States dealership field specialists using AR at Mercedes-Benz. Specialists wear AR glasses and share their views with remote experts and ingest holographic instructions or suggestions similar to the telestrator drawings in real time.

AR can be used for the identification of assets and equipment in warehouses. Leveraging CV tools AR can track a particular asset or equipment accurately. Each machine's history of maintenance and need for replacement can be presented on the devices. The data about heating, ventilation, and air conditioning (HVAC) can be visualized in AR goggles. These kinds of data can be extracted from various sensors ofInternet of Things (IoT) devices.

Prototyping and refining products can be done using AR. Interaction with 3D virtual assets allows the designing of new products easily. AR can track logistics and supply chains and visualize the process in real time. Spatial information about ongoing processes can be superimposed on each device in detail. This enables workers not to look at separate instructions and displays while working. For instance, Wärtsilä Seals & Bearings is utilizing AR technology for marine maintenance services [73].

AR technology has numerous applications in robotics [74] as well. For instance, AR was utilized for the teleoperation of maintenance robots [75]. The system helped to control a real maintenance robot from a distance using a handheld controller and to visualize the position and orientation of the robot using AR. Safety aura visualization was developed to overlay safe and dangerous regions around the robots using AR [76].

2.12 Industrial Safety

Industrial environments carry hazards such as injuries and damage to equipment. The safety of workers, machinery, and facilities is critically important for various stakeholders such as the government, non-governmental organizations, worker unions, and employers. Hazards such as fire, electrical shock, radiation, moving machinery parts, excessive noise, high temperatures, and poor air quality can harm workers' health. To decrease the risk of getting harm to the workers and machinery, prudent safety practices should be applied in the industrial sector.

2.12.1 A Brief History of Industrialization

Starting in 18th century Europe, agriculture-based economies have transformed into industrial or manufacturing-based economies. Steam-powered machinery increased productivity in Great Britain. The manufacturing industry transforms raw materials into final products. The amalgamation of automated machinery, equipment, and proficient workers lead many countries to economic growth, better living standards, and greater average incomes. Cities were founded around workshops and factories. Mass production of goods caused complex division of labor between workers as well [77].

Industrial development can be divided into four industrial revolutions in history, as illustrated in Figure 2.6. Industry 1.0 began with the invention of steam and water power in the 18th century. These inventions lead to the mechanization of manufacturing and industries. Industry 2.0 occurred in the late 19th century and was driven by advancements in electric power.

The advent of computers and automation led to a new revolution in the 20th century. Integration of digital systems into industrial processes increased efficiency and accuracy.

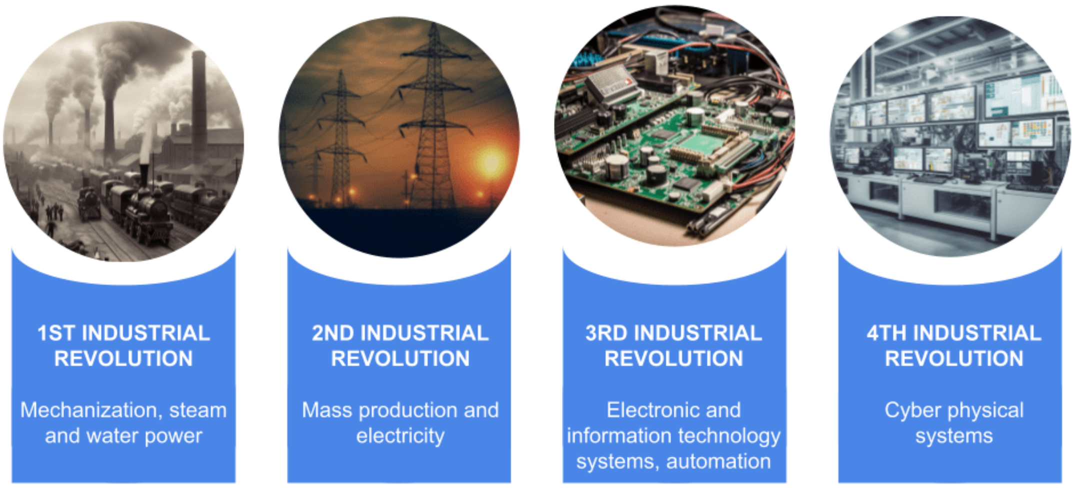


Figure 2.6: Industrial revolutions in human history

The digital revolution, Industry 3.0, made the manufacturing industry automated using robotics and other smart systems. The current state of manufacturing in developed countries is Industry 4.0 which converges new technologies such as the IoT, AR/VR, and AI. Industry 4.0 makes a significant shift towards the interconnection of all industrial processes and data-driven decision-making.

This field is evolving very fast in the last decade with the integration of new technologies such as the IoT, big data, ML, CV, AR, and robotics. Workers are performing inspection way faster using CV than it was a decade ago. And workers move heavy particles with pick-and-place robots easily without the cooperation of many workers and causing pain and trauma. Monitoring the working process is performed using digital twins and telepresence. The processes are automated thanks to autonomous machines. Embracing the new technologies will increase profit and decrease downtime.

2.12.2 Industrial Incidents

Industrial incidents are unforeseen events that occur by accident, negligence, or incompetence. The incidents can cause equipment damage, injuries, spread of disease, or death. The incidents are recorded in factories and their causes are analyzed. Repeated incidents with similar circumstances indicate a failure of risk management policies.

2.12.3 Distraction or Inattention

Distraction is a loss of concentration on a primary task. People can get distracted or their attention may get dispersed by external and internal factors [78]. External distractions can be phone calls, social media, surfing on the internet, noises, or other people. Intrusions of other people with unscheduled conversations can be the impetus for interruptions. Internal distractions can be originated from fatigue, stress, or ruminations.

Inconvenient body positioning or temperature can lead to distraction or inattention. Working without breaks and extended periods also results in fatigue which can lead to distraction. The interruptions can be divided into sequential and concurrent according to the degree of multitasking [79]. Sequential multitasking is switching from one task to another and disengaging completely from the previous tasks. Concurrent multitasking is performing multiple tasks almost simultaneously.

In industrial settings, there are many disruptions and distractions that draw the workers' attention. An operator's attention can be diverted for long periods of time by digital distractions such as phone calls, messages, or emails, according to [80]. Moreover, industrial employees' attention is diverted by auditory distractions such as low-intensity noises [81] and impromptu talks [82].

It was found that working alone with interruptions lasting for around three seconds result in twice as many mistakes as working alone without interruptions [83]. When minds and eyes are not concentrated on the ongoing task, the risk of injury increases. In addition to the serious risk of harm or death, accidents brought on by distractions result in service downtime of the manufacturing equipment, lowering productivity and causing significant financial losses.

2.12.4 Human Sensory System

There are five primary human senses: vision, smell, hearing, taste, and touch. Sensory receptors convey signals about different stimuli happening in the external world. The signals are in the form of electrical impulses that can be interpreted by the brain [84].

Human eyes have vision receptors called rods and cones that perceive light and color. The detected signals are transmitted through optic nerves to the brain. Human ears contain hair cells to sense sounds and auditory nerves to transmit impulses to the brain. The tongue has receptors that detect different flavors such as sweet, sour, salty, bitter, and umami. Scents can be detected by olfactory receptors and transmitted by the olfactory nerve to the brain. The human olfactory system collects data from a range of distances. Odorous molecules from objects in the environment go through the nasal cavity and contact olfactory receptors. Receptors are located next to the nasal cavity and nerve fibers of the olfactory system transmit signals to the brain. Our skins can sense touch, temperature, and pain. Touch is sensed by mechanoreceptors, while thermoreceptors sense temperature. Nociceptors detect pain and tissue damage.

Modern neuroscientists claim that we have more than the five classical senses [85]. For instance, equilibrioception refers to the perception of balance [86], i.e., the orientation of the body perceived with respect to gravity. Proprioception is awareness of body parts without looking into them [87]. Kinaestia refers to the sense of movement of body parts. Chronoception is the perception of duration or time passing [88].

There are sensing capabilities that are found in animals but not in humans. For example, the electroreception of sharks helps find their prey by means of the electrical field around them even if they don't see it [89]. Magnetoreception helps bats orient with respect to the magnetic field of the Earth [90]. Polarized waves of light are used by insects and birds for navigation [91]. Dogs have extraordinary smelling capabilities which allows them to find the source and history of the specific smell[92]. Echolocation is the ability of dolphins to orient in space using sound waves bouncing from the objects [93]. Migratory birds sense magnetic fields to navigate during long-distance migrations [94].

Nowadays, modern intelligent systems utilize sensors that are beyond naturally possible. For instance, light detection and ranging (Lidar) sensors can create a 3D map of the environment emitting laser pulses and measuring the time or phase of returning light [95]. Accelerometer and gyroscopes allow the system to measure the orientation and speed of the attached object [96]. One of the main goals of AR technology is to expand human perception. This technology has the potential to connect the aforementioned sensing capabilities to humans.

The reaction time of humans to auditory cues is around 150 ms [97] while people react to visual cues under 200 ms [98]. Reaction time to olfactory stimuli differentiates according to the type of olfactory and adaptation to the environment [99]. Iravani et al. [100] showed that humans react faster to negative smells, which, in turn, triggers a fast avoidance response. It is estimated that the response to negative odors is roughly 300 ms. The reaction time of the human gustatory system to different tastes varies between 200 ms and 700 ms [101].

Warning modalities should address the challenges and issues of the specific industrial environment. For instance, stench gas is used to warn workers about coming hazards in the mining industry since noise and dust pollution don't allow workers to hear or see well in mines and quarries. Presumably, expanding human perception with cutting-edge sensors and technologies such as AR and AI can be the solution to the challenges of the industrial environment.

By fusing data coming from various sensors, AI can represent the information in a way that humans can respond faster. AI integrated with AR can represent the relevant information directly in the field of view of the user, trigger tactile alerts, and imitate a spatial sound indicating the position of the coming hazard. Predictive analytics of AI can be implemented by using the history of events and issues. When the system gets data similar to past incident data, the system can convey alerting signals to humans using AR technology.

After mentioning the advantages of modern technologies, the limitations should be discussed thoroughly. Modern mobile and wearable AR device lacks computing power. Data coming from the sensors can be processed in edge computing devices. However, this adds latency to transmitting data between devices. These can be dangerous in emergency situations.

2.12.5 Prevention of Industrial Injuries

2.12.5.1 Safety Measures

Safety measures are implemented to protect workers and prevent accidents [102]. First of all, workplaces should be inspected regularly by safety experts to identify risks and their solutions. If workers report potential hazards, these should also be considered. Workers should make sure they are working in a safe environment and should be encouraged to report risky conditions.

Regular training of employees is mandatory according to industry regulations in developed countries [103]. The machine and equipment are allowed to be used by only authorized workers with proper training. Personal protective equipment such as safety glasses, helmets, respirators, gloves, steel-toed boots, earplugs, and protective clothing should be worn properly in the workplaces. Machines that present major hazards should be isolated by guard rails equipped with interlocks and emergency stop switches.

Safety signs, labels, and warnings should be placed in visible places [104]. Fire extinguishers and first aid kits should be available for emergency situations. Emergency response plans should be developed and emergency exits should be arranged for rapid evacuation. Ergonomic principles of using tools and lifting heavy assets should not be omitted to avoid the risk of getting musculoskeletal disorders or other injuries. Storage, handling, and disposal of hazardous materials should be implemented according to industrial standards. Labeling containers and management of hazardous assets are included in these regulations. Employers and employees should support developing a good safety culture in the workplace. Safety measures should be improved continuously.

2.12.5.2 Effective Training

Employee training is an imperative process to improve productivity and prevent accidents in the manufacturing industry. Training employees on how to utilize equipment and machine could save time, financial resources, and, most importantly, human lives. Untrained or undertrained employees would decrease productivity and endanger their safety in the factory. The operator should be aware of the policies of the factory and have proficient knowledge regarding their job. Knowledge deficit should be eliminated to improve the performance of workers.

Persistent training would help to avoid many incidents [105]. The training process should be interactive and hands-on to improve employees' practical skills. Demonstrating all the tasks or showing examples without rushing is vital for fostering important skills in the working process. Presentations or printed instructions might be boring and employees might forget them in the near time. The environment where employees will practice should be safe and controlled. In particular, the tasks with semi-automated machines must involve monitoring by experienced trainers. Otherwise, employees might get injured in one of the first practices with the machine. The practical training sessions provided in such a way will increase the confidence and problem-solving skills of the workers. Training should be continuous and periodic to consider the new hazards that the working process might bring.

Employers, team leads, or trainers should be clear in their instructions and should give constructive feedback to employees. The feedback should not be based on emotions. Feedback sometimes should be in the form of advice and not only as criticism. The employee should be allowed to give feedback on the manager's wrong interpretations or be able to inform the manager of dangerous situations or conditions. Otherwise, they may hide some dangerous situations until they got to incidents. Employers should have regular safety meetings with their workers where all hazardous situations should be discussed. Inspections of equipment, environment, and products should be conducted according to regulations.

Emergency response training is significant to develop life-saving skills in the event of impending or occurring disasters [106]. The employee must have roles and responsibilities in emergencies the same as in normal times. For instance, some employees should be responsible for performing first aid in the event someone gets injured. The workers need to understand the type of alerts and actions that should be taken. Employees should be trained to find and use emergency equipment in disasters. Wrong actions during an emergency might cause more injuries or fatalities. Right emergency response actions can mitigate future incidents.

In case of an incident, the severity of the incident should be evaluated and any hazards should be mitigated. First aid should be provided and other emergency departments should be called if needed. All the details of incidents should be investigated to identify the causes. Preventive measures should be implemented to avoid similar incidents in the future.

2.12.5.3 Personal Protective Equipment

Professional attire should not be neglected in the industry. The proper clothing might help workers to avoid injuries or fatalities. The term personal protective equipment (PPE) refers to clothing worn to minimize exposure to chemical, electrical, radiological, and biological hazards. PPEs protect the eyes, head, ears, hands, feet, body, respiratory system, or whole body (see Fig. 2.7). PPEs can be in the form of hard hats, goggles, gloves, coveralls, shoes, respirators, earplugs, and earplugs. Before wearing PPEs workers should make sure that they are not deformed, torn, or dented. Well-fitted PPEs can protect the body from burns, electric shocks, radiation, or chemical

Figure 2.7: Types of personal protective equipment.

subjects. PPEs should be stored in storage rooms that are clean and protected from all kinds of damage.

Hard hats are must-have protection in factories that risk objects falling to the head. Requiring to wear hard hats and training workers to wear hats properly is the task of employers. The damaged hard hats should be replaced timely.

Eyes should be protected with proper safety glasses during work. After work, the workplace should provide eyewash facilities where workers can be exposed to chemicals. Safety goggles such as for welding protect eyes from extreme light, radiation, and flying particles of metal.

Gloves are worn for protecting hands from diseases, hot objects, poisonous objects, and vibrations that might lead to cuts from the machine such as a saw machine. Gloves are made of different materials such as leather, rubber, vinyl, nitrile, or latex. Gloves can be long-lasting or disposable. Disposable or contaminated gloves should be changed on time.

Ear muffs or earplugs are worn to protect ears from high noise levels in the industry. If not worn regularly in noisy factory settings, workers can develop hearing problems. The earplugs should be clean and should fit the ears. Workers are recommended to wear earplugs where the level of noise extends 85 decibels [107].

2.12.5.4 Smart IoT Solutions for Industrial Safety

Industry 4.0 aims for widespread digitalization of the manufacturing and industry [108]. Intelligent technologies such as cyber-physical systems (CPS) and IoT allow manufacturing to increase efficiency and decrease cost. CPS as the basis of Industry 4.0 refers to the system that bridges computation and physical systems. The twin term IoT interconnects the devices, machines, sensors, and computing devices such as cloud computing and node computing for smart solutions. It is predicted that by 2025 the number of devices that are connected to the internet will be 55.7 billion in the world [109]. Software solutions with AI, AR, big data, cloud computing, and data analysis can push the manufacturing renaissance further.

The data obtained during machine operation allows manufacturers to analyze the quality of products, refine the services, streamline performance, and increase production. In addition, a data-driven supply chain helps to supply on-demand products. Machine-to-machine connectivity supports quality data service and visibility of all processes. 5G networks can be utilized for the interconnection of equipment and for conveying information to operators with high bandwidth and low latency.

Cloud computing enables considering on-demand products in changing marketplace and planning production. VR and AR technologies allow customers to interact with holographic products. These technologies have the potential to revolutionize the customer experience. Using various sensors for tracking the environment in real-time will give a digital representation of the factory. The potential hazards can be analyzed and future accidents and injuries can be avoided. Smart systems trigger on-time equipment maintenance warnings to decrease downtime in the factory. Warehouse operations and inventory management can be automated using IoT technology.

The safety of workers and equipment can be enhanced using various sensors in the factory. Arduous and repetitive jobs can be automated to decrease work-related repetitive strain injury (RSI). First-person view instructions can be given using AR and VR. Some inspections and monitoring of operators can be handed over to smart systems. However, smart systems did not reach the technical maturity to do all inspections and monitoring yet.

2.13 AR for Industrial Safety

AR technology can be utilized to warn operators of possible dangers and hazards. The headsets can convey potential hazards in the field of view of the user. The information from IoT devices can be displayed in real-time. AR warnings with different cues can be initiated for people with disabilities.

2.13.1 AR-based Training

Novice workers can be trained with AR assistants in semi-virtual environments in factories. AR can represent information about current processes with realistic visualizations. The responsibilities of each team member can be taught safely in a simulated environment. The AR system can provide cues if the worker's progress is slow. The performance of workers can be increased in this way. Personalized feedback to each worker can be conveyed instantly.

Safety training is an essential part of industrial training. AR allows inexperienced workers to make mistakes with no serious consequences. Real-time feedback on each hazardous scenario can diminish future errors.

2.13.2 AR Glasses as Personal Protective Equipment

There were companies that developed AR headsets as PPE. Specifically, Daqri presented an AR-based smart helmet that protects the head, eyes, and ears in 2016 [110]. The helmet integrated state-of-the-art sensors such as thermal and 360-degree field of view cameras. However, the project failed in 2019 and the company went off business reportedly [111].

There are very few companies that introduced new HMDs as PPES or PPE versions of their existing AR goggles. Note-worthily, Microsoft is one of the companies that presented a PPE version of their famous Hololens 2 glasses. Trimble XR10 [112] is a hard hat integrated with Hololens 2. It protects the head and eyes whilst giving immersive information about the industrial environment. VisualLive hardhat adapters were introduced for attaching Hololens 2 AR glasses to regular hard hats [113].

There were AR applications where virtual assistants showed how to don PPEs properly. For instance, Honeywell introduced an AR application called "Honeywell Connected Plant Skills Insight Immersive Competency" which provides virtual training [114]. The application was developed for AR headsets such as Microsoft Hololens.

RUMEN AR glasses were created to be worn on the helmets of industrial workers [115]. Realwear presented a ruggedized Assisted Reality-powered wearable Tablet that has various sensors such as thermal and environmental cameras [116]. Attached to the hardhat, the device allows controlling the menu with voice commands. Realwave glasses are productively used by electric companies to restore service after storms and hurricanes [117]. The glasses are powered with geographic information systems which have detailed information about each component of the infrastructure. HMT-1Z1 glasses of RealWave received certifications to be used in hazardous environments such as the mining, oil, and gas industries [118]. British Petroleum, Honeywell, Shell, and ExxonMobil are major industrial companies that utilize these glasses. Youbiquo from Italy also released its Talen Halo AR headset for industrial maintenance [119].

2.14 Systems for Preventing Distractions

Researchers have suggested a number of methods, including the Interruption Recovery Assistant (IRA) [120] and the Interruption Assistance Interface (IAI) [121] to deal with interruptions and diversions and Attention Manager (AM) [122]. IRA was created for mission commanders to monitor and manage convoys and unmanned aerial vehicles (UAVs) to better defend ground forces [120]. The IRA gives the commander the ability to rewatch historical events by providing an interactive timeline of significant occurrences. In simulation experiments, the system assisted participants in picking up where they left off more quickly and making more informed decisions. Another system, IAI, offered two types of support to participants in a work involving UAV monitoring and replanning: the bookmarked aid mode displayed the tactical position of UAVs when a bookmark was clicked, and the animated assistance mode played back an animated sequence of events.

With both modes, participants' decisions were more accurate. In contrast, AM was created to keep the user's focus from notifications until it determines an appropriate time to interrupt them with the subsequent application message [122]. It was recommended that task models be uploaded to the manager or that a pipeline be set up to gradually learn task models for creating AMs [123]. For naval air combat, the Change History Explicit (CHEX) tool was created to regain situational awareness following interruptions. The tool was created to keep track of changes in a change history table. This system enabled participants to easily recover the most recent changes after experiment interruptions. These interruption recovery systems assisted in restoring the user's previous actions but did not concentrate on preventing interruptions.

2.15 Warning Systems

To avoid distractions, warning systems have been created in various areas [124, 125, 126, 127]. In the automotive sector, a lane departure warning system has been developed that triggers the user with haptic and auditory warnings [124]. On a head-up display (HUD), the EUCLIDE anti-collision warning system for passenger automobiles alerts the driver both audibly and visually [125]. To avoid collisions with pedestrians, another driver assistance system with visual cues has been suggested [126]. The efficacy of an AR warning system on driving behavior at crosswalks was studied by Calvi et al. [127]. Visual and audiovisual (a beep sound) alerts were examined in a driving simulator as two separate warning modalities. In [128], participants performed driving tasks with a pedestrian warning system in a realistic scenario. A HUD, installed in a vehicle, warned the driver of a crosswalk with a visual sign within 13 m, while the system alerted the driver with an alarm within 8 m. A warning system for cyclists was suggested in [129] to avoid collisions with the opening doors of parked cars.

2.16 Warning Modalities

AR systems can include output devices with auditory, visual, haptic, somatosensory, and olfactory modalities. Warning systems are integrated with visual and haptic cues for lane keeping in modern cars of Mercedes-Benz [124]. The system shows visual warnings and vibrates the steering wheel to pull the driver's attention back to driving.

Stench gas is productively used to warn workers about emergencies in noisy environments such as mining [130]. These provoke memories of danger and initiate a trained response.

Visual alerts can be text, signs, or animations with clear instructions or information. Visual warnings are essential for people with hearing disorders. In noisy environments, visual warnings can help workers to convey about emergency situations [131]. Visual warnings can be shown on smartphones, PCs, big digital screens, or television screens.

Voice alerts can be spoken warnings or different sound signals to warn individuals about potential dangers. Sound tones include horns, sirens, chimes, or beep sounds. Spoken messages are pre-recorded or digitally synthesized. It can be activated on smartphones, voice-enabled devices, or public address systems. It is valuable for people with hearing disabilities.

Wireless Emergency Alerts (WEA) can be broadcasted to specific geographical areas if an emergency affects a big area [132]. Emergency alerts on phones can be in visual and audio forms.

The combination of multiple modalities allows for the effective conveyance of warnings. Visual, auditory, textual, and tactile cues each bring unique benefits in various situations. The flexibility to choose and combine different modalities can help to optimize the impact of warnings and mitigate risks. Adaptable and comprehensive warning systems can convey vital information in emergency situations. However, research should be conducted to take into account the diverse peculiarities of each modality and ensure efficient communication between the worker and the warning systems.

2.17 Formulation of research questions

Industrial incidents can be caused by distractions, such as unscheduled conversations and phone calls. To address inattention, various approaches have been developed such as the Interruption Recovery Assistant (IRA) [120] and the Interruption Assistance Interface (IAI) [121]. To prevent inattention in the automotive industry, eye-tracking approaches were implemented as well [133, 134, 135]. Eye tracking approaches can be divided into target-, direction-, and purpose-based eye movement interpretations [136]. Glance targets are coded with types such as bicyclist, pedestrian, etc. in the target-based approach [137], whereas the direction-based approach tracks the direction of eye movements [138]. The purpose-based approach analyzes the probable reason for glances in humans [139]. Amidst these approaches, we followed the target-based approach in eye-tracking to identify the inattention of the user. We integrated eye-tracking technology into our system, seeking to combat inattention in dynamic industrial environments. **This way, we aimed to determine if our system reduces human operator inattention in our first research question (RQ1)**.

For our warning system, we devised three cues: auditory, visual, and audiovisual. We decided to compare these warning types according to the perception of human participants in the experiment. Human perception can be assessed using questionnaires or interviews. There are studies in the literature that evaluate human subjective measurements using questionnaires. For instance, DeMelo et al. [140] investigated the cognitive load that participants experienced with and without virtual aids in a simulated AR military environment. The AR framework included embodied virtual assistants and auditory assistants. The findings indicated that participants experienced a lower cognitive load when they received assistance from an auditory assistant. Chromaglasses AR-based camera system was developed specifically for individuals with color blindness [141]. The usability of the system and participants' cognitive load was evaluated using a questionnaire. Additionally, the realism of the experience was asked in the questionnaire. Erickson et al. [142] developed an AR application to share gaze rays with collaborators. The questionnaire included questions assessing the usability, cognitive load, ease of use, and perceived realism of the AR system. **We intended to understand how individuals perceive the different modalities (auditory, visual, and audiovisual) in industrial environments in the second research question(RQ2)**.

The third research question (RQ3) aimed to assess the advantage of our system over conventional warning systems. Various warning systems are employed in the industry. We compared our system with a sensorized safety mat using objective measurements. Safety mats are primarily used to prevent the entry of human operators into manipulator areas [143] and serve as an anti-fatigue sensor and warning device for prolonged standing tasks in the industryy [144, 145].

The World Health Organization (WHO) has published burn injury prevention plans covering various aspects of life, including tradition, lifestyle, and country regulations [146]. Standard safety plans in the industry may not guarantee the absence of burns. Exposure of industrial workers to high temperatures can result in heat stress, a condition where the body cannot maintain a standard temperature balance between 36°C and 37°C [147].

To prevent heatstroke, wearable technology has been developed, such as the nanosheet skin-based patch containing pH sensors. This skin-adhesive and small device is designed for burn injury prevention [148]. Additionally, a battery-free and AI-enabled sensorized patch was proposed for wound monitoring [149]. The PETAL (Paper-like Battery-free In situ AI-enabled Multiplexed) sensorized patch can detect the level of burn injury and consists of five colorized sensors, holding potential to enhance burn injury prevention in the industry. However, these sensors may lack the ability to understand the temperature of objects placed far from the user.

Multispectral cameras have been employed to detect heat in objects [68]. IR and ultraviolet cameras integrated with Microsoft Hololens AR glasses conveyed information about the thermal condition of objects in the field of view. After experiments with human participants, a debriefing session revealed that our system simplified tasks that would have otherwise required more time or effort. Discussions also touched upon the potential of novel signals for interpersonal communication, daily activities, or personal health. To assess human participants' subjective perceptions, we conducted semi-structured interviews after the experiments. **We evaluated the perception of participants in the fourth research question (RQ4)**.

Chapter 3

An AR-Based Warning System for Industrial Safety

Distraction or inattention is one of the major causes of accidents in the industry. In order to prevent distractions in industrial settings, we introduce an AR-based warning system in this work (see Fig. 3.1). The system provides auditory, visual, and audiovisual alerts when the user averts the gaze from the machine. The efficacy of the system was examined in a scenario involving a desktop computer numerical control (CNC) machine.

The novelty of this method is due in part to a number of factors:

- It is the first AR warning system developed for industrial safety.
- We present DL-based object detection for AR systems.
- Our solution was built for untethered AR smart goggles (Microsoft HoloLens 2), which are neither location-specific like spatial AR nor require hands like mobile AR.
- In addition, our system seamlessly displays notifications within the user's field of perception. On the other hand, notifications on the screens of mobile devices and tablets can divert users' attention.

3.1 System Architecture

The flowchart of the system operation is shown in Fig. 3.2. The system is launched on wearable AR glasses and establishes a wireless connection to a laptop-based object detection server. Next, the AR smart glasses start sending frames from the built-in camera to the server in real-time for frame-by-frame object detection. In this way, the computing constraint of the AR HMD is bypassed. The server returns data about the items it has detected to the AR glasses. When the CNC machine is detected (SI2), the smart glasses determine the position of the machine using spatial perception (SI1) and overlay the machine with a virtual hologram (SI3). The user is asked to confirm the correctness of the machine's position with hand gestures and dwelling their head toward the direction of the machine for a few seconds (SI4)

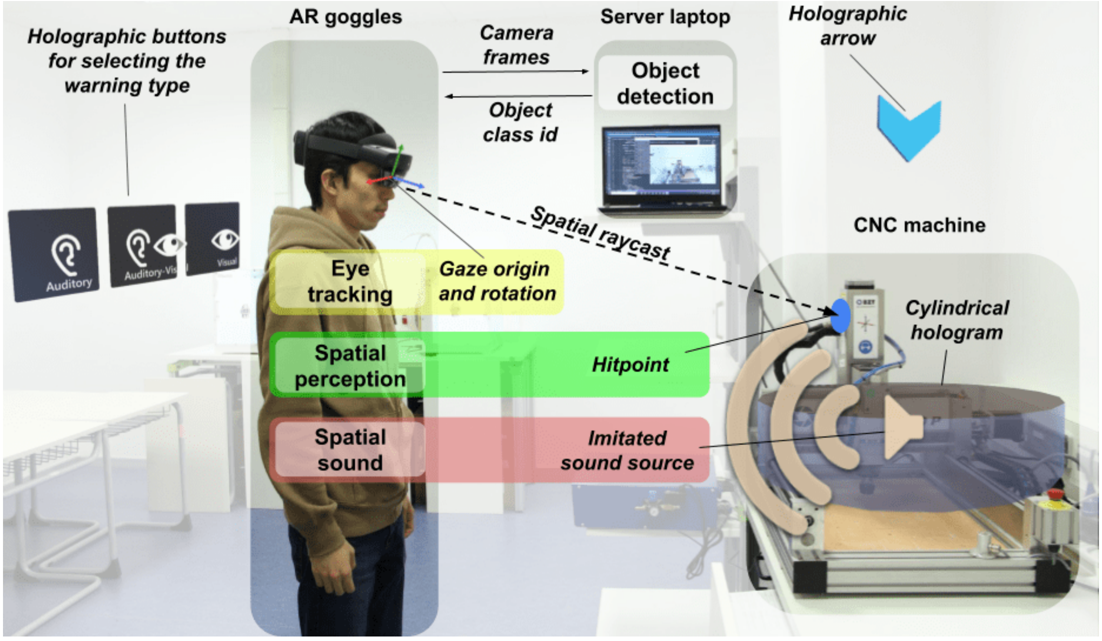


Figure 3.1: Overview of the warning system showing the holographic buttons, cylindrical hologram, holographic arrow, object detection server, and the user wearing AR goggles

Three virtual buttons with warning-type icons (auditory, visual, and audiovisual) are displayed on the machine's opposite side. The operator selects a warning type by pressing the associated holographic button (SI5). After the operator has set the type of warning, the AR glasses begin to track the operator's gaze using the eye-tracking cameras. Once the operator has looked at the machine for the first time (SO1), the warning system is activated. The system issues a warning with the selected modality as soon as the operator looks away from the machine (SO2). In particular, the virtual arrow (SO3B1) in the visual warning system (SO3B) displays the direction of the machine, while the auditory system (SO3A) beeps a warning from the machine's location (SO3A1) to warn others. The audiovisual warning concurrently integrates both warning modalities (SO3C). The purpose of the AR warning system was to prevent operator inattention while operating an industrial machine, which can result in accidents and injuries.

3.2 System Implementation

3.2.1 Augmented Reality

The warning system comprises two main components, namely the system initialization and system operation parts, as diagrammed in Fig. 3.2. The system initialization part encompasses several elements, including spatial perception (Subsection 3.2.3), machine detection (Subsection 3.2.2), holographic overlay over the machine, verification of the

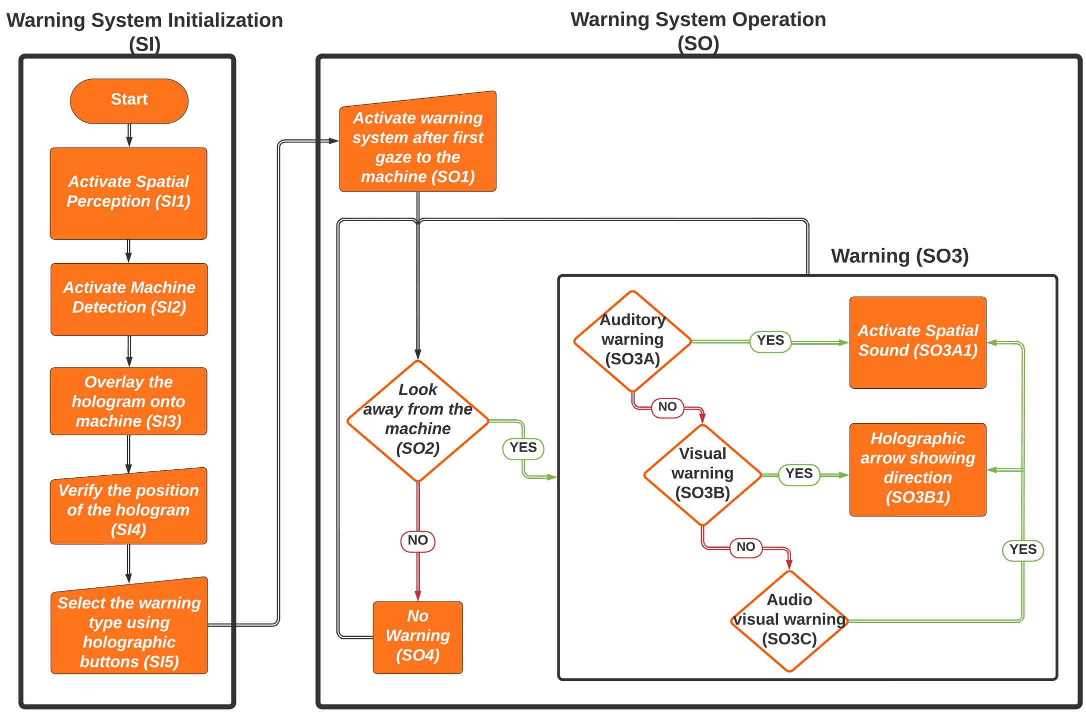


Figure 3.2: The flowchart of the AR-based industrial warning system illustrating the logic of the system operation

hologram's position, and selection of the warning type (Subsection 3.2.4). The AR headset employed in this system allows hand tracking of 26 joints and can recognize various hand gestures such as air tap, touch, push, and hand ray. To confirm the machine's position, we utilized the air-tap gesture. The system's subsequent hand interaction was choosing the type of warning by pressing the holographic buttons with the index finger. In the system operation part, the warnings were predicated on the direction of the user's gaze, obviating the need for hand interactions or voice commands. To ascertain the user's gaze direction, we leveraged the eye-tracking functionality of the AR glasses, as discussed in Subsection 3.2.5. Auditory warnings in the form of spatial sound were delivered through the headphones of the AR goggles, as elucidated in Subsection 3.2.6. Similarly, visual warnings in the shape of holographic arrows were displayed on the visor of the AR glasses.

The warning system employed a client/server setup, comprising the Microsoft HoloLens 2 AR goggles and a laptop computer, as depicted in Figure 3.1. These components connected with each other via a wireless network adapter, specifically the Cisco Aironet 1800. As for the ML server, we utilized a laptop computer (Intel Core i5 9300H central processing unit, 8 GB DDR3 memory, and Nvidia GTX 1650 GPU) running on Windows 10 operating system. The development of the HoloLens 2 application was carried out on the Unity 2018.4.36 platform, employing the MRTK v2.0.0. Subsequently, the application was deployed on the AR goggles through Microsoft Visual Studio 2019. On the server side, we employed a Python 3.7 script to

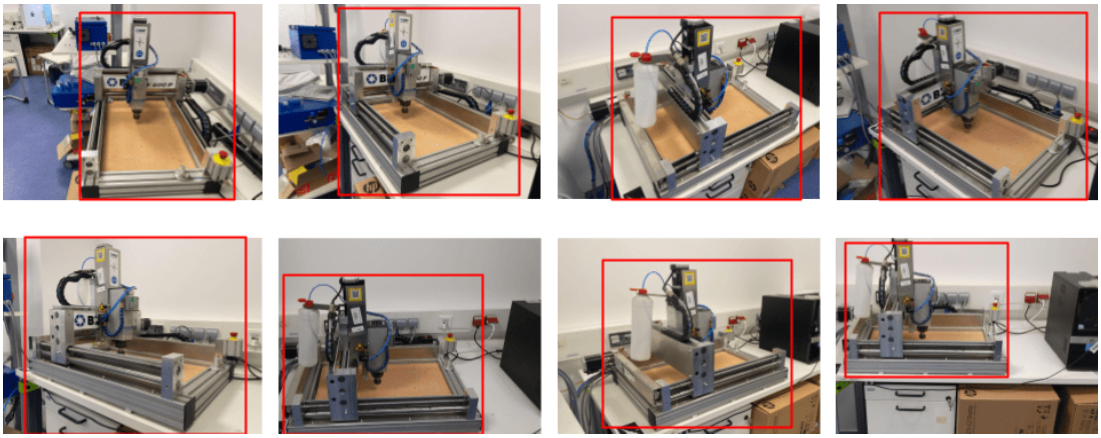


Figure 3.3: Sample images of the CNC machine with bounding boxes from the dataset

execute the deep neural network responsible for object detection. The script utilized OpenCV libraries, enabling seamless integration of computer vision capabilities.

3.2.2 AI-Based Machine Detection

The significant role of cloud and edge computing in delivering substantial computational resources to mobile devices like smartphones, smartwatches, and smart glasses is indispensable. In our research, we further harnessed the power of edge computing for a computer vision task. To be specific, the AR goggles captured RGB frames with a resolution of 896×504, which were then transmitted to the ML server via a TCP/IP socket. We deliberately chose this lower resolution instead of the maximum resolution of the 2272×1278 to minimize the wireless transmission time of the images to the server. Subsequently, a Python script executed the object detection process.

We employed the YOLOv4-tiny model for object detection, as documented by Bochkovskiy et al. in [150]. The YOLOv4-tiny model stands out due to its exceptional combination of speed and accuracy, along with an improved classifier compared to earlier iterations of YOLO. This choice ensures that our system benefits from the advancements and refinements introduced in this version. To achieve optimal performance, we trained the model using transfer learning. Transfer learning is an ML technique that enables the transfer of knowledge acquired from one task to another [151]. By leveraging this method, we harness the existing knowledge within the YOLOv4-tiny model and adapt it to our specific object detection task.

In our transfer learning approach, we utilized the pre-trained weights of the first 29 layers of YOLOv4-tiny, which were originally trained on the COCO dataset [152]. The COCO dataset is a large-scale dataset created by Microsoft for object detection, segmentation, and captioning tasks. This allowed us to leverage the knowledge distilled from the COCO dataset and apply it to our specific task. Additionally, we made adjustments to the number of filters in the convolutional layers that precede the two

YOLO layers of YOLOv4-tiny, setting it to 18. This modification was made to optimize the model's performance and align it with the requirements of our object detection task. To train our DL model, we collected a dataset consisting of 100 images of the CNC machine from various angles and distances, as depicted in Figure 3.3. We manually labeled the bounding boxes around the CNC machine using the LabelImg program with the only class label "CNC machine". The dataset was divided into a training set (70%) and a test set (30%) to evaluate the model's performance. During the training process, we configured the batch size to be 2 and used a subdivision value of 8 to facilitate efficient training. These choices were made to make a balance between computational resources and model optimization.

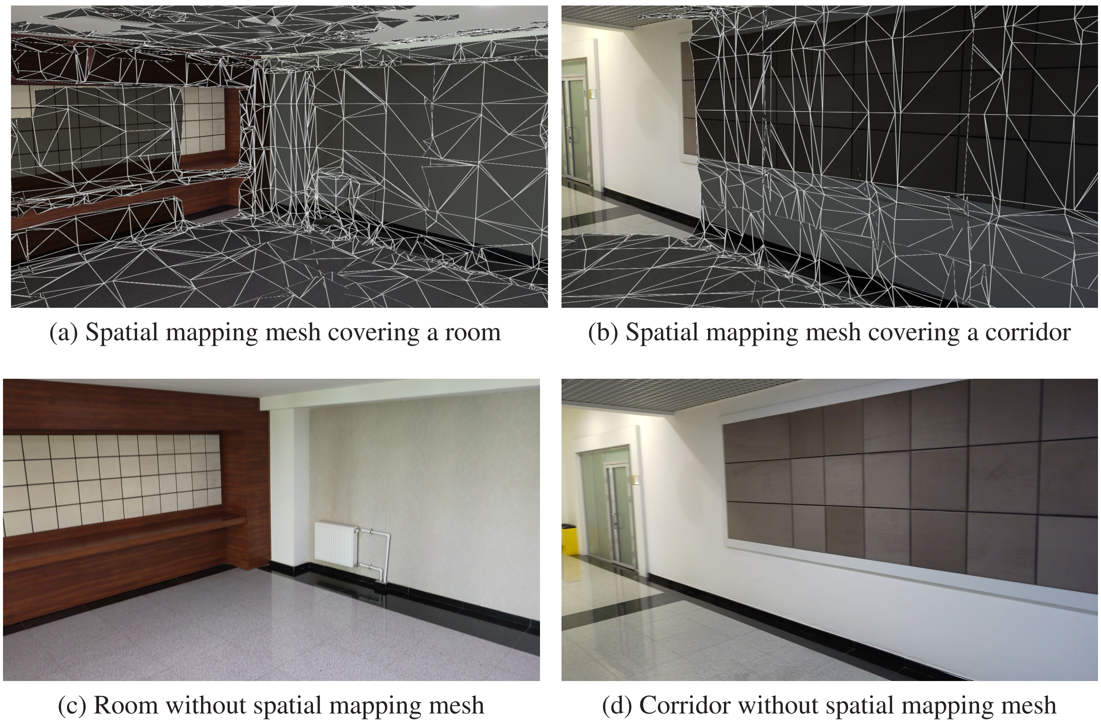
The model underwent training for 2,000 epochs using specific hyperparameters. These included a learning rate of 0.00261, a decay rate of 0.0005, and a momentum of 0.9. To enhance the training process, we employed the mosaic augmentation technique, which expanded the dataset to a total of 16,000 images. Mosaic data augmentation was introduced in YoloV4 and enabled the model to generalize and teach to recognize objects without relying on the context. The training was conducted for a duration of 30 minutes utilizing Google Colab on a Tesla K80 GPU, taking advantage of its computational power. Following the training phase, we evaluated the model's performance on the test set.

The common metric for measuring the performance of object detection is the Average Precision (AP). AP was used to assess the performance and accuracy of the Faster R-CNN, SSD, and Yolo object detection algorithms. AP is based on sub-metrics such as confusion matrix, precision, recall, and Intersection over Union (IoU). The confusion matrix consists of true positive, true negative, false positive, and false negative parameters. In a confusion matrix, a true positive occurs when the model accurately predicts a positive class. False positive refers to the model incorrectly predicting a positive class when it should have been negative. True negative happens when the model correctly predicts a negative class. False negative, on the other hand, is when the model incorrectly predicts a negative class. Precision quantifies how well the model finds true positives out of all positives whilst recall measures true positives out of all predictions. IOU calculates the overlap between the predicted bounding box or segmentation mask and the ground truth bounding box or mask. It is calculated by dividing the area of intersection between the predicted and ground truth regions by the area of union between them. The IOU value ranges between 0 and 1, with higher values indicating higher accuracy. With an IOU threshold of 0.5, the AP achieved by the model was measured to be 90.3%. These results indicate the effectiveness of the model in accurately detecting objects within the specified IOU threshold.

The duration required for transmitting frames from the AR goggles to the ML server on the laptop depended on the performance of the wireless network adapter. On average, it took approximately 210 ms to send a single image. The subsequent object detection process, on the other hand, had an average execution time of 92 ms. As a consequence, the system achieved an average operational rate of three frames per second. These metrics provide insight into the efficiency and speed at which the system processed and analyzed the frames.

3.2.3 Spatial Perception

To gather environmental and hand interaction data, we utilized a combination of four environmental cameras, IMU sensors, and depth cameras, including both near and far-range capabilities. Within the MRTK, the Spatial mapping module played a key role. This module effectively scanned the real-world environment and generated triangle meshes, which were then attached to the corresponding surfaces In the Unity platform, the Physics API was employed to establish the physics properties of GameObjects and Spatial mapping. Specifically, the Physics API offered a function called *Raycast*, which emitted a virtual ray from the head-mounted display (HMD) in a specified direction. This function determined whether the ray intersected with the virtual hologram or the spatial mapping mesh. The precise point at which the ray collided with the mesh was referred to as the *hitpoint*, as shown in Figure 3.1. For the purpose of machine position determination, we utilized raycasting through the spatial map. By casting rays and detecting collisions, we were able to identify the exact location of the machine.



(a) Spatial mapping mesh covering a room

(b) Spatial mapping mesh covering a corridor

(c) Room without spatial mapping mesh

(d) Corridor without spatial mapping mesh

Figure 3.4: Examples of spatial mapping meshes in Microsof Hololens 2 covering a room (3.4a, 3.4c) and a corridor (3.4b, 3.4d)

Consequently, we superimposed a transparent cylindrical hologram onto the hitpoint generated by the raycast, creating a visual representation aligned with the machine's position.

3.2.4 Warning Modalities

We integrated three holographic buttons into the system using the MRTK. These buttons were designed with distinct names and icons corresponding to different warnings, as depicted in Figure 3.5. By pressing one of these buttons, the user activates the selected warning modality. Additionally, the operator has the flexibility to switch between warning modalities by pressing another button. The auditory warning system, leveraging the spatial sound feature of the AR goggles, emits a beeping sound from the position of the hologram. This auditory cue can effectively alert the user of potential hazards. In the visual warning system, a holographic blue arrow is employed to indicate the direction of the CNC machine. This visual cue provides clear visual guidance to the user, facilitating situational awareness. Furthermore, we implemented an audiovisual warning system that combines both auditory and visual warning types. By utilizing these two warning modalities simultaneously, we wanted to enhance the effectiveness of conveying critical information to the user.

3.2.5 Eye Tracking

Eye tracking plays a significant role in understanding the user's visual attention, intent, and focus. In the case of the Microsoft HoloLens 2, the eye-tracking functionality relies on two infrared cameras in conjunction with infrared LEDs to accurately track the user's gaze. Within the software framework, MRTK provides the means to extract essential data about the origin and direction of gaze.

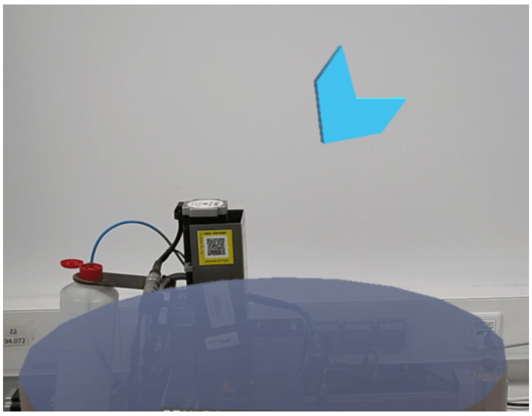
Leveraging the capabilities of the MRTK, we utilized decision structures and handlers to determine whether the user was gazing toward the hologram or elsewhere. To achieve this, we developed separate functions for each condition and incorporated them into the decision structure statements (see Appendix C). During our experimental study, we incorporated the eye calibration settings of the AR goggles to derive the inter-pupillary distance (IPD) for each individual participant. This calibration step ensured that the eye-tracking system accurately captured and interpreted the user's gaze behavior. By leveraging eye-tracking technology and employing tailored decision structures, we aimed to gain valuable insights into the user's visual interaction with the holographic elements, ultimately enhancing the safety of the user in our research.

3.2.6 Spatial Sound

Spatial sound is a reproduction of audio in a 3D sound environment, mimicking the way sound is heard in the real world. Spatial sound gives the illusion of surrounding sound in 3D space whilst traditional audio utilizes left and right stereo channels of speakers and earphones. One of the breakthroughs of spatial sound, Dolby Atmos technology for cinemas, can be free from channels and distribute 128 discrete audio tracks and up to 64 unique speaker feeds [153]. The spatial sound helped visually impaired people in indoor environments [154]. The deviation of localizing sound was less than 10° in the experiment with visually impaired and sighted people. Surgical Simulator was developed to train novice surgeons using Microsoft Hololens 2 for Hip Arthroplasty [155]. The simulator utilized the spatial sound capabilities of the AR glasses. Participants in the experiment reported that spatial sound made the experience more immersive.

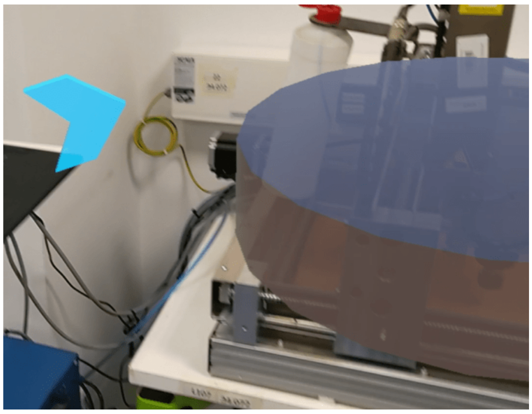
The head-related transfer function (HRTF)-based technology is used in the Microsoft HoloLens 2 to create spatial sound experiences. This technology is pivotal in reproducing sound as perceived from a specific position and orientation within a spatial environment. It is essential to note that individuals possess distinct head sizes, ear sizes,

Figure 3.5: Holographic buttons for selecting the warning types with the MRTK holographic mesh overlaid to the hand



(a) Top side

(b) Bottom side



(c) Right side

(d) Left side

Figure 3.6: Holographic arrow pointing to the machine from different positions during warning system operation in the visual mode

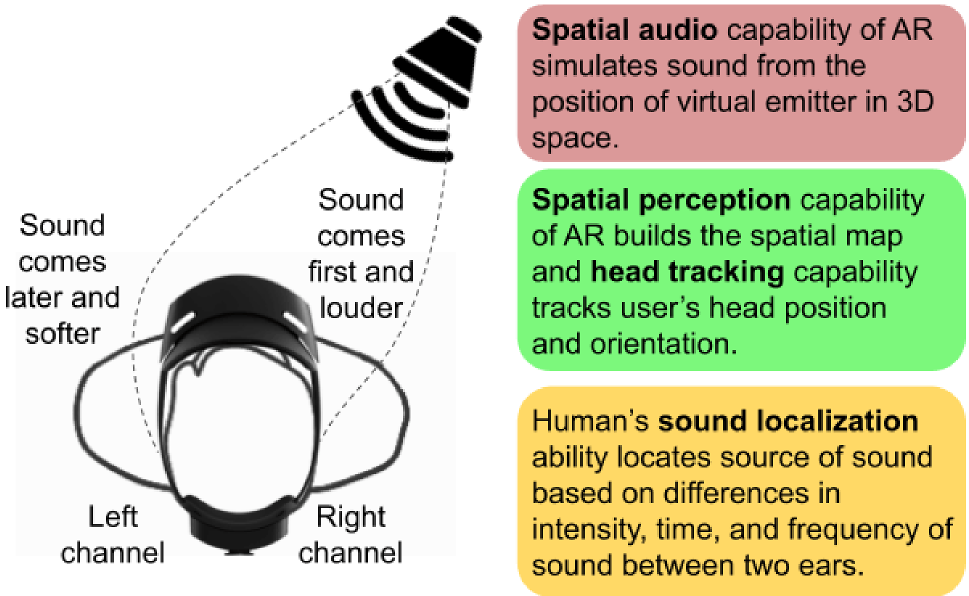


Figure 3.7: Spatial sound, head tracking, and spatial perception capabilities of AR glasses and sound localization ability of human

shapes (pinnae), and positions, which impact their auditory perception. To account for these individual differences, the Microsoft HoloLens 2 dynamically adjusts the HRTF based on the user's head size, taking into consideration their IPD. HRTF-based customization of sound spatialization enables the sound localization ability of humans to locate the source of sound and gives a personalized immersive sound experience (see Fig. 3.7). In our implementation, we attached an audio source component to the hologram overlaying the CNC machine. By enabling the spatialize feature provided by the MRTK, we were able to leverage the HRTF technology and achieve spatial sound rendering. During system operation, a spatial beeping sound is emitted from the position of the CNC machine. This audio cue serves as a guidance mechanism, directing the user's attention and helping them orient themselves toward the machine's location.

3.2.7 Safety Mat as a Conventional Industrial Safety System

To conduct a comparative analysis between our AR-based warning system and a commonly used industrial safety system, we integrated a safety mat into our experimental setup, as illustrated in Figure 3.8. The purpose of the safety mat is to activate an alarm when the human operator exits the designated working zone. This is accomplished through the utilization of buzzers and red blinking LED lights, which serve as visual and auditory warning indicators. For the implementation of the safety mat, we incorporated an array of force-sensitive resistors (FSRs) positioned between two rubber mats, each measuring 575×455 mm. These FSRs, specifically the FSR 408 model manufactured by Interlink Electronics, were connected to the analog inputs of a microcontroller, in our case, the Arduino Uno. The microcontroller continuously monitored the forces detected by the FSRs. When a low force reading was registered, indicating that the user had moved away from the device, the microcontroller triggered commands to activate the buzzers and the LED array. This immediate response served to alert both the user and nearby individuals of a potential safety breach. By introducing this safety mat as a reference system, we aimed to assess and compare the effectiveness and usability of our AR-based warning system in ensuring user safety and compliance within industrial environments.

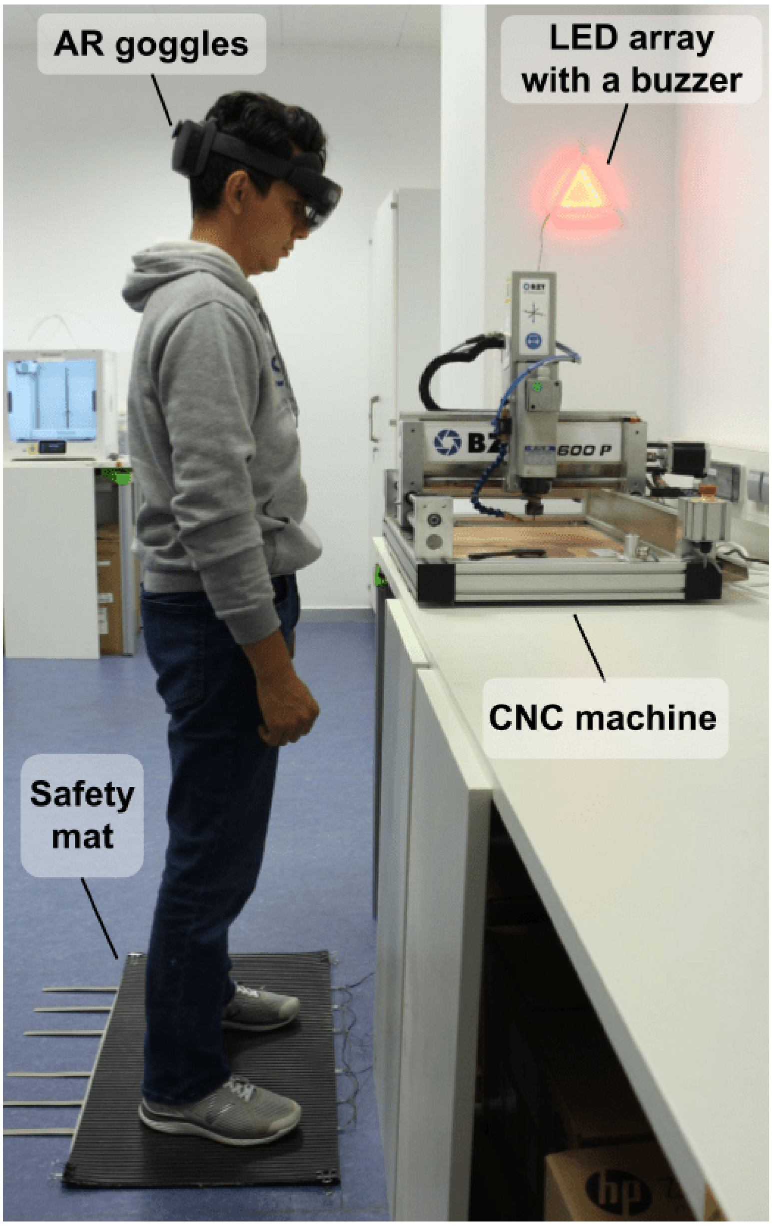


Figure 3.8: Overview of the alternative industrial warning system illustrating the machine, LEDs with a buzzer, and the user with the AR goggles on the safety mat. In this experiment, AR goggles are only used to track the eye movements of the user.

BLANK

Chapter 4

Experiment 1: Objective measurements

To comprehensively assess the effectiveness of the system and the different warning types, we conducted two distinct sets of experiments involving human participants.

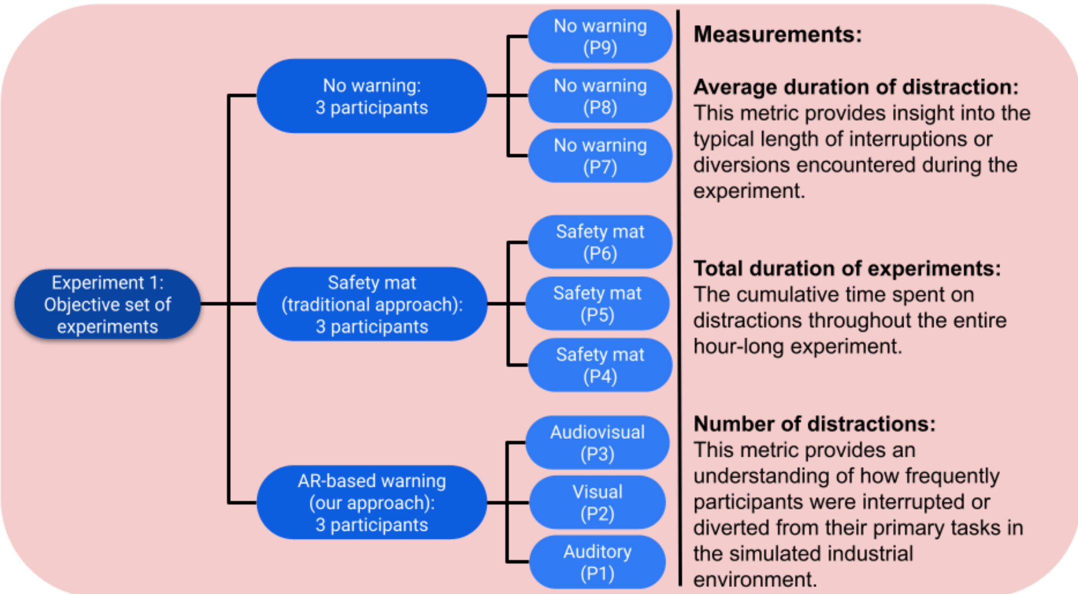


Figure 4.1: Conditions and measurements in Experiment 1 (objective)

In Experiment 1, our focus was on objective measurements. We designed three conditions as the independent variable: the AR condition (our approach), the safety mat condition (traditional approach), and the no warning condition (see Fig.4.1). Within the AR condition, we further divided participants into three subgroups, each experiencing a specific warning modality (auditory, visual, or audiovisual). This allowed us to evaluate the impact of different warning modalities on participant behavior. For each subgroup, a single warning modality was activated during the experiment. Similarly, in the safety mat and no warning conditions, three participants were assigned to each condition. To assess the effectiveness of the warning system, we measured several dependent variables related to distractions. These variables included the number of distractions, the average duration of distractions, and the total duration of distractions experienced by participants. By analyzing these objective measurements, we aimed to gain insights into how different warning conditions influenced participants' attention.

4.1 Rationale for Experiments

In the literature, researchers have explored auditory, visual, and haptic warnings in various domains. Early studies in the aerospace industry involving experienced pilots demonstrated that auditory signals elicit faster response times compared to visual signals [156]. In the context of truck drivers, spatial audio warnings were found to enhance situational awareness [157]. However, a conflicting study in the automotive industry by Dingus et al.[158] contradicted the findings of Reinecke et al.[156], suggesting that visual warnings are more effective in improving driver performance than auditory warnings. Similarly, Yang et al. [159] found that HUD visual warnings conveyed more driver assistance information compared to acoustic warnings. Interestingly, their experiment divided participants into auditory, visual, and audiovisual warning groups, and contrary to expectations, the results indicated the superiority of visual warnings over the other modalities.

In a simulator study by Detmann et al.[160], the combination of auditory and visual warnings led to reduced driver reaction times. Rauterberg[161] conducted an experiment using a simulator application of a manufacturing plant to examine the effects of auditory alarms. The experiment included two conditions: visual alarms and audiovisual alarms. For each machine or robot in the plant, one out of 32 types of sound alarms was selected. The alarms could be triggered simultaneously in the event of a malfunction, designed not to mask each other. Participants were required to enter specific codes for each machine when it encountered a malfunction. The experiment demonstrated a significant performance improvement when using audiovisual alarms compared to visual alarms. However, it should be noted that the experiment had limitations, including a small number of participants (eight) and non-realistic simulated conditions (spatial alarms with two speakers and a simulated manufacturing plant on a desktop computer).

4.2 Experimental Scenario

In the first experiment, individuals were engaged in a one-hour simulated repetitive task on a machine. The task was performed under three conditions: an AR warning system, a safety mat warning system (already employed in industry) for comparative purposes, and without any warning system. Nine participants were involved in the study, with three individuals using the safety mat warning, three without any warning, and three utilizing a specific type of AR warning (auditory, visual, or audiovisual). Additionally, the participants wore AR glasses throughout the experiment, even during the task without AR warning, to track their eye movements. In our experimental protocol, human participants were tasked with executing tool changes in the CNC machine's spindle every 5-7 minutes in an hour-long experiment. Participants were instructed by experimenters to change the tool. The process involved the meticulous removal of the existing tool from the machine head, followed by the precise installation of a new tool. This procedure necessitated the use of two wrenches for secure fixation. Notably, five distinct tools were substituted during the experimental tasks. The primary objective of this experiment was to gather quantitative data on attention by analyzing eye movements and exploring the impact of AR on attention.

The participants visited the experimental site, which was a laboratory equipped with a table-top CNC machine, for a single session. Due to the ongoing COVID-19 pandemic, all participants were required to wear masks throughout the duration of the experiments to ensure safety.

The study protocol received approval from the Institutional Research Ethics Committee of Nazarbayev University, ensuring compliance with ethical standards. Before their participation, all individuals provided informed consent for their involvement in the study.

4.3 Results

The following section presents the quantitative results obtained from Experiment 1. Participants performed simulated tasks for an hour in Experiment 1. The raw data for the objective results of nine participants of the experiment are provided in Table 4.3. Participants performed tasks with three types of AR-based warnings, safety mat, and without warnings in Experiment 1.

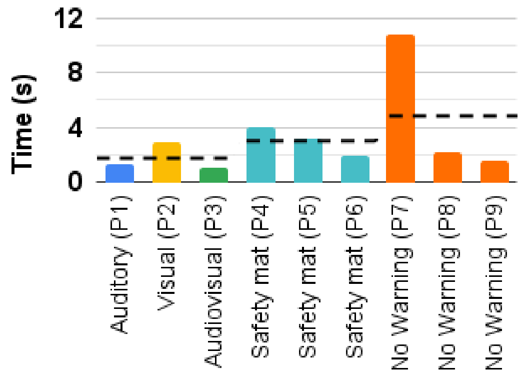
In Experiment 1, participants' performance showed improvement when using AR warning compared to the other two conditions, i.e., safety mat warning and no warning (see Fig. 4.2c). Participants who received the safety mat warning or no warning were more frequently and for longer durations distracted compared to participants who received the AR warning. However, the safety mat warnings allowed participants to perform better than those who received no warning.

The average for the total duration of distractions for all three AR-based warnings (M=162.95 s, SD=90.38) was approximately 6 times shorter than with the safety mat warning (M=1000.82 s, SD=646.49). Participants without warnings (M=1183.44 s, SD=439.50) got distracted slightly for longer periods in total than participants with safety mat warnings (M=1000.82 s, SD=646.49). Participant 3 with audiovisual warning performed tasks with the shortest distraction period (only 75.27 s in total) in Experiment 1. This was roughly 13 times less than with safety mat warnings (M=1000.82 s, SD=646.49) and 16 times less than without warning (M=1183.44s, SD=439.50) conditions on average. Total distraction time with auditory warnings (157.76 s) was less than with visual warnings (255.83 s).

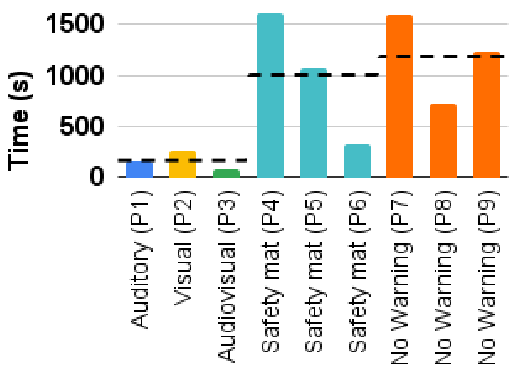
The average duration of distractions of AR-based warnings was quite consistent which was between 1 and 2.8 seconds. The longest average duration was in the distractions with visual warnings among AR-based warnings. This shows that reaction time to visual warnings was slower than to auditory and multimodal (audiovisual) warnings.

Participants using safety mat warnings differed by performance. Participant 6 showed significantly better results than Participants 4 and 5. The number of distractions was 170 and the average duration of distractions was 1.89 s which was slightly worse than AR-based warnings only. However, Participants 4 and 5 got distracted for a significantly longer duration and frequently.

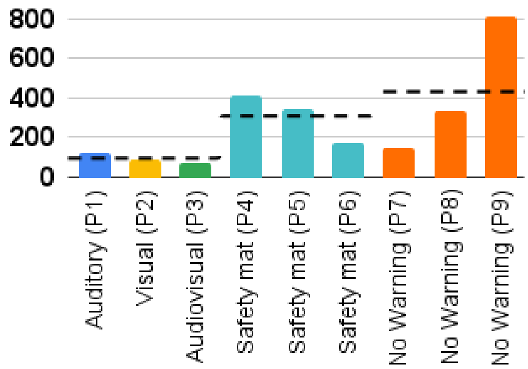
Participants without warning got distracted either for longer periods or frequently. For instance, Participant 7 got distracted for 10.77 s on average in each of the distractions however for 149 times only. At the same time, the distraction of Participant 9 without warning was frequent (810 times) and, however, with short distraction periods (1.525 s on average).



(a) Average duration of distractions



(b) Total duration of distractions



(c) Number of distractions

Figure 4.2: Statistics for different warning types in Experiment 1 for nine participants (P1-P9): a) average duration of distractions, b) total duration of distractions, and c) number of distractions. The black dashed lines indicate the average of the participants in different groups.

Among the various AR warnings, the audiovisual warning proved to be more

Table 4.1: Results of Experiment 1

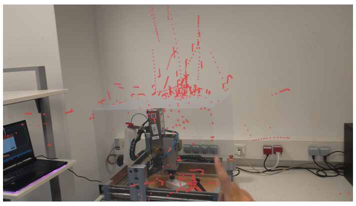
Type of warning	Number of distractions	Total duration of distractions (s)	Average duration of distractions (s)
Auditory (P1)	123	157.75	1.28
Visual (P2)	89	255.82	2.87
Audiovisual (P3)	75	75.27	1.00
Safety mat (P4)	407	1610.39	3.95
Safety mat (P5)	347	1069.22	3.08
Safety mat (P6)	170	322.84	1.89
No Warning (P7)	149	1594.74	10.77
No Warning (P8)	335	720.32	2.15
No Warning (P9)	810	1235.25	1.52

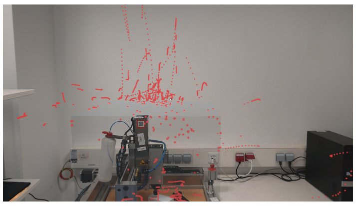
effective in helping participants maintain focus on the task at hand. Participants experienced fewer distractions with visual warnings compared to auditory warnings (refer to Fig. 4.2a), but, on average, the distractions lasted for a longer duration (refer to Fig. 4.2b). To visually analyze the participants' performance, the 2D gaze points were plotted in 3D space (refer to Fig. 4.3). This visualization clearly demonstrates that participants were more likely to concentrate on the machine when assisted by the AR warning system. Safety mat warnings caused fewer distractions than without warning, on average. However, Participants 4 and 5, who utilized a safety mat, experienced numerous distractions, leading to outcomes comparable to the no-warning condition in the experiment. The superior performance of the AR warning system in comparison to the traditional safety mat can be attributed to the AR system's ability to directly track the user's attention using AI-based object recognition and gaze tracking. On the contrary, the safety mat only verifies the presence of the human operator near the machine and does not monitor their attention.

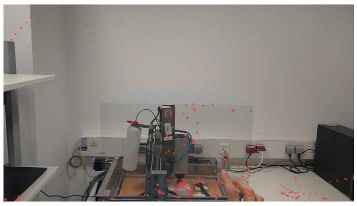
4.3.1 Discussion

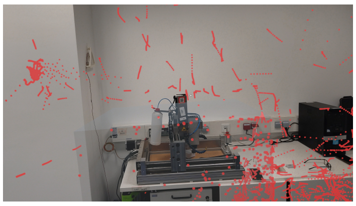
Our study examined the impact of different warning modes, as well as the absence of warnings, and the number, average duration, and total duration of distractions experienced by participants operating an industrial machine. When audiovisual warnings were employed, participants diverted their gaze from the machine less frequently and for shorter periods compared to other warning modes.

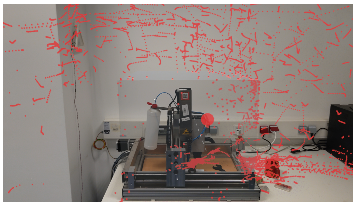
During visual warnings, participants looked away from the machine less frequently than during auditory warnings. Participants responded more quickly after receiving auditory warnings than after visual warnings, consistent with the study of Jain et al. [97]. AR-based warnings preceded safety mat warnings in the duration of distractions. This indicated the effectiveness of our system over alternative solutions that are used productively in business. These results could provide future perspectives for research in

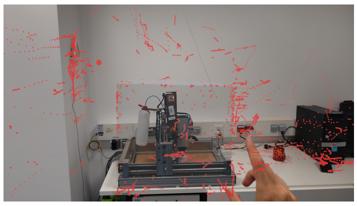

(a) AR with auditory (P1)

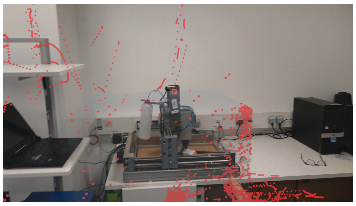

(b) AR with visual (P2)

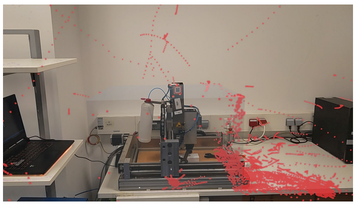

(c) AR with audiovisual (P3)


(d) Safety mat (P4)


(e) Safety mat (P5)


(f) Safety mat (P6)


(g) No warning (P7)


(h) No warning (P8)

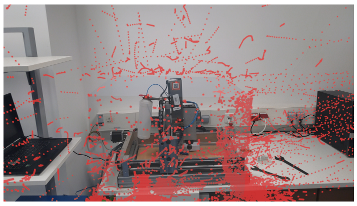

(i) No warning (P9)

Figure 4.3: The gaze points of participants overlaid on the real world environment in Experiment 1

industrial safety. Integrating cutting-edge technologies such as eye tracking, AR, and AI could make warning systems more efficient in different domains.

The warning system implemented in our study facilitated increased attention during the work process, similar to the findings reported by Brown et al. [124]. Increased attention is imperative in many domains including automotive and manufacturing industries. AR-based warning systems could be adapted to other industries in the future.

Experiment 1 had limitations regarding the number and contingent of participants. Firstly, the number of female participants (2) was less than that of male participants (7). Secondly, the number of participants (9) was limited since each session lasted for an hour. Thirdly, participants consisted of university faculty or students. Further studies could be investigated with gender-balanced groups of participants from different domains such as automotive and manufacturing industries. The number of participants can be increased to decrease the subjectivity of the study.

On the other hand, the study had advantages concerning the study design. The experiments were conducted in a between-subjects design. Different participants were assigned to different types of warning conditions. This design has several benefits over within-subjects design where the same participant performs tasks with all conditions. For instance, it reduces learning effects. The same participant performing the tasks with other conditions could have become more efficient in the next tasks. Also, conducting multiple long sessions of experiments could make participants tired or bored in our case.

Experiment 1 was conducted in a laboratory equipped with table-top CNC isolated from all kinds of hazards. The real-world case studies in industrial environments could enable us to evaluate the applicability and effectiveness of AR-based warning systems further.

BLANK

Chapter 5

Experiment 2: Subjective measurements

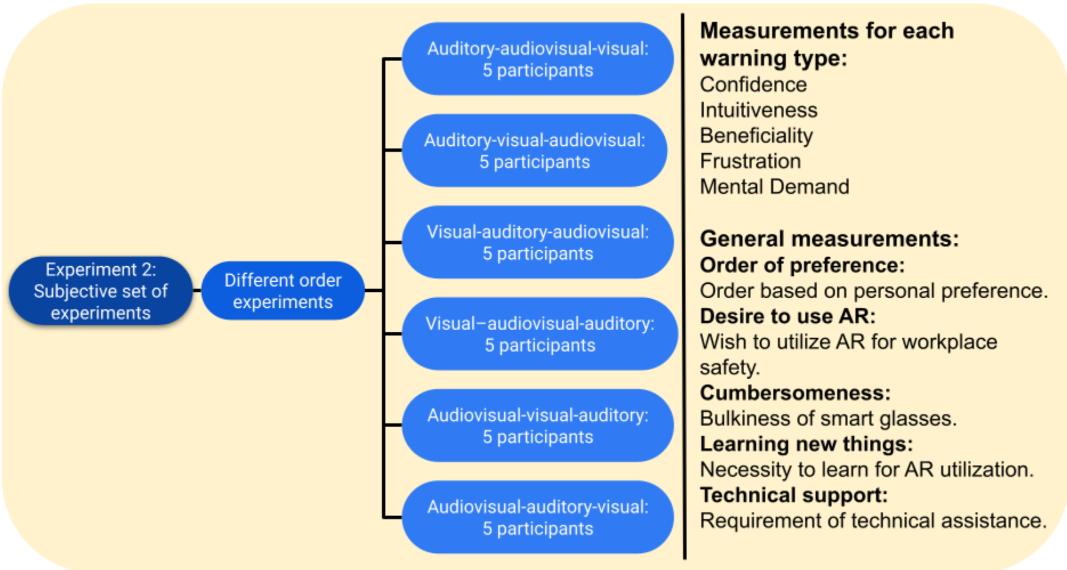


Figure 5.1: Conditions and measurements in Experiment 2 (subjective experiment)

Given the objective measurements, we aimed to compare the user experience and mental load associated with auditory, visual, and audiovisual modalities of an AR warning system for industrial safety. By conducting this comparative analysis, we aimed to contribute to the understanding of which warning modality is most effective in this specific context.

Experiment 2 encompassed the initialization of the system and the execution of three distinct tasks: 1) visual warning, 2) auditory warning, and 3) audiovisual warning. Initially, participants received instructions regarding the various warning modalities and the operation of the AR goggles. Subsequently, participants wore AR goggles and underwent eye calibration to ensure accurate eye tracking. They also adjusted the video brightness and auditory warning intensity to their preference.

After becoming acquainted with the device, the experimental tasks commenced. During system initialization, participants were positioned near the CNC machine with their heads initially turned to the left, and then instructed to turn their heads towards the machine. Upon detection of the machine, a cylindrical hologram was superimposed onto it. Participants confirmed the machine's position by performing a finger tap if

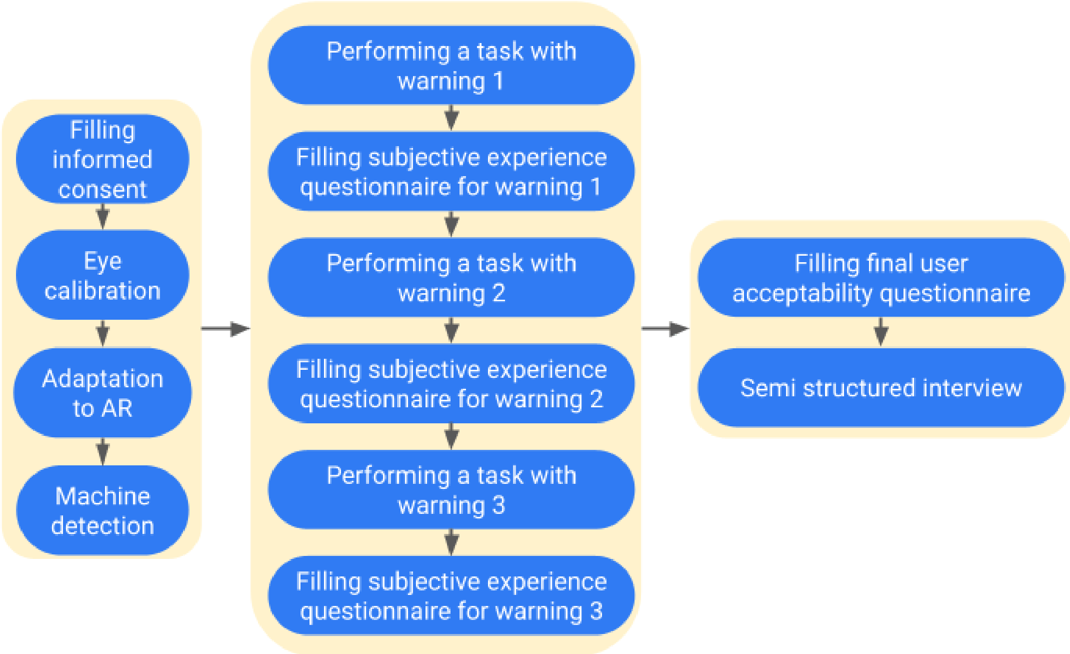


Figure 5.2: The scenario of Experiment 2 (subjective experiment)

it was correctly indicated. Subsequently, the hologram remained fixed on the CNC machine. The warning systems varied depending on the type of warning used. In each task, participants were instructed to perform a sequence of visual orientation tasks in relation to the CNC machine. Specifically, they were directed to sequentially look right and return their gaze to the CNC machine, then left and back to the CNC machine, upward and back to the CNC machine, and downward and back to the CNC machine. Subsequently, participants were required to repeat these gaze movements without physically turning their heads, but instead, moving only their eyes. Additionally, participants were tasked with executing a complete 360-degree rotation, starting from facing the CNC machine and concluding upon facing it again. Each task had a duration of approximately five minutes.

Each participant performed tasks under varying conditions, including auditory, visual, and audiovisual warnings with different orders. The warning stimuli were initiated when participants averted their gaze from the CNC machine and ceased once participants redirected their visual attention toward it. After each task, participants were asked to fill out the subjective experience questionnaire (see Table 5.1). After performing tasks with all three types of warnings, they were asked to fill out the final user acceptability questionnaire (see Table 5.2). Finally, semi-structured interviews were conducted with participants.

DeMelo et al. [140] examined the cognitive load experienced by participants in a simulated AR military environment, both with and without the presence of virtual assistants. The cognitive load was assessed using the NASA Task Load Index (NASA TLX) [162]. The AR framework included embodied virtual assistants and auditory

Table 5.1: The subjective experience questionnaire

ID	Items
SUS9	I felt very confident using the warning system.
SUS3	I found the warning intuitive.
CL1	The warning task was mentally demanding.
UE1	I found the system beneficial for the working process.
CL6	I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system.

Table 5.2: The final user acceptability questionnaire

ID	Items
UE2	Which warning do you prefer? Sort them in order of preference.
SUS1	I would like to use the AR smart glasses for safety in the working environment.
SUS9	The smart glasses are cumbersome to use.
SUS10	I needed to learn a lot of things before I could get going with this system.
SUS4	I thought that I would need the support of a technical person to be able to use this system.

assistants. The findings indicated that participants experienced a lower cognitive load when they received assistance from an auditory assistant.

In [141], the Chromaglasses AR camera system was developed specifically for individuals with color vision deficiency. The usability of the system was evaluated using the System Usability Scale (SUS) questionnaire and the NASA TLX to measure participants' cognitive load. The SUS questionnaire included the first question to assess usability, while the NASA TLX incorporated the first three questions to analyze cognitive load. Additionally, a custom question was included to evaluate the realism of the experience.

Erickson et al. [142] developed an AR application for the purpose of sharing gaze rays with collaborators. To assess subjective measures in the experiment, the authors designed a questionnaire incorporating five questions from the NASA TLX and SUS [163]. The questionnaire encompassed questions from NASA TLX regarding the cognitive workload, perceived difficulty, and level of annoyance experienced during personal interactions with AR system. Additionally, it included questions assessing the ease of use (from SUS) and perceived realism (custom question) of AR system.

Our research involved conducting various tasks using different warning methods, and subsequently, participants were requested to complete Likert scale questionnaires consisting of five points. The post-experiment questionnaires encompassed items derived from the NASA TLX and SUS questionnaires in order to evaluate subjective experiences, as outlined in Table 5.1 and Table 5.2. Irrespective of the warning system employed, the questionnaires contained identical sets of five questions. The NASA TLX questions focused on assessing participants' mental demand (CL1) and frustration

(CL6) while utilizing the system, while the SUS questionnaire gauged user confidence (SUS9) and system intuitiveness (SUS3). To assess the effectiveness of the warning, we introduced a custom question (UE1). Following the completion of all tasks and the post-experiment questionnaires, participants responded to a final user acceptability questionnaire, which consisted of general inquiries as well as a ranking question (UE2) concerning the warning modalities. The general questions encompassed items from the SUS questionnaire, which evaluated participants' inclination to use the system (SUS1), ease of use (SUS9), need for acquiring new knowledge (SUS10), and technical support requirements (SUS4). These questions aimed to assess participants' readiness to utilize the system in a professional setting, the level of inconvenience caused by the AR glasses, whether participants needed to acquire new knowledge to operate the system, and if they required technical support. Lastly, we conducted semi-structured interviews with the participants to gather more qualitative insights and perspectives.

We employed the Shapiro-Wilcoxon test with a significance level of 5% to test for the normality of the ratings obtained from the Likert scale questionnaire in Experiment 2. The test indicated that the ratings were not normally distributed. Consequently, we opted for a Friedman test (α= .05), a non-parametric ANOVA test for repeated measures, to analyze the results instead of a parametric ANOVA.

Specifically, in Experiment 2, we tested three hypotheses about the participants' perception of the system:

- Do the participants find the AR warning system beneficial? (H1)
- Do the participants prefer auditory warnings to visual warnings? (H2)
- Is the audiovisual warning system the preferred option? (H3)

In Experiment 2, we recruited a total of 30 participants, comprising 15 males and 15 females. The participants had an average age of 24.9 years (SD = 5.78). Among them, 17 participants had previous experience using AR/VR systems, while 14 participants had prior familiarity with operating industrial machinery. They performed the three tasks in a randomized order. During Experiment 2, five participants wore glasses, while two participants opted for contact lenses to accommodate their visual needs.

The participants visited the experimental site, which was a laboratory equipped with a table-top CNC machine, for a single session. Due to the ongoing COVID-19 pandemic, all participants were required to wear masks throughout the duration of the experiments to ensure safety.

The study protocol received approval from the Institutional Research Ethics Committee of Nazarbayev University, ensuring compliance with ethical standards. Before their participation, all individuals provided informed consent for their involvement in the study.

5.1 Results

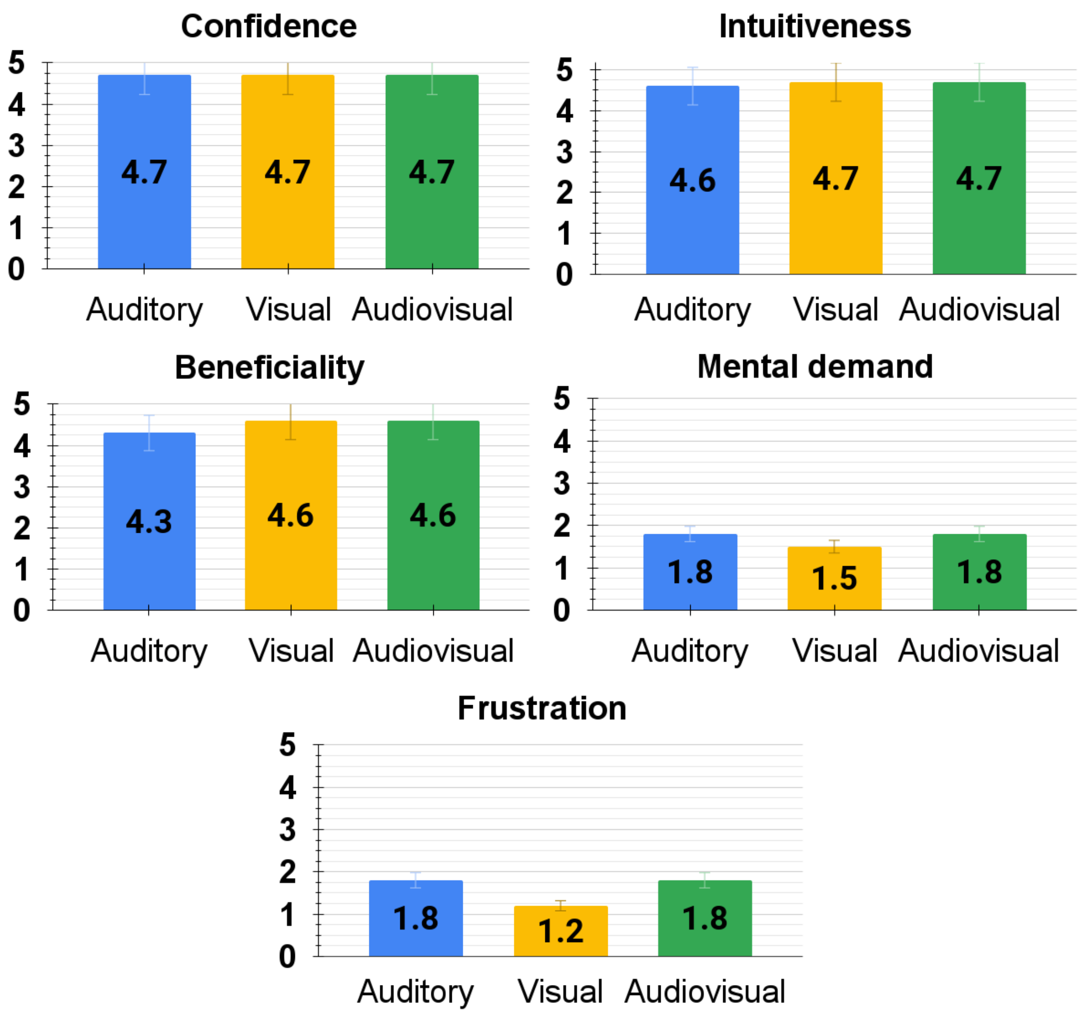


Figure 5.3: Subjective ratings (ranked between 1 and 5) for each of the three cues: confidence (top left), intuitiveness (top center), beneficiality (top left), frustration (bottom left), and mental demand (bottom right)

In this section, we will delve into the findings of Experiment 2, which involves the analysis of subjective questionnaires completed by the participants after the experiments. The questionnaires assessed factors such as participants' confidence, frustration levels, mental demand, system intuitiveness, and ratings regarding the benefits of the different warning types. Figure 5.3 displays the graphical representations of these measures.

The participants' levels of confidence in using the different types of warnings were found to be nearly equal. However, they perceived the audiovisual and visual warnings to be more intuitive and beneficial to their work processes. In contrast, the visual warnings elicited less frustration and mental demand compared to the other types of warnings. The participants considered auditory warnings to be the least intuitive and beneficial.

Statistical analysis using Friedman test revealed no significant differences between

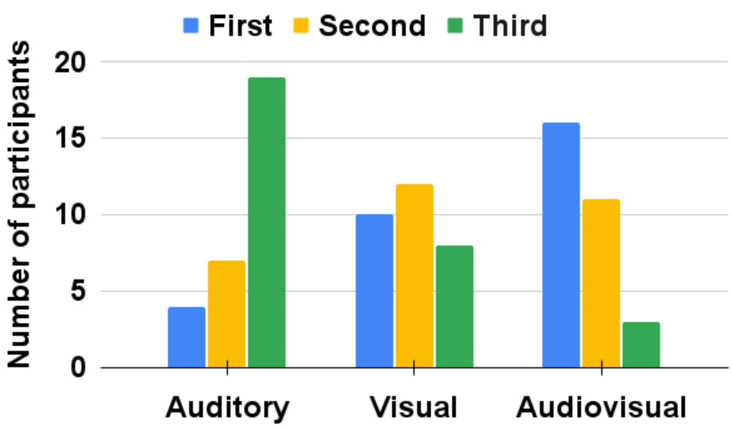


Figure 5.4: Participants' preference rankings of warning modalities

warning types in terms of confidence (χ^2 (2) = 0.317, p = .853), intuitiveness (χ^2 (2) = 0.217, p = .897), mental demand (χ^2 (2) = 2.117, p = .347), and beneficiality (χ^2 (2) = 1.55, p = .46). However, there was a significant difference in frustration (χ^2 (2) = 6.717, p = .035) among the warning types.

Participants were asked to rank warning types by placing them in order of their preference in the first question of the general acceptability questionnaire (see Table 5.2). We calculated the ranking score by summing ratings multiplied by weights for each choice in a specific type of warning. For example, the first choice of participants weighed three points, whereas one point was given to the least preferred. Ranking scores for auditory, visual, and audiovisual warnings were 45, 62, and 73, respectively. Participants preferred the audiovisual warning (confirming H3) to others, while the auditory warning was considered the least preferred (contradicting H2) (refer to Fig. 5.4).

The findings from the final questionnaire (SUS1), which included general inquiries about the system, indicated a high level of acceptance for the AR glasses as a safety measure in the workplace (M = 4.03, SD = 1.06), thus supporting hypothesis H1. Participants perceived AR headsets to be moderately burdensome during the working process (M = 2.5, SD = 1.04). They expressed a neutral stance on the necessity of technical support (M = 2.25, SD = 1.26) and the requirement to learn new skills in order to use the system (M = 2.83, SD = 1.38).

5.1.1 Discussion

Upon completing Experiment 2, semi-structured interviews were conducted. Participant 1 highlighted that the dark visor of the AR glasses could impede operators from adapting to the working process. Participant 2 observed that the auditory warning lacked sufficient information but induced a sense of urgency, while the visual warning conveyed understandable information but did not elicit the same level of urgency.

Consequently, the participant suggested that integrating both modalities would enhance both situational awareness and urgency. The transparency of the cylindrical hologram generated differing opinions among participants, with some suggesting increased transparency and others preferring greater opacity. Participant 6 specifically mentioned the need to reduce the size of the hologram. To address this issue, a software feature could be implemented to allow adjustment of both transparency and size.

Participant 9 recommended replacing the beeping sound with spatial voice commands, such as "Look at me." Notably, the same voice command, "Look at me," was employed in the CARA cognitive assistant for guiding individuals with visual impairments [164]. Participant 16 suggested incorporating a melodious tone into the audio warning, as continuous warnings were perceived as intrusive in a previous study [165]. Audio configurations could be added to accommodate individual usability preferences. Participants 12 and 30 expressed the opinion that the arrow hologram in the visual warning should be red to effectively alert the operator. They cited examples of red visual alarms being employed as high-urgency alerts in aircraft cockpit systems [166]. Participant 14 requested the use of a path instead of an arrow hologram. Additionally, some participants recommended displaying a white hologram to cover the user's field of view when the operator is not looking at the machine. Incorporating various types of visual aids could enhance the intuitiveness of alerts. Many participants noted that the combined audiovisual warning was both rapid and informative.

We did not find a significant difference in the mental demand experienced by participants when comparing auditory, visual, and audiovisual warnings, as also pointed out by Detmann et al. [160]. However, visual warnings were less frustrating to participants compared to auditory and audiovisual warnings.

As mentioned by Li et al.[167], the AR glasses used in our study were perceived as moderately bulky, which can lead to fatigue during prolonged use. The acceptance of our system aligned with the high levels reported in [160]. Directional warnings, proposed by authors [160], were found to be very supportive by participants for the advanced drive assistance system (ADAS).

Several limitations are associated with this experiment. Firstly, the participant pool was limited to university students and faculty members. It is recommended for future studies to include participants from diverse age groups and occupations, such as industrial operators, to enhance the generalizability of the findings.

Secondly, the auditory warning mode utilized a continuous beeping sound. Future research could explore the effectiveness of employing sounds with different frequencies, intensities, or even verbal commands within the auditory warnings. As reported by [97], the reaction time of participants to auditory cues was less than to visual cues. Using sound cues as warning systems can be highly effective in rapidly alerting people to potential dangers or emergencies.

Furthermore, the study focused on visual and auditory warning modalities, but it would be valuable to investigate the potential benefits of incorporating other modality cues, such as tactile cues, into the warning system. For instance, according to [124], participants demonstrated faster response times with haptic warnings than with auditory warnings. This could provide a more comprehensive understanding of how different cues contribute to situational awareness and response.

Chapter 6

Thermal Perception Using AR for Industrial Safety

To the best of our knowledge, AR goggles integrated with thermal sensors have not yet been utilized to improve industrial safety. We aimed to fill this research gap and investigate the potential of AR goggles equipped with thermal sensing for industrial safety in this study. Specifically, we present an AR-based safety system that seamlessly informs an industrial operator about potentially dangerous hot objects in their workspace. Our main contributions are as follows:

- We developed an untethered embedded system for a thermal sensing array (Flir Lepton 2.5) and ergonomically integrated it into AR goggles (Microsoft Hololens 2), as shown in Fig. 6.2.
- We calibrated the homography matrix between the thermal array and the visual camera of the AR goggles and achieved an average corner error of the homography estimation of 7.6 pixels. This allowed us to display the thermal holograms with high spatial accuracy.
- In real-world experiments conducted in a simulated industrial environment with a table-top CNC machine where a heated bed (220×220mm) was placed at different locations, we demonstrated that the operator receives real-time visual feedback on hot objects with injury risk.

6.1 Thermal Imaging

A thermal camera is a non-contact device that captures IR radiation, representing the temperature distribution in space as a visible image. There is a wide range of thermal cameras available, varying in form factors such as handheld (pistol-grip), monocular, phone-connected, and thermal imaging smartphones.

For our study, we chose to integrate the FLIR Lepton 2.5 thermal array into our AR system due to its compact size (8.5×11.7×5.6 mm). To accomplish this, we developed a 3D-printed module that housed both the thermal camera and a single-board computer. This module was connected to the AR goggles, as depicted in Fig. 6.2.

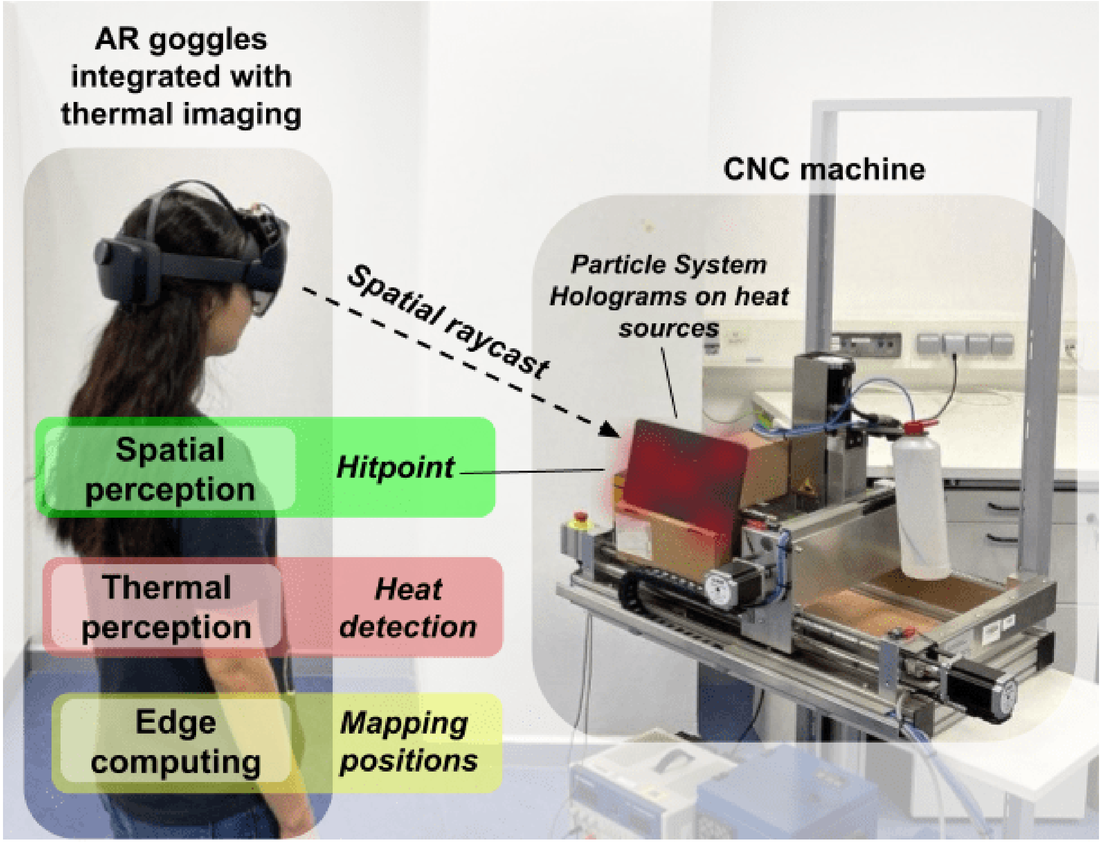


Figure 6.1: Overview of our industrial safety system illustrating the user wearing AR goggles integrated with thermal sensing, holographic temperature information embedding, and a CNC machine with a simulated hot part.

In our setup, the thermal camera served as a first-person view camera, enabling the system to detect and perceive heat sources within the user's field of view. The thermal camera had a resolution of 80×60 pixels and a horizontal field of view (HFOV) of 50°, whereas the RGB cameras of the Microsoft Hololens 2 had an HFOV of 64.69°. The thermal camera operated within the wavelength range of 8 to 14 μm and provided temperature values as pixel data.

To establish connectivity, the thermal camera was linked to a single-board computer (Raspberry Pi 3) via the serial peripheral interface (SPI). After undergoing preprocessing, the acquired data was transmitted to the AR goggles over a WiFi connection facilitated by the single-board computer.

The acquired raw grayscale data, consisting of 14-bit values, represented temperature measurements. These values were subsequently converted into corresponding temperature values (e.g., 7390 corresponding to 23° C and 8590 corresponding to 70° C). To identify the hot pixels with temperatures exceeding a threshold of 40° C, their coordinates, and temperature values were stored in an array for further analysis.

To position the holograms accurately in the 3D space, we employed the Raycast API from Unity's Physics library. This API projects an imaginary ray originating from the pixel position in the image plane and detects the precise intersection point in the world

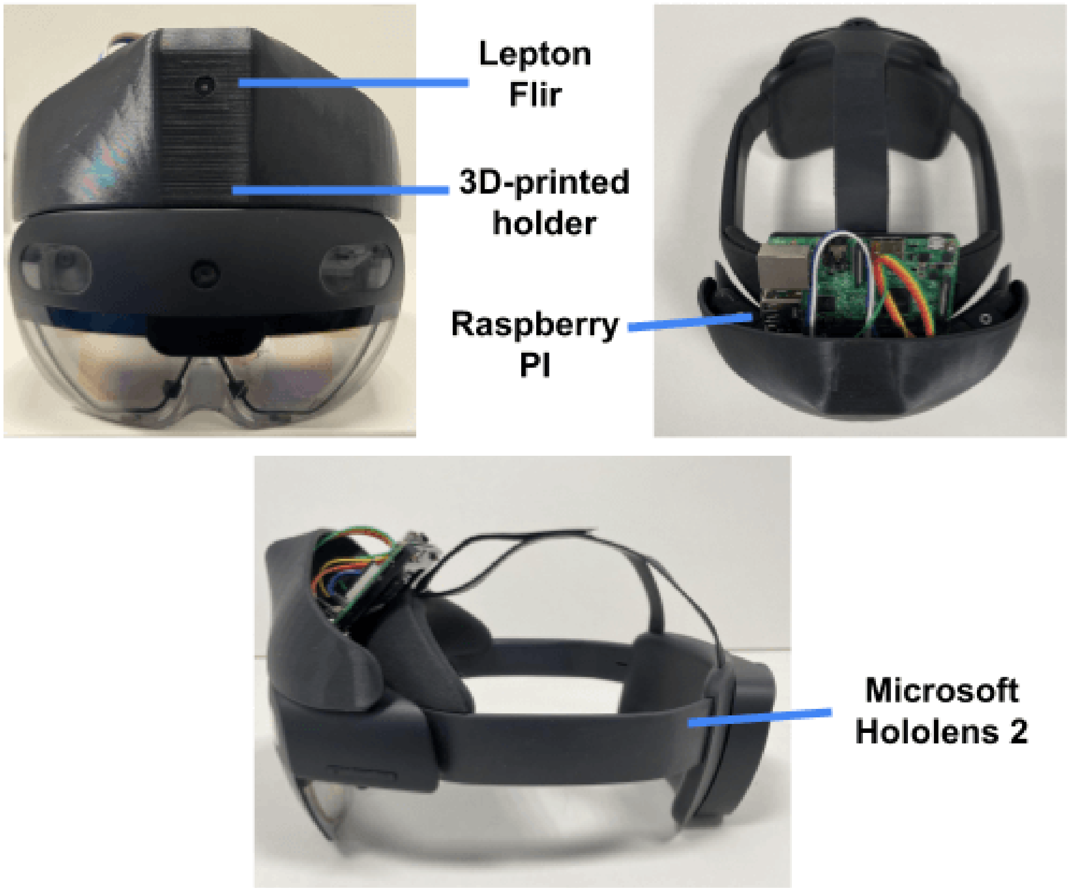


Figure 6.2: Side, top, and frontal views of the AR system with thermal imaging

frame. An increasing number of raycasts causes a computational burden on AR glasses even though it would yield a smoother result. AR glasses have limited computational capabilities in holographic processing units. Depending on our scenario of usage where the raycasts are processed every frame, we had to trade off the number of raycasts against the smoothness of meshes. Therefore, to optimize computational efficiency for the AR glasses, a random selection was made from the array, involving the extraction of five pixels. By limiting the number of pixels, the overall computation time for raycasting against spatial surfaces of the AR glasses was reduced. Before transmitting these selected pixels to the AR glasses, it was necessary to map their coordinates onto the AR image plane, ensuring proper alignment and integration with the overall AR display.

6.2 Mapping the Thermal Image to the Holographic Space

To align the pixel positions in the thermal array, where the temperature values exceeded the threshold, with the AR image plane, a mapping process was carried out. This

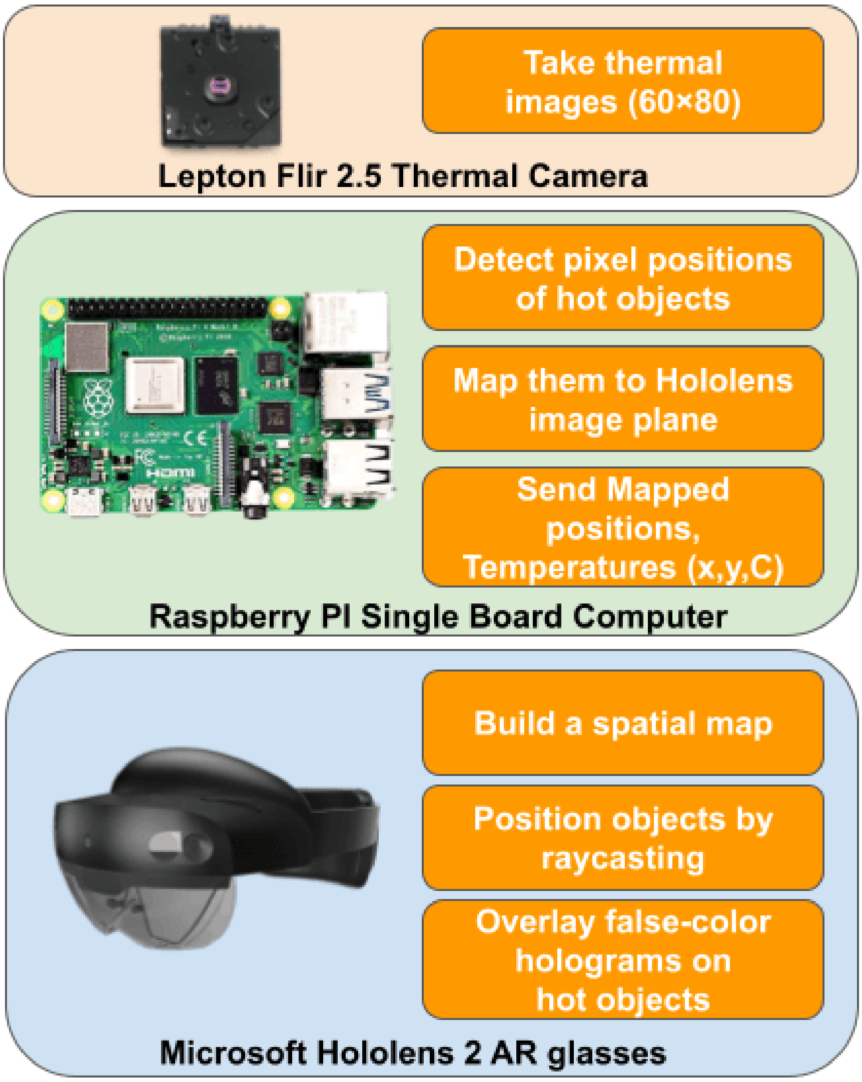


Figure 6.3: Hardware and software architecture of the system.

mapping required the determination of a homography matrix between two camera frames through calibration, as depicted in Figure 6.4.

Homography refers to the transformation between two image planes when both planes can be projected onto the same planar surface in space [168]. By calculating the homography matrix H, which is a 3×3 matrix, the destination image points can be obtained by multiplying the source image point with the matrix H:

$$\begin{bmatrix} x_h \\ y_h \\ 1 \end{bmatrix} = H \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} \tag{6.1}$$

where x_h and y_h are the point-pixel coordinates in the Hololens 2 camera images and x_t and y_t are the corresponding point-pixel coordinates in the thermal array images. This equation can be transformed into:

$$Ah = 0 \tag{6.2}$$

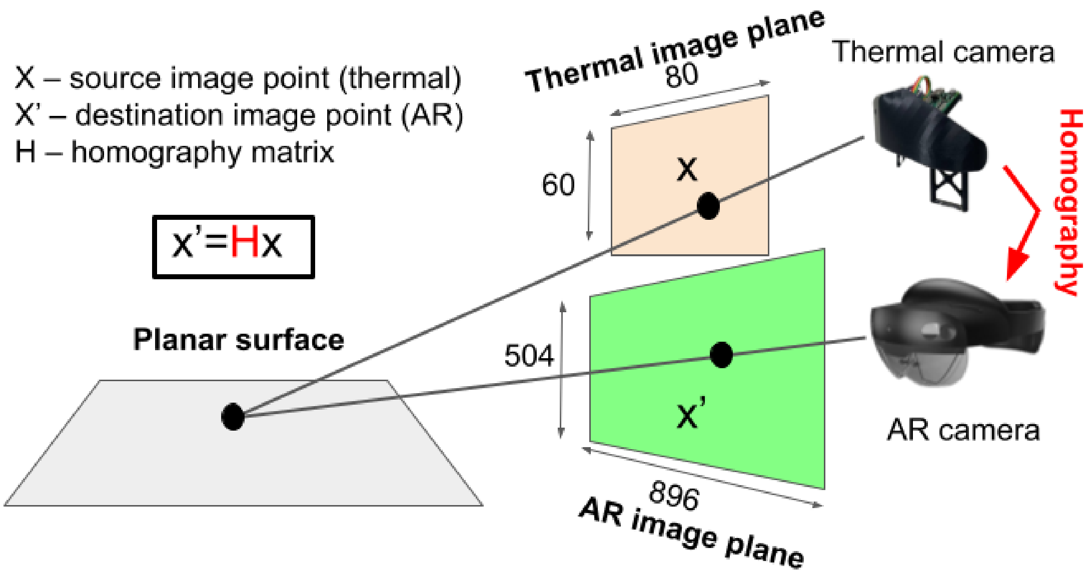


Figure 6.4: Image planes of thermal and Hololens 2 RGB cameras and a planar surface.

where

$$A = \begin{bmatrix} x_t & y_t & 1 & 0 & 0 & 0 & -x_h x_t & -x_h y_t & -x_h \\ 0 & 0 & 0 & x_t & y_t & 1 & -y_h x_t & -y_h y_t & -y_h \end{bmatrix} \tag{6.3}$$

and

$$h = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} \tag{6.4}$$

$$\begin{bmatrix} x_t & y_t & 1 & 0 & 0 & 0 & -x_h x_t & -x_h y_t & -x_h \\ 0 & 0 & 0 & x_t & y_t & 1 & -y_h x_t & -y_h y_t & -y_h \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{6.5}$$

To determine the homography matrix, it is necessary to identify a minimum of four corresponding points in each of the image planes. Because the homography matrix

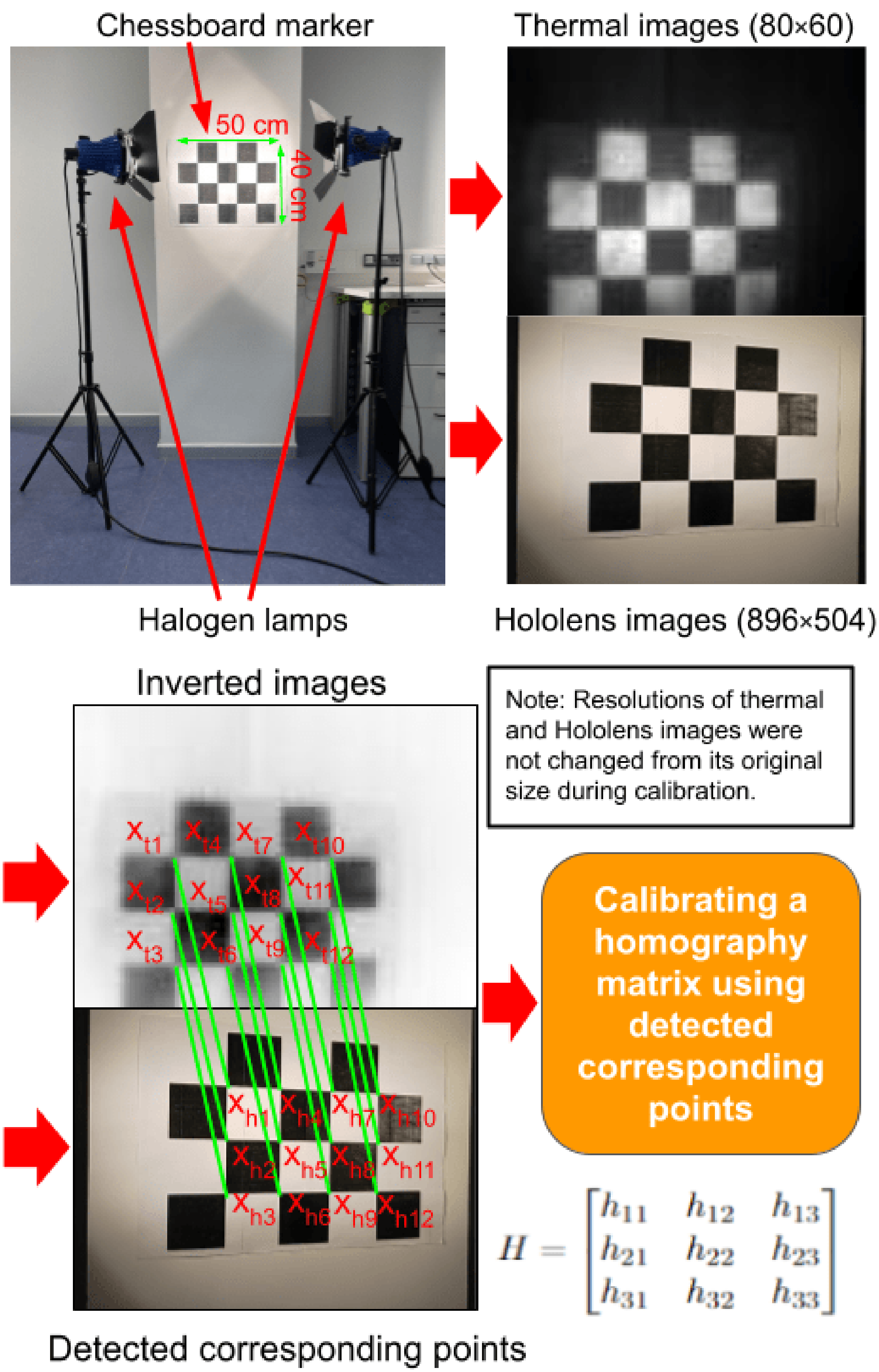


Figure 6.5: The calibration process of the homography matrix.

has 8 degrees of freedom. Each of our points contains x and y, consequently, 4 points make 8 equations. To increase the accuracy of the homography matrix, more than four points can be utilized and the least squares equations can be derived to solve the overdetermined system. Therefore, we utilized 12 images from each thermal camera and RGB camera of Hololens 2. Checkerboards' 12 points (3×4) from each image were extracted for calculation of the homography matrix.

The equation presented in (6.5) is particularly useful when dealing with multiple pairs of corresponding points. Specifically, when there are n pairs of corresponding points, (6.5) can be expressed as:

$$\begin{bmatrix} x_{t1} & y_{t1} & \cdots & -x_{h1}x_{t1} & -x_{h1}y_{t1} & -x_{h1} \\ 0 & 0 & \cdots & -y_{h1}x_{t1} & -y_{h1}y_{t1} & -y_{h1} \\ x_{t2} & y_{t2} & \cdots & -x_{h2}x_{t2} & -x_{h2}y_{t2} & -x_{h2} \\ 0 & 0 & \cdots & -y_{h2}x_{t2} & -y_{h2}y_{t2} & -y_{h2} \\ & & \vdots & & & \\ x_{tn} & y_{tn} & \cdots & -x_{hn}x_{tn} & -x_{hn}y_{tn} & -x_{hn} \\ 0 & 0 & \cdots & -y_{hn}x_{tn} & -y_{hn}y_{tn} & -y_{hn} \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{6.6}$$

Corresponding points can be identified using various techniques, such as checkerboard patterns (Zhang's method [23]) or feature matching algorithms like speeded-up robust features (SURF) and scale-invariant feature transform (SIFT). These methods have not only been employed to calculate homography matrices between two RGB cameras but have also been utilized for determining homography matrices between cameras from different domains.

For example, Baek et al. [169] successfully calibrated a homography matrix for stereo gamma radiation cameras using these methods. In [170], an aluminum calibration board was utilized to calibrate IR, ultraviolet, and RGB cameras, resulting in the determination of their respective homography matrices.

To calibrate the homography matrix, we employed Zhang's method with a checkerboard. Checkerboard is a board of black and white square patterns that are used to calibrate the sensors in a camera. The calibration process involved heating the checkerboard pattern, printed on paper, using two halogen lamps. This generated a temperature difference between the white and black sections of the pattern, as illustrated in Fig. 6.5. To detect the corners of the pattern, the captured image was inverted. By obtaining the set of corresponding corners from both camera images, we were able to extract the homography matrix using the OpenCV library in Python. The resulting homography matrix played a crucial role in mapping the plane of the thermal image to the plane of the Hololens image. This mapping process facilitated the alignment and integration of the thermal image within the AR display of Hololens.

6.3 Edge Computing

Edge computing is a technology that reduces latency and enhances performance by placing computing devices in close proximity to the data source. It finds widespread application in various AR image-processing tasks. An example of its utilization can be found in [171], where an embedded DL system was employed for AR in firefighting applications. This system incorporated RGB, depth, and IR camera streams. The acquired data from the cameras was processed by the Nvidia Jetson TX2, which employed the Faster R-CNN object detection algorithm. The processed data was subsequently transmitted to Microsoft Hololens 1, where it was visualized on the 2D plane for the user.

To facilitate the detection of hot object positions and temperatures in thermal images, as well as the mapping of these positions to the AR image plane and data transmission to the AR goggles, we selected a Raspberry PI 3 Model B single-board computer. This computer is equipped with a Quad Core 1.2 GHz Broadcom BCM2837 64-bit CPU and 1 GB of RAM. For the purpose of achieving untethered operation, the Raspberry PI 3 was powered by a power bank (22.5 W and 10000 mAh)

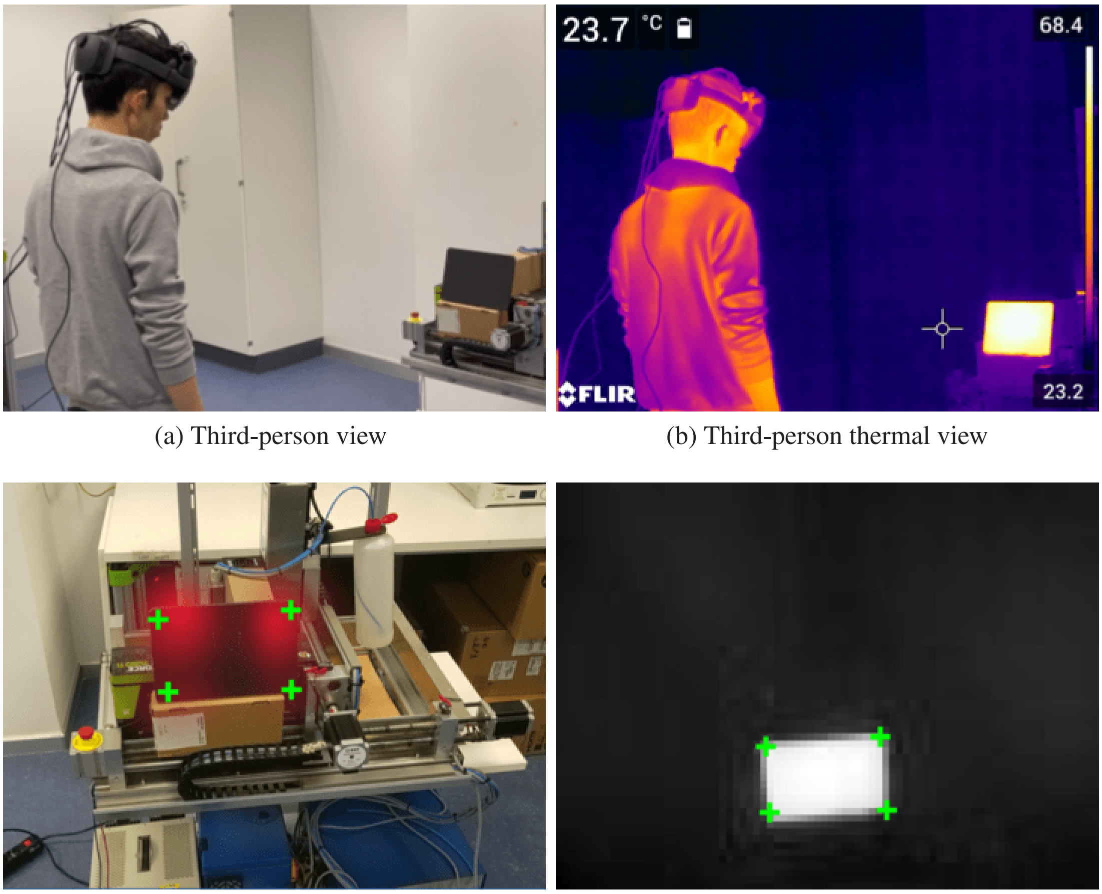
To acquire thermal images, a Python script leveraging the PyLepton library (referenced as [172]) was utilized. The camera captured thermal images at a rate of 9 frames per second. Subsequently, the acquired thermal images were processed using the OpenCV library [173]. These processing steps enabled the extraction of relevant information and facilitated the subsequent tasks of object position and temperature detection.

6.4 Creating Thermal Holograms on the AR Scene

Cutting-edge AR glasses possess the capability to generate a three-dimensional (3D) representation of the physical environment, facilitating precise placement of holograms. An illustration of this functionality can be found in [174], where an AR application utilized the spatial mapping feature of the AR glasses to detect objects and accurately position name annotations over them.

To develop our AR application, we utilized Unity 2020.3.24 and MRTK version 2.8.2.0. The application was deployed to Microsoft Hololens 2 using Visual Studio 2019. The AR system received five positions with temperature values above a predetermined threshold, randomly selected from the embedded system, and mapped them to the Hololens plane.

Due to the computational limitations of the AR glasses, only a limited number of points were received to be processed. The Microsoft Hololens glasses were utilized to perceive the user's spatial environment and overlay false-color holograms over



(a) Third-person view (b) Third-person thermal view

(c) First-person Hololens view (d) First-person Lepton Flir view

Figure 6.6: Views from different perspectives and spectra (Green "+" signs were added to indicate mapped points in first-person view subfigures).

Figure 6.7: The color gradient of false-color particle systems shifting across a range of 40° C to 70° C

areas identified as potentially hazardous due to elevated temperatures. The false-color representation was determined based on a color gradient that gradually transitioned from yellow to red, corresponding to increasing temperatures (see Fig. 6.7). Particle systems were placed at the positions where the raycast hit the physical objects in the environment, as depicted in Figure 6.6. Particle systems are used in computer graphics to form a "fuzzy" objects such as fire, smoke, or fireworks using small 2D sprites or 3D models. These particles had a semi-transparent spherical shape with a radius of 10 cm.

BLANK

Chapter 7

Experiment and Results

The purpose of our safety system was to provide industrial operators with alerts regarding hot objects in their work environment. These hot objects were augmented with holograms that exhibited a color gradient ranging from yellow to red based on their temperature, as depicted in Figure 7.1. We followed the guidance of Lloyd-Smith and Mendelssohn [175], who indicated that the pain threshold for hot objects begins at a temperature of 43°C. Consequently, objects with temperatures exceeding 40°C were overlaid with yellow holograms, while those surpassing 70°C were overlaid with red holograms, signifying a high risk of severe injury. This approach allowed us to effectively communicate temperature information to the user.

The calibration process, which involved utilizing halogen lamps and heated checkerboard patterns, yielded precise outcomes. By detecting the corresponding corners in each camera plane, we were able to calculate the homography matrix (as can be seen in Eq. 7.1) that mapped the thermal image plane to the image plane of the AR glasses. The mean error of the corners in relation to the homography matrix was determined to be 7.6 pixels.

$$H = \begin{bmatrix} 92.195 & 0 & 43.107 \\ 0 & 92.574 & 32.063 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.1}$$

To evaluate the system's performance qualitatively, we conducted tests in a simulated industrial setting featuring a CNC machine and a heated aluminum plate. The plate was heated to various temperatures to visualize the holograms using a false-color representation. During the experiments, a male participant with industrial experience carried out different tasks, including assuming various positions and observing the machine. Following the completion of the experiments, we conducted semi-structured interviews with the participant to gather qualitative feedback regarding their individual experiences.

During the semi-structured interviews, the participant expressed positive feedback about our system, stating that he would be interested in using it in an industrial environment and felt confident while using it. He believed that people would easily adapt to the system without requiring much assistance or the need to learn many new concepts. The participant also appreciated the intuitive nature of the false-color representation and found that the holograms did not obstruct his view of the machine.

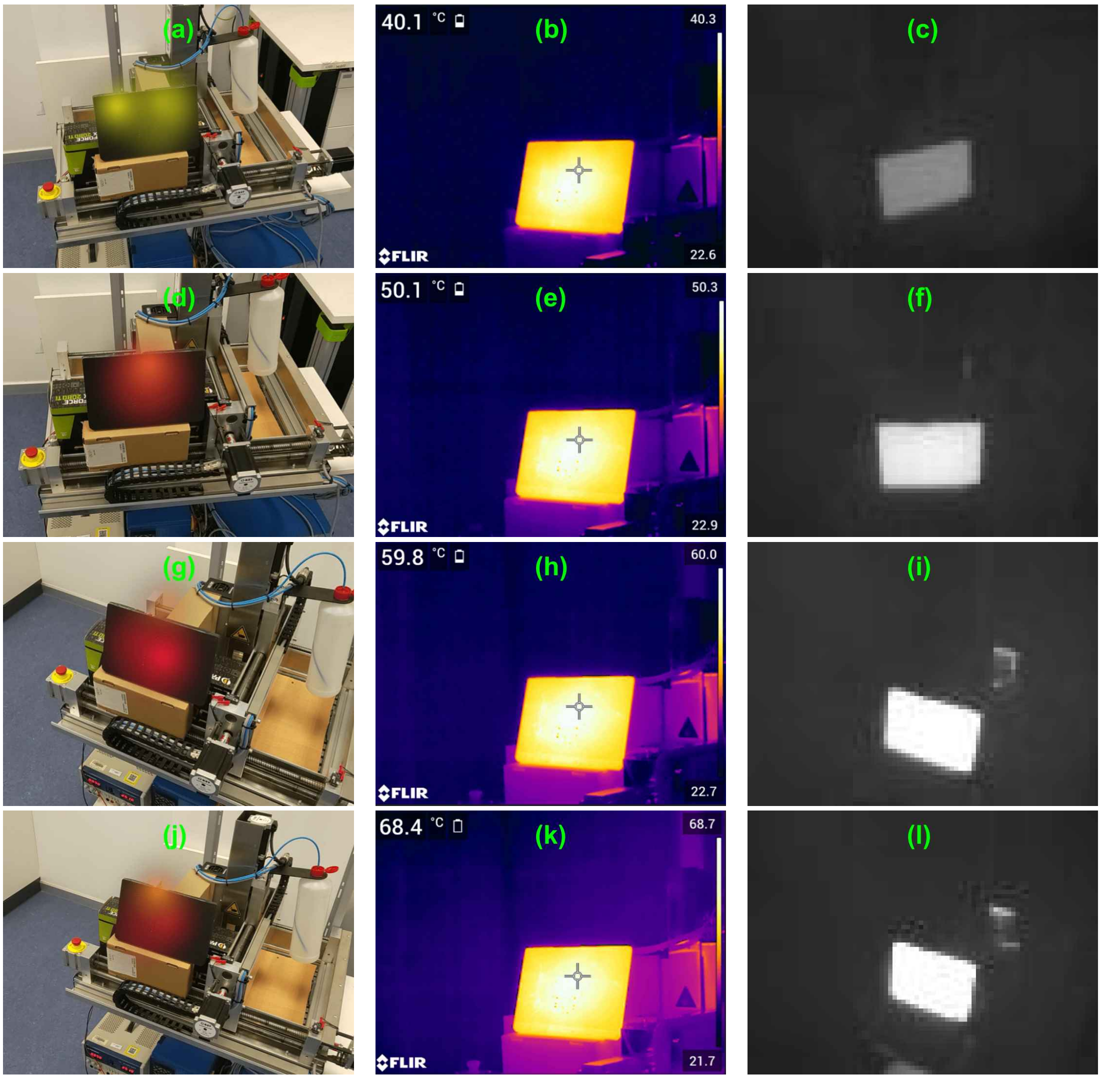


Figure 7.1: Views from Microsoft Hololens AR goggles (a-d), first-person thermal camera (e-h), and third-person thermal camera (i-l) at different temperatures: 40°C ((a), (c), and (i)), 50°C ((b),(f), and (j)), 60°C ((c),(j), and (k)), and 70°C ((d),(h), and (l)).

However, he did mention that the AR glasses were perceived as "bulky" and there were occasional latency issues.

It is important to acknowledge the limitations of our system, as highlighted during the interviews. The equipment in our setup, including the thermal camera with a low-frequency rate (9 Hz) and low-resolution thermal images (80×60 pixels), presents certain constraints. The combination of the thermal camera, Raspberry PI, and holder attached to the Hololens adds bulk and reduces the ergonomic comfort of the wearable device. The total weight of the Hololens 2 and the embedded system amounts to 0.75 kg, which may cause discomfort during prolonged usage. Knight and Baber [176, 177] showed that weight loaded to the front of the head such as HMD can have detrimental effects on the musculoskeletal system. These included neck extensors and sternocleidomastoid muscles. The increase in the perceived pain of participants or discomfort started in less than 8 minutes of wearing helmets with different loads

Figure 7.2: Thermal cameras compared to human finger and hand: a) Flir Lepton and b) Flir SC655

(from 0.5 kg to 2 kg) in the experiments. To address these concerns, integrating a low-weight high-resolution thermal camera into modern wearable AR goggles could be advantageous for industrial settings.

Utilizing thermal cameras with higher resolution could allow our safety system to detect hazards accurately in the environment. However, thermal cameras with high resolution are heavier and have larger volumetric envelopes. For instance, FLIR SC655 can capture images with high resolution (640×480) and at a faster rate (50Hz) but this camera is heavy (0.7kg) and consumes more power (24 W). The size of SC655 is much bigger than Flir Lepton (see Fig. 7.2). Increased power consumption necessitates the inclusion of larger batteries in the system. This can increase the weight of our AR safety system, which is already not lightweight. Miniaturization of thermal sensor arrays could help us to develop more efficient and ergonomic systems.

Furthermore, the system employs raycasting to position the hot object holograms. Due to the limited computing power of the AR glasses, the system operates at a slower update rate of approximately 0.5 frames per second. Enhancing the computing capabilities of modern AR glasses could significantly reduce the time required for hologram positioning.

Despite these limitations, the system exhibits the potential in enhancing worker safety within the industrial sector by augmenting human sensory perception. The seamless integration of the embedded system ensures the mobility and untethered operation of the AR glasses, further contributing to their practicality in real-world applications.

We tested the applicability of our system with a heated bed with different shapes. Experiments with heatable objects of different shapes can show the subtleties or limitations of our safety system. For instance, objects commonly used in factories such as pipes can be heated for experiments.

Integration of other sensors could provide more accurate knowledge about the environment. Temperature sensors are employed to identify instances of equipment and machinery overheating, while humidity sensors enable the monitoring of moisture levels in the air. To detect toxic gases, gas sensors are utilized, and pressure sensors are employed to measure the pressure of liquids and gases in pipes and tanks for visualization in AR. Light sensors are utilized to assess ambient light levels, and occupancy sensors detect the presence of individuals in designated areas. The data collected from these sensors are interconnected via IoT systems and subsequently conveyed to AR smart glasses, enriching the user's perception. This could allow the system to distinguish potential hazards and normal occurrences in the factory. This would increase the reliability of the system among users. To decrease the mass of the wearable AR system, we could leverage IoT for fusing information coming from different sensors in a factory. For instance, in the agricultural field, multiple cameras and sensors were positioned around a farm to visualize the various sensor data, such as coordinates of crops, soil moisture content, temperature, water level, nutrients, and luminance [178]. Multiple RGB cameras were positioned for stability and accuracy of visualizations and data was conveyed through the internet using IoT and Wireless Sensor Network technologies.

Different visualization can be tested to find intuitive interfaces. For instance, textual cues or warnings in different colors could trigger faster responses in participants of the experiment. Also, adding notifications with other cues could improve reaction time to warnings. Auditory and haptic modalities have the potential to react faster than other visual warnings. Safety instructions can be conveyed using different cues in AR goggles.

Designing our system to be adaptable for multiple industries could be the future perspective for our research. Different industries may have specific requirements for the safety of their workers. For instance, noisy and dusty industries might require to trigger olfactory warnings instead of auditory and visual warnings.

Chapter 8

Conclusion

In conclusion, this research has presented innovative approaches for enhancing industrial safety through the integration of AR and AI technologies. The AR warning system was developed to address the issue of inattention caused by interruptions during the working process. By harnessing AI and gaze-tracking technologies, the system provided auditory, visual, and audiovisual warnings to alert industrial operators of potential risks.

This method introduces a novel approach on several fronts. Firstly, it marks the inception of an AR-enhanced warning system specifically designed for industrial safety. Secondly, it showcases the implementation of DL-based object detection within the framework of AR systems. Notably, our solution is tailored for untethered AR smart goggles, specifically the Microsoft HoloLens 2. These goggles offer distinct advantages as they are not confined to specific locations like spatial AR and do not necessitate the use of hands, as seen in mobile AR setups. Furthermore, our system ensures a seamless display of notifications directly within the user's field of perception, contrasting with the potential distraction posed by notifications on the screens of mobile devices and tablets.

The findings of this study have successfully answered the research questions posed. Participants utilizing the AR system demonstrated reduced instances and durations of gazing away from the working space compared to those who did not use the AR system, indicating the system's effectiveness in promoting attention (RQ1). Questionnaire responses revealed that the audiovisual AR warnings were preferred by participants, underscoring their intuitiveness and effectiveness (RQ2). Moreover, the AR warning system outperformed the conventional safety mat in minimizing distractions in terms of their number and duration (RQ3). Notably, participants reported less frustration when visually alerted.

Overall, the acceptance and perceived benefits of the AR warning system have confirmed its effectiveness in enhancing industrial safety (H1). Further research is warranted to objectively evaluate the efficacy of audiovisual warnings in real-world industrial settings. This study has demonstrated the significant potential of AI-enhanced AR systems in improving industrial safety.

Furthermore, this research has introduced an AR-based safety assistant that overlays holograms on hot objects in 3D space, augmenting human perception. By leveraging edge computing and thermal imaging technologies, the system enhances the detection and visualization of potential hazards.

Our technical contributions involve the development of an untethered embedded system designed for thermal sensing, utilizing the Flir Lepton 2.5 thermal sensing array. This system was seamlessly integrated into AR goggles. A crucial aspect of our contribution lies in the calibration of the homography matrix between the thermal array and the visual camera of the AR goggles. This calibration process yielded an average corner error of 7.6 pixels for the homography estimation. This meticulous calibration is instrumental in ensuring high spatial accuracy when displaying thermal holograms.

The experiment demonstrated the system's capability to provide operators with real-time visual feedback on hot objects, effectively conveying the risk of potential injury. This integration of thermal sensing into the AR environment contributes to enhanced safety measures by empowering operators with immediate awareness of the thermal conditions in their surroundings.

The participant's feedback during the experiments indicated a sense of safety when utilizing the system, highlighting the viability and potential of thermal imaging with AR for industrial safety (RQ4). However, it is essential to acknowledge the limitations imposed by the current state of AR technologies. These limitations can be addressed through the development of new and advanced AR systems.

In conclusion, this study has contributed to the advancement of industrial safety through the integration of AR and AI technologies. The findings underscore the effectiveness and potential of these technologies in promoting attention, minimizing distractions, and augmenting human perception. It is hoped that this research will inspire further investigations and developments in the field, ultimately leading to safer working environments in various industrial settings.

The long-term use of both warning systems and safety systems with thermal imaging can have different effects on people. The systems can not be used for daily use because of the bulkiness and battery life of smart glasses. Therefore, the real-world implementation and long-term effects should be explored further. Conducting studies and experiments in real industrial settings can give valuable insights into the further implementation of AR-based safety systems.

Novel AR technologies should be explored that could enable spatial and seamless computing. Interactions with the system can be more immersive and accurate and AR-based safety systems could operate faster than present with newer technologies. For instance, a slow sample rate of AR rendering can cause cybersickness in users [179]. This can be bypassed using fast and powerful computing units.

Developing AI algorithms tailored for industrial safety applications can improve accuracy and decrease processing time. Real-time risk assessment and predictive analytics algorithms that anticipate potential hazards can prevent accidents. The safety measures can be changed from situation to situation. AI models for adaptive safety measures can be a focus of further investigations.

Training industrial workers in simulated hazardous scenarios can be an avenue for future research. AR technology can be utilized to simulate hazardous scenarios. Circumventing real-world hazards with holographic hazardous environments in different scenarios could enhance the employee's reaction time and improve situational awareness in case of real accidents.

AR-based assembly assistants can be developed to surpass paper-based instructions according to time and other parameters in different tasks(see Fig. 8.1). Inspection systems can be integrated into the assistant to increase the quality of products and the safety of participants. Machine inspection can be faster and more accurate than a human worker on inspection. AR-based assembly assistants with inspection options can overcome human limitations by presenting information and instructions on top of the product and tools.

The spatial mapping capabilities of modern AR glasses are limited in the number of meshes due to the computing capacity. Spatial mapping of AR glasses can render coarse triangle meshes on the environment. This hinders developers and researchers from positioning holographic objects accurately. The computing capabilities of AR goggles limit us in the number of ray casts. Consequently, the number of detected objects in 2D images that can be positioned in a 3D map is also limited. Algorithms and devices for optimization or refining of spatial meshes should be explored for accurate positioning and an increased number of holograms.

Integrating AR and AI with IoT can enhance the capabilities of industrial safety. IoT sensors can provide real-time information about the status of equipment and the environment where the workers are working. AI can detect potential hazards and AR can be equipped to place holographic content on top of the detected places.

Human attention and sensory systems have limitations in detecting hazards in the surrounding environment regardless of how employees were trained. The systems that collaborate with human workers could increase the productivity of workers and decrease accidents in the industry. Intuitive interactions for human workers can be investigated for human-machine collaboration. For instance, gestures that are intuitive for humans or similar to human-human interaction can make systems more ergonomic. The usability of the system can be increased following these types of studies.

The cognitive load of human workers can be explored by using different safety systems. Mental fatigue and physical strain can be reduced by understanding the cognitive load of workers. Especially, conducting experiments with the systems in real-world scenarios can provide future trends in research on industrial safety with AR and AI. AR can be improved to reduce annoying content in the FOV of the user whereas the physical and mental state of human workers can recognized using AI.

Data visualization has been studied in computers that we use for many years. The outcomes of these studies can be tested with AR visualizations. In case improvements

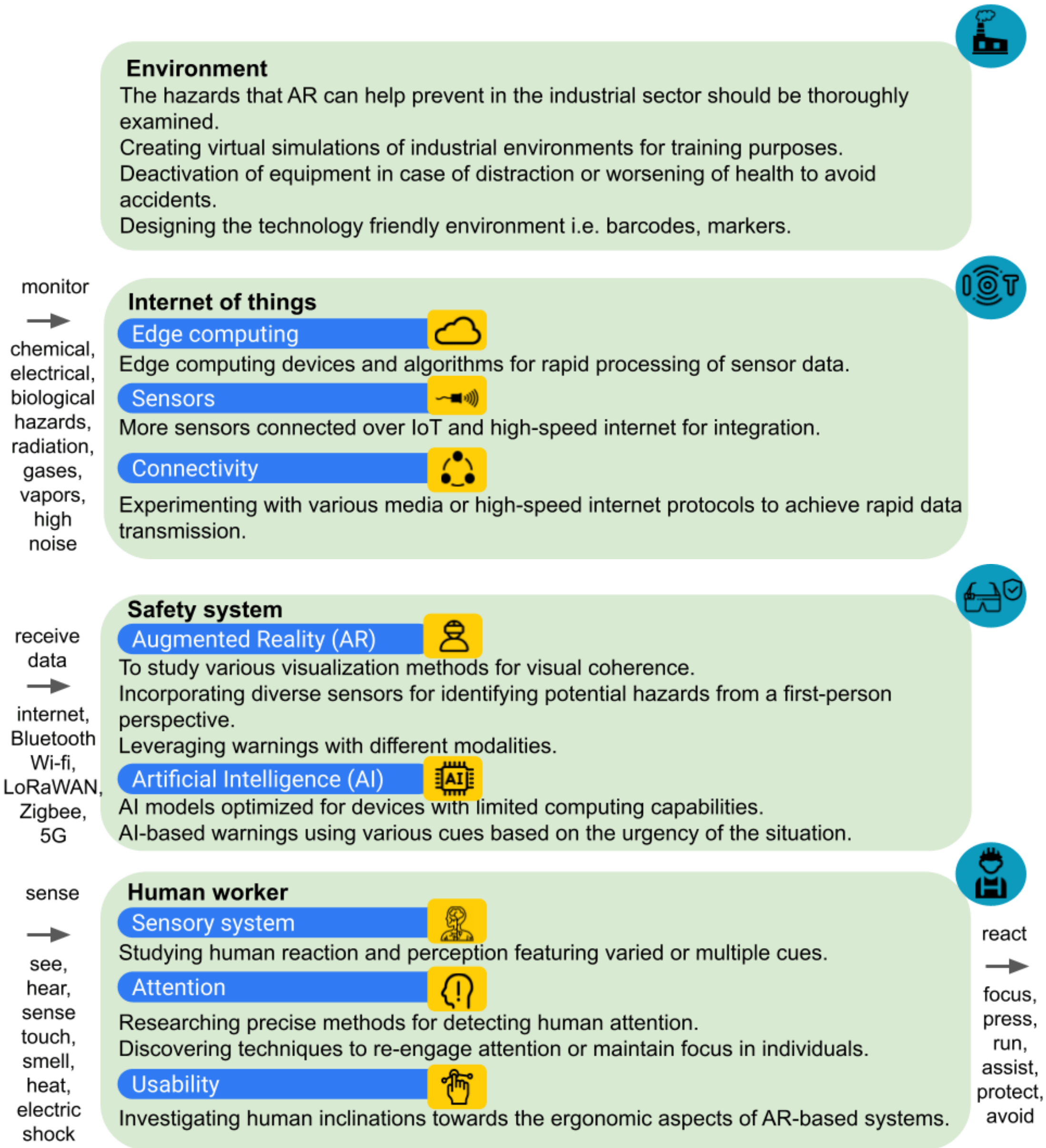


Figure 8.1: Future perspectives in the research of AR-based safety systems

are needed, the visualizations can be adapted for holographic content. It may happen that many approaches to visualizations that are used in personal computers, tablets, and smartphones might not work in AR visualizations. In that case, novel approaches particular for AR should be developed for visualizing content in AR.

Detecting the human worker's attention is imperative for industrial safety. Finding accurate ways of detecting the state of attention is an active research area. We utilized human gaze direction in order to understand the attention of humans. However, more accurate indicators might be required for very delicate jobs or cases. Accurate attention indicators are essential for surgeons, air traffic controllers, nuclear power plant operators, and astronauts, as they perform tasks that require precise focus and concentration. Physiological and behavioral indicators can be monitored using various sensors. For instance, electoroencephalography (EEG) measures electrical brain activity, electrocardiography (ECG) measures heart rate and heart variability, and electrodermal activity (EDA) sensors measure the electrical conductance of skin. All these measurements can accurately indicate changes in attention.

The AR-based warning system proposed in the study helped avoid inattention that might lead to accidents. However, there are a number of human factors that might contribute to accidents. These include fatigue, stress, pressures such as tight deadlines, lack of training, complacency, poor communication between workers, physical impairments, substance abuse, and lack of situational awareness. The impact of human factors should be minimized for the safety of workers. AR-based systems that can detect human factors that contribute to accidents can be explored. AI algorithms can be integrated for analyzing the actions of users and other members. These processes must adhere to industry regulations and align with ethical principles, including safeguarding individual privacy.

During the experiments, the preferences of human participants were different according to the results of semi-structured interviews. Investigating human preferences during tasks could provide options for future AR-based safety systems.

The adaptability of AR-based systems to different industrial environments, human preferences, or equipment should be tested. AR-based systems can be either adapted to different domains manually or designed with preference settings in the beginning. Designing adaptable AR-based systems could allow people to utilize them for numerous tasks. Furthermore, it could be enhanced with AI-based algorithms to set configurations according to the industrial environment, user interactions, or type of equipment. That kind of smart adaptable AR-based safety system could help avoid dangerous situations for novice workers.

In the industrial setting, there are numerous hazards that can lead to serious and even fatal consequences for workers. These hazards encompass a wide range of risks, such as chemical exposure, machinery accidents, falls, and electrical incidents. To address these issues, different approaches should be tested. Especially, AR-based solutions could be advantageous over business-as-usual solutions. To fully harness the potential of AR-based solutions, a systematic exploration of all processes is necessary. This can include regulatory compliance, cost-benefit analysis, usability, human-computer interaction, hardware, and software. Collaboration of experts from different fields such as industrial engineers, computer scientists, biomedical engineers, and ergonomics in designing AR-based safety systems is essential. Their collective knowledge and expertise lead to innovative, user-centered solutions that enhance industrial safety, improve human performance, and create safer and more productive working environments.

BLANK

Bibliography

[1] Bureau of National Statistics of Kazakhstan, "On work-related injuries and occupational diseases in the Republic of Kazakhstan for 2022." https://stat.gov.kz/ru/news/o-travmatizme-svyazannom-s-trudovoy-deyatelnostyu-i-professionalnykh-zabolevaniyakh-v-respublike-kaz/?sphrase_id=113422, 2022. [Online; accessed 02-June-2023].

[2] International Labour Organization, "Accidents at work statistics." https://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang--en/index.htm, 2022. [Online; accessed 02-June-2023].

[3] Eurostat, "The enormous burden of poor working conditions." https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_statistics, 2022. [Online; accessed 02-June-2023].

[4] U. B. O. L. STATISTICS, "2.8 million nonfatal workplace injuries and illnesses occurred in private industry in 2019." https://www.bls.gov/opub/ted/2020/2-8-million-nonfatal-workplace-injuries-and-illnesses-occurred-in-private-industry-in-2019.htm, 2022. [Online; accessed 02-June-2023].

[5] Y. Chinniah, "Analysis and prevention of serious and fatal accidents related to moving parts of machinery," *Safety Science*, vol. 75, pp. 163–173, 2015.

[6] M. Dźwiarek and A. Latała, "Analysis of occupational accidents: Prevention through the use of additional technical safety measures for machinery," *International Journal of Occupational Safety and Ergonomics*, vol. 22, no. 2, pp. 186–192, 2016. PMID: 26652689.

[7] C. Safety, "New report: Burns and scalds injuries in the workplace." https://cesafety.co.uk/news/new-report-burns-and-scalds-injuries-in-the-workplace/, 2019. [Online; accessed 02-June-2023].

[8] U.S. Bureau of Labour Statistics, "Injuries, illnesses, and fatalities." https://www.bls.gov/web/osh/cd_r1.htm, 2020. [Online; accessed 02-June-2023].

[9] F. L. Ready, Y. D. Gebremedhem, M. Worku, K. Mehta, M. Eshte, Y. L. GoldenMerry, F. E. Nwariaku, S. E. Wolf, and H. A. Phelan, "Epidemiologic

shifts for burn injury in Ethiopia from 2001 to 2016: Implications for public health measures," *Burns*, vol. 44, no. 7, pp. 1839–1843, 2018.

[10] S. S. Kudryavtsev, P. V. Yemelin, and N. K. Yemelina, "The development of a risk management system in the field of industrial safety in the Republic of Kazakhstan," *Safety and Health at Work*, vol. 9, no. 1, pp. 30–41, 2018.

[11] F. Mohammadi Amin, M. Rezayati, H. W. van de Venn, and H. Karimpour, "A mixed-perception approach for safe human–robot collaboration in industrial automation," *Sensors*, vol. 20, no. 21, 2020.

[12] I. E. Sutherland, "The ultimate display," in *Proc. International Federation of Information Processing Congress (IFIP)*, pp. 506–508, Macmillan and Co.: London, UK, 1965.

[13] N. J. Wade, "Charles Wheatstone (1802–1875)," 2002.

[14] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, pp. 355–385, 08 1997.

[15] P. C. Thomas and W. David, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," in *Proc of the Hawaii international Conference on System Sciences*, vol. 2, ACM SIGCHI Bulletin, 1992.

[16] L. Rosenberg, "Virtual fixtures: Perceptual tools for telerobotic manipulation," in *Proc. of IEEE Virtual Reality Annual International Symposium*, pp. 76–82, 1993.

[17] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," in *Proc. of the IEEE and ACM International Workshop on Augmented Reality (IWAR)*, pp. 85–94, 1999.

[18] "The Wikitude website." https://www.wikitude.com/, 2023. [Online; accessed 02-June-2023].

[19] "The Pokemon Go website." https://pokemongolive.com/?hl=en, 2023. [Online; accessed 02-June-2023].

[20] J. Hong, "Considering privacy issues in the context of Google Glass," *Communications of the ACM*, vol. 56, no. 11, p. 10–11, 2013.

[21] F. Tavares, "NASA's using augmented reality to transform air traffic management." https://www.nasa.gov/feature/ames/augmented-reality-air-traffic-management, 2022. [Online; accessed 2-June-2023].

[22] Microsoft, "RE'FLEKT case study." https://partner.microsoft.com/en-sa/case-studies/reflekt, 2022. [Online; accessed 02-June-2023].

[23] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[24] U. Neumann and S. You, "Natural feature tracking for augmented reality," *IEEE Transactions on Multimedia*, vol. 1, no. 1, pp. 53–64, 1999.

[25] G. Reitmayr and T. W. Drummond, "Going out: Robust model-based tracking for outdoor augmented reality," in *Proc. of the IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 109–118, 2006.

[26] P. Hübner, M. Weinmann, and S. Wursthorn, "Marker-based localization of the Microsoft Hololens in building models," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 42, no. 1, 2018.

[27] J. Rekimoto, "Matrix: A realtime object identification and registration method for augmented reality," in *Proc. of the IEEE Asia Pacific Computer Human Interaction*, pp. 63–68, IEEE, 1998.

[28] C. Xiao and Z. Lifeng, "Implementation of mobile augmented reality based on Vuforia and Rawajali," in *Proc. IEEE International Conference on Software Engineering and Service Science*, pp. 912–915, IEEE, 2014.

[29] C.-H. Lin, Y. Chung, B.-Y. Chou, H.-Y. Chen, and C.-Y. Tsai, "A novel campus navigation app with augmented reality and deep learning," in *Proc. IEEE International Conference on Applied System Invention (ICASI)*, pp. 1075–1077, IEEE, 2018.

[30] P. Ren, X. Qiao, Y. Huang, L. Liu, S. Dustdar, and J. Chen, "Edge-assisted distributed dnn collaborative computing approach for mobile web augmented reality in 5G networks," *IEEE Network*, vol. 34, no. 2, pp. 254–261, 2020.

[31] M. R. Marner, R. T. Smith, J. A. Walsh, and B. H. Thomas, "Spatial user interfaces for large-scale projector-based augmented reality," *IEEE Computer Graphics and Applications*, vol. 34, no. 6, pp. 74–82, 2014.

[32] O. Bimber and R. Raskar, *Spatial augmented reality: Merging real and virtual worlds*. CRC Press, 2005.

[33] A. Erickson, K. Kim, G. Bruder, and G. F. Welch, "Exploring the limitations of environment lighting on optical see-through head-mounted displays," in *Proc.*

Symposium on Spatial User Interaction, (New York, NY, USA), Association for Computing Machinery, 2020.

[34] Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, S. Qian, X. Liu, Y. Wu, F. Dong, C.-W. Qiu, *et al.*, "Artificial intelligence: A powerful paradigm for scientific research," *The Innovation*, vol. 2, no. 4, p. 100179, 2021.

[35] K. Cao, Y. Liu, G. Meng, and Q. Sun, "An overview on edge computing research," *IEEE Access*, vol. 8, pp. 85714–85728, 2020.

[36] L. Qian, Z. Luo, Y. Du, and L. Guo, "Cloud computing: An overview," in *Cloud Computing* (M. G. Jaatun, G. Zhao, and C. Rong, eds.), (Berlin, Heidelberg), pp. 626–631, Springer Berlin Heidelberg, 2009.

[37] K. Lyytinen and Y. Yoo, "Ubiquitous computing," *Communications of the ACM*, vol. 45, no. 12, pp. 63–96, 2002.

[38] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. USA: Pearson, 4th ed., 2020.

[39] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[40] O. Pauly, B. Diotte, P. Fallavollita, S. Weidert, E. Euler, and N. Navab, "Machine learning-based augmented reality for improved surgical scene understanding," *Computerized Medical Imaging and Graphics*, vol. 41, pp. 55–60, 2015.

[41] K. Ichihashi and K. Fujinami, "Estimating visibility of annotations for view management in spatial augmented reality based on machine-learning techniques," *Sensors*, vol. 19, no. 4, p. 939, 2019.

[42] R. E. Sutanto, L. Pribadi, and S. Lee, "3D integral imaging based augmented reality with deep learning implemented by faster R-CNN," in *Proc. of the International Conference on Mobile and Wireless Technology*, pp. 241–247, Springer, 2017.

[43] B. Hoppenstedt, K. Kammerer, M. Reichert, M. Spiliopoulou, and R. Pryss, "Convolutional neural networks for image recognition in mixed reality using voice command labeling," in *Proc. of the International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 63–70, Springer, 2019.

[44] B. Huynh, J. Orlosky, and T. Höllerer, "Semantic labeling and object registration for augmented reality language learning," in *Proc. of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 986–987, IEEE, 2019.

[45] K. Lebeck, E. Cuervo, and M. Philipose, "Collaborative acceleration for mixed reality," Tech. Rep. MSR-TR-2018-20, Technical report, Microsoft, 2018. 49, July 2018.

[46] F. Göbel, K. Kurzhals, M. Raubal, and V. R. Schinazi, "Gaze-aware mixed-reality: Addressing privacy issues with eye tracking," in *CHI 2020 Workshop on Exploring Potentially Abusive Ethical, Social and Political Implications of Mixed Reality in HCI (CHI 2020)*, 2020.

[47] F. Roesner, T. Kohno, and D. Molnar, "Security and privacy for augmented reality systems," *Communications of the ACM*, vol. 57, no. 4, pp. 88–96, 2014.

[48] Microsoft, "Hololens 2 privacy and data protection." https://learn.microsoft.com/en-us/hololens/hololens2-privacy, 2022. [Online; accessed 02-June-2023].

[49] F. Ragusa, A. Furnari, S. Battiato, G. Signorello, and G. M. Farinella, "EGO-CH: Dataset and fundamental tasks for visitors behavioral understanding using egocentric vision," *Pattern Recognition Letters*, vol. 131, pp. 150–157, 2020.

[50] H. Dhiman, S. Büttner, C. Röcker, and R. Reisch, "Handling work complexity with AR/deep learning," in *Proc. of the Australian Conference on Human-Computer-Interaction*, pp. 518–522, Association for Computing Machinery (ACM), 2019.

[51] M. Runz, M. Buffier, and L. Agapito, "Maskfusion: Real-time recognition, tracking and reconstruction of multiple moving objects," in *Proc. of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 10–20, IEEE, 2018.

[52] "Matterport implementation of mask-rcnn." https://github.com/matterport/Mask_RCNN. Accessed: 2020-04-25.

[53] K. Židek, P. Lazorík, J. Pitel', and A. Hošovskỳ, "An automated training of deep learning networks by 3D virtual models for object recognition," *Symmetry*, vol. 11, no. 4, p. 496, 2019.

[54] S. Li, Y. Dou, J. Xu, Q. Wang, and X. Niu, "mmCNN: A novel method for large convolutional neural network on memory-limited devices," in *Proc. of the IEEE Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1, pp. 881–886, IEEE, 2018.

[55] Y. Ma, K. Liu, Z. Guan, X. Xu, X. Qian, and H. Bao, "Background augmentation generative adversarial networks (BAGANs): Effective data generation based on gan-augmented 3D synthesizing," *Symmetry*, vol. 10, no. 12, p. 734, 2018.

[56] Q. H. Gao, T. R. Wan, W. Tang, and L. Chen, "Object registration in semi-cluttered and partial-occluded scenes for augmented reality," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15079–15099, 2019.

[57] J. Zubizarreta, I. Aguinaga, and A. Amundarain, "A framework for augmented reality guidance in industry," *The International Journal of Advanced Manufacturing Technology*, vol. 102, no. 9-12, pp. 4095–4108, 2019.

[58] K. Yang, X. Gong, S. Song, T. Xing, X. Chen, and D. Fang, "SpeAR: A fast AR system with high accuracy deployed on mobile devices," in *Proc. of the IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 113–120, IEEE, 2017.

[59] R. Moezzi, D. Krcmarik, J. Hlava, and J. Cýrus, "Hybrid SLAM modelling of autonomous robot with augmented reality device," *Materials Today: Proceedings*, vol. 32, pp. 103–107, 2020. DAS2019.

[60] M. Garon, P.-O. Boulet, J.-P. Doiron, L. Beaulieu, and J.-F. Lalonde, "Real-time high resolution 3D data on the hololens," in *Proc. of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 189–191, IEEE, 2016.

[61] K.-B. Park, S. H. Choi, M. Kim, and J. Y. Lee, "Deep learning-based mobile augmented reality for task assistance using 3D spatial mapping and snapshot-based RGB-D data," *Computers & Industrial Engineering*, vol. 146, p. 106585, 2020.

[62] R.-M. Rahal and S. Fiedler, "Understanding cognitive and affective mechanisms in social psychology through eye-tracking," *Journal of Experimental Social Psychology*, vol. 85, p. 103842, 2019.

[63] M. Wedel and R. Pieters, "A review of eye-tracking research in marketing," *Review of Marketing Research*, pp. 123–147, 2008.

[64] P. Majaranta and A. Bulling, *Eye Tracking and Eye-Based Human–Computer Interaction*, pp. 39–65. London: Springer London, 2014.

[65] S. Kapp, M. Barz, S. Mukhametov, D. Sonntag, and J. Kuhn, "ARETT: Augmented reality eye tracking toolkit for head mounted displays," *Sensors*, vol. 21, no. 6, 2021.

[66] "Unity website." https://docs.unity3d.com/Packages/com.unity.xr.openxr@1.0/manual/features/eyegazeinteraction.html, 2022. [Online; accessed 02-June-2023].

[67] S. Naspetti, R. Pierdicca, S. Mandolesi, M. Paolanti, E. Frontoni, and R. Zanoli, "Automatic analysis of eye-tracking data for augmented reality applications: A prospective outlook," in *Augmented Reality, Virtual Reality, and Computer Graphics* (L. T. De Paolis and A. Mongelli, eds.), (Cham), pp. 217–230, Springer International Publishing, 2016.

[68] A. Erickson, K. Kim, G. Bruder, and G. F. Welch, "Beyond visible light: User and societal impacts of egocentric multispectral vision," in *Virtual, Augmented and Mixed Reality: 13th International Conference, VAMR 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings*, pp. 317–335, Springer, 2021.

[69] A. Erickson, K. Kim, R. Schubert, G. Bruder, and G. Welch, "Is it cold in here or is it just me? analysis of augmented reality temperature visualization for computer-mediated thermoception," in *Proc. of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 202–211, IEEE, 2019.

[70] J. Orlosky, P. Kim, K. Kiyokawa, T. Mashita, P. Ratsamee, Y. Uranishi, and H. Takemura, "Vismerge: Light adaptive vision augmentation via spectral and temporal fusion of non-visible light," in *Proc. of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 22–31, 2017.

[71] P. Lincoln, A. Blate, M. Singh, T. Whitted, A. State, A. Lastra, and H. Fuchs, "From motion to photons in 80 microseconds: Towards minimal latency for virtual and augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 4, pp. 1367–1376, 2016.

[72] B. Briggs, "Vroom with a view: Hololens 2 powers faster fixes for Mercedes-Benz USA." https://news.microsoft.com/source/features/digital-transformation/vroom-with-a-view-hololens-2-powers-faster-fixes-mercedes-benz-usa/, 2020. [Online; accessed 02-June-2023].

[73] "The Wärtsilä Seals & Bearings website." https://www.wartsila.com/media/news/31-07-2018-augmented-reality-creates-a-new-dimension-in-marine-maintenance-services, 2022. [Online; accessed 02-June-2023].

[74] Z. Makhataeva and H. A. Varol, "Augmented reality for robotics: A review," *Robotics*, vol. 9, no. 2, 2020.

[75] A. Yew, S. Ong, and A. Nee, "Immersive augmented reality environment for the teleoperation of maintenance robots," *Proc. of the CIRP Conference on Life Cycle Engineering*, vol. 61, pp. 305–310, 2017.

[76] Z. Makhataeva, A. Zhakatayev, and H. A. Varol, "Safety aura visualization for variable impedance actuated robots," in *Proc. of the IEEE/SICE International Symposium on System Integration (SII)*, pp. 805–810, 2019.

[77] H. B. Chenery, S. Robinson, M. Syrquin, and S. Feder, *Industrialization and growth.* Oxford University Press New York, 1986.

[78] G. Mark, V. M. Gonzalez, and J. Harris, "No task left behind? examining the nature of fragmented work," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, (New York, NY, USA), p. 321–330, Association for Computing Machinery, 2005.

[79] K. A. Fletcher, S. M. Potter, and B. N. Telford, "Stress outcomes of four types of perceived interruptions," *Human Factors*, vol. 60, no. 2, pp. 222–235, 2018. PMID: 29131659.

[80] M. A. Orhan, S. Castellano, I. Khelladi, L. Marinelli, and F. Monge, "Technology distraction at work. Impacts on self-regulation and work engagement," *Journal of Business Research*, vol. 126, pp. 341–349, 2021.

[81] C. P. Beaman, "Auditory distraction from low-intensity noise: A review of the consequences for learning and workplace environments," *Applied Cognitive Psychology*, vol. 19, no. 8, pp. 1041–1064, 2005.

[82] J. Olson, "Research about office workplace activities important to US businesses - and how to support them," *Journal of Facilities Management*, vol. 1, pp. 31–47, 01 2002.

[83] E. M. Altmann, J. G. Trafton, and D. Z. Hambrick, "Momentary interruptions can derail the train of thought.," *Journal of Experimental Psychology: General*, vol. 143, no. 1, p. 215, 2014.

[84] P. Sterling and S. Laughlin, *Principles of neural design.* MIT press, 2015.

[85] Margaret Crable, "Exactly how many senses do we really have?." https://neurosciencenews.com/sense-intuition-consciousness-22296/, 2023. [Online; accessed 02-June-2023].

[86] C. Matteo, S. Marina, B. Patrizia, *et al.*, "Equilibrioception: a method to evaluate the sense of balance," in *Proc. of the Consortium, Posters and Demos at CHItaly (DCPD@ CHItaly 2017), Cagliari, Italy, September 18-20, 2017*, pp. 143–147, CEUR-WS, 2017.

[87] J. C. Tuthill and E. Azim, "Proprioception," *Current Biology*, vol. 28, no. 5, pp. R194–R203, 2018.

[88] P. Pagliano, "The multisensory environment–time to sense," *Everyday Multisensory Environments, Wellness Technology and Snoezelen*, p. 18, 2014.

[89] M. G. Paulin, "Electroreception and the compass sense of sharks," *Journal of Theoretical Biology*, vol. 174, no. 3, pp. 325–339, 1995.

[90] R. A. Holland, J. L. Kirschvink, T. G. Doak, and M. Wikelski, "Bats use magnetite to detect the earth's magnetic field," *PLoS One*, vol. 3, no. 2, p. e1676, 2008.

[91] K. P. ABLE, "Skylight polarization patterns and the orientation of migratory birds," *Journal of Experimental Biology*, vol. 141, no. 1, pp. 241–256, 1989.

[92] A. Kokocińska-Kusiak, M. Woszczyło, M. Zybala, J. Maciocha, K. Barłowska, and M. Dzięcioł, "Canine olfaction: physiology, behavior, and possibilities for practical applications," *Animals*, vol. 11, no. 8, p. 2463, 2021.

[93] A. Kaveh and N. Farhoudi, "A new optimization method: Dolphin echolocation," *Advances in Engineering Software*, vol. 59, pp. 53–70, 2013.

[94] H. Mouritsen and P. Hore, "The magnetic retina: Light-dependent and trigeminal magnetoreception in migratory birds," *Current Opinion in Neurobiology*, vol. 22, no. 2, pp. 343–352, 2012. Neuroethology.

[95] X. Liu, "Airborne LiDAR for DEM generation: Some critical issues," *Progress in Physical Geography*, vol. 32, no. 1, pp. 31–49, 2008.

[96] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," in *Proc. of the International Joint Conference on Neural Networks (IJCNN)*, pp. 381–388, IEEE, 2016.

[97] A. Jain, R. Bansal, A. Kumar, and K. Singh, "A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students," *International Journal of Applied and Basic Medical Research*, vol. 5, no. 2, p. 124, 2015.

[98] T. Hülsdünker, M. Ostermann, and A. Mierau, "The speed of neural visual motion perception and processing determines the visuomotor reaction time of young elite table tennis athletes," *Frontiers in Behavioral Neuroscience*, vol. 13, p. 165, 2019.

[99] J. E. Cometto-Muniz and W. S. Cain, *Olfactory adaptation*, pp. 257–281. New York: Marcel Dekker, 1995.

[100] B. Iravani, M. Schaefer, D. A. Wilson, A. Arshamian, and J. N. Lundström, “The human olfactory bulb processes odor valence representation and cues motor avoidance behavior,” *Proc. of the National Academy of Sciences*, vol. 118, no. 42, p. e2101209118, 2021.

[101] T. Yamamoto and Y. Kawamura, “Gustatory reaction time in human adults,” *Physiology & Behavior*, vol. 26, no. 4, pp. 715–719, 1981.

[102] H. A. Hamid, M. N. Abdullah, M. Asmoni, M. A. A. Lokman, and N. Shaari, “An overview of the management commitment to safety elements for mitigating accidents in the construction industry,” *Jurnal Teknologi*, vol. 74, no. 2, pp. 1–8, 2015.

[103] International Labour Organization, “Learning and training for work in the knowledge society.” https://www.ilo.org/public/english/standards/relm/ilc/ilc91/pdf/rep-iv-1.pdf, 2002. [Online; accessed 02-June-2023].

[104] K. L. Chan and A. H. Chan, “Understanding industrial safety signs: Implications for occupational safety management,” *Industrial Management & Data Systems*, vol. 111, no. 9, pp. 1481–1510, 2011.

[105] J. Komaki, A. T. Heinzmann, and L. Lawson, “Effect of training and feedback: Component analysis of a behavioral safety program.,” *Journal of Applied Psychology*, vol. 65, pp. 261–270, 1980.

[106] J. K. Ford and A. M. Schmidt, “Emergency response training: strategies for enhancing real-world performance,” *Journal of Hazardous Materials*, vol. 75, no. 2-3, pp. 195–215, 2000.

[107] The National Institute for Occupational Safety and Health (NIOSH), “Noise and occupational hearing loss.” https://www.cdc.gov/niosh/topics/noise/default.html#:~:text=85%20decibels,Recommended%20Exposure%20Limit%20(REL)., 2023. [Online; accessed 02-June-2023].

[108] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, “Industry 4.0,” *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, 2014.

[109] “The International Data Corporation website.” https://www.idc.com/getdoc.jsp?containerId=prAP46737220, 2020. [Online; accessed 02-June-2023].

[110] Brian Barrett, “Daqri smart helmet.” wired.com/2016/01/daqri-helmet/, 2016. [Online; accessed 02-June-2023].

[111] Adi Robertson, “Augmented reality headset company Daqri is reportedly shutting down.” https://www.theverge.com/2019/9/13/20864556/daqri-ar-headset-

smart-glasses-startup-shutdown-asset-sale-layoffs, 2019. [Online; accessed 02-June-2023].

[112] "The Trimble website." https://fieldtech.trimble.com/en/products/mixed-reality/trimble-xr10-with-hololens-2, 2022. [Online; accessed 02-June-2023].

[113] "The VisualLive website." https://visuallive.com/hard-hat-for-hololens-2-system/, 2022. [Online; accessed 02-June-2023].

[114] "The Honeywell website." https://pmt.honeywell.com/us/en/about-pmt/newsroom/featured-stories/a-virtual-world-optimizing-industrial-training-and-operations, 2022. [Online; accessed 02-June-2023].

[115] "The RUMEN website." https://www.behance.net/gallery/148209271/RUMEN-AR-HMD-Device, 2023. [Online; accessed 02-June-2023].

[116] "The RealWear website." https://www.realwear.com/, 2023. [Online; accessed 02-June-2023].

[117] C. Warren, "The future of energy will be augmented." https://eprijournal.com/the-future-of-energy-will-be-augmented/, 2018. [Online; accessed 02-June-2023].

[118] T. J. Website, "AR headsets give oil and gas sector the quicker fix." https://jpt.spe.org/ar-headsets-give-oil-and-gas-sector-quicker-fix, 2018. [Online; accessed 02-June-2023].

[119] "The Youbiquo website." https://www.youbiquo.eu/, 2022. [Online; accessed 02-June-2023].

[120] F. Sasangohar, "Improving interruption recovery in human-supervisory control (HSC)," Master's thesis, University of Waterloo, 2009.

[121] S. D. Scott, S. Mercier, M. L. Cummings, and E. Wang, "Assisting interruption recovery in supervisory control of multiple UAVs," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 5, pp. 699–703, 2006.

[122] B. P. Bailey, J. A. Konstan, and J. V. Carlis, "The effects of interruptions on task performance, annoyance, and anxiety in the user interface," in *Proc. of the International Conference on Human-Computer Interaction*, (Tokyo, Japan), pp. 593–601, 2001.

[123] P. D. Adamczyk and B. P. Bailey, "If not now, when? The effects of interruption at different moments within task execution," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 271–278, 2004.

[124] D. E. Brown, B. Reimer, B. Mehler, and J. Dobres, "An on-road study involving two vehicles: Observed differences between an auditory and haptic lane departure warning system," in *Proc. of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, (New York, NY, USA), p. 140–145, Association for Computing Machinery, 2015.

[125] L. Macchi, M. Martinetto, and P. C. Cacciabue, "Assessing a driver support system by field test and subjective data analysis: The case of EUCLIDE anti-collision warning system," in *Proc. of the Annual Conference on European Association of Cognitive Ergonomics*, p. 155–158, University of Athens, 2005.

[126] V. Frémont, M.-T. Phan, and I. Thouvenin, "Adaptive visual assistance system for enhancing the driver awareness of pedestrians," *International Journal of Human–Computer Interaction*, vol. 36, no. 9, pp. 856–869, 2020.

[127] A. Calvi, F. D'Amico, C. Ferrante, and L. Bianchini Ciampoli, "Effectiveness of augmented reality warnings on driving behaviour whilst approaching pedestrian crossings: A driving simulator study," *Accident Analysis & Prevention*, vol. 147, p. 105760, 2020.

[128] H. Kim, X. Wu, J. L. Gabbard, and N. F. Polys, "Exploring head-up augmented reality interfaces for crash warning systems," in *Proc. of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, (New York, NY, USA), p. 224–227, Association for Computing Machinery, 2013.

[129] T. von Sawitzky, T. Grauschopf, and A. Riener, "No need to slow down! A head-up display based warning system for cyclists for safe passage of parked vehicles," in *Proc. of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, (New York, NY, USA), p. 1–3, Association for Computing Machinery, 2020.

[130] W. H. Pomroy and T. L. Muldoon, *Improved Stench Fire Warning for Underground Mines*, vol. 9016. US Department of the Interior, Bureau of Mines, 1985.

[131] S. Wu, L. Hou, G. K. Zhang, and H. Chen, "Real-time mixed reality-based visual warning for construction workforce safety," *Automation in Construction*, vol. 139, p. 104252, 2022.

[132] H. Mitchell, J. Johnson, and S. LaForce, "Wireless emergency alerts: An accessibility study.," in *ISCRAM*, 2010.

[133] M. Kowal, G. Sandison, L. Yabuki-Soh, and R. la Bastide, "Region tracking in an image sequence: Preventing driver inattention," *CoRR*, vol. abs/1908.08914, 2019.

[134] K. Horak, "Fatigue features based on eye tracking for driver inattention system," in *2011 34th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 593–597, 2011.

[135] T. D'Orazio, M. Leo, C. Guaragnella, and A. Distante, "A visual approach for driver inattention detection," *Pattern Recognition*, vol. 40, no. 8, pp. 2341–2355, 2007. Part Special Issue on Visual Information Processing.

[136] C. Ahlström, K. Kircher, M. Nyström, and B. Wolfe, "Eye tracking in driver attention research—how gaze data interpretations influence what we learn," *Frontiers in Neuroergonomics*, vol. 2, 2021.

[137] K. Kircher and C. Ahlström, "Attentional requirements on cyclists and drivers in urban intersections," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 68, pp. 105–117, 2020.

[138] T. W. Victor, J. L. Harbluk, and J. A. Engström, "Sensitivity of eye-movement measures to in-vehicle task difficulty," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 2, pp. 167–190, 2005. The relationship between distraction and driving performance: towards a test regime for in-vehicle information systems.

[139] K. Kircher and C. Ahlstrom, "Minimum required attention: A human-centered approach to driver inattention," *Human Factors*, vol. 59, no. 3, pp. 471–484, 2017. PMID: 27738279.

[140] C. M. de Melo, K. Kim, N. Norouzi, G. Bruder, and G. Welch, "Reducing cognitive load and improving warfighter problem solving with intelligent virtual assistants," *Frontiers in Psychology*, vol. 11, p. 3170, 2020.

[141] T. Langlotz, J. Sutton, S. Zollmann, Y. Itoh, and H. Regenbrecht, *ChromaGlasses: Computational glasses for compensating colour blindness*, p. 1–12. New York, NY, USA: Association for Computing Machinery, 2018.

[142] A. Erickson, N. Norouzi, K. Kim, R. Schubert, J. Jules, J. J. LaViola, G. Bruder, and G. F. Welch, "Sharing gaze rays for visual target identification tasks in collaborative augmented reality," *Journal on Multimodal User Interfaces*, vol. 14, no. 4, pp. 353–371, 2020.

[143] V. Gopinath, F. Ore, S. Grahn, and K. Johansen, “Safety-focussed design of collaborative assembly station with large industrial robots,” *Procedia Manufacturing*, vol. 25, pp. 503–510, 2018. Proceedings of the 8th Swedish Production Symposium (SPS 2018).

[144] M. Mustafa Kamal, H. Kamarularifin, and C. M. F. Che Mat, “Anti- fatigue rubber mat for industrial workplace: Load pressure analysis at different type of mats,” *Journal of Advanced Research in Applied Mechanics*, vol. 50, p. 12–16, Nov. 2020.

[145] N. Wiggermann and W. M. Keyserling, “Effects of anti-fatigue mats on perceived discomfort and weight-shifting during prolonged standing,” *Human Factors*, vol. 55, no. 4, pp. 764–775, 2013. PMID: 23964416.

[146] M. D. Peck, J. A. Molnar, and D. Swart, “A global plan for burn prevention and care.,” *Bulletin of the World Health Organization*, vol. 87 10, pp. 802–3, 2009.

[147] A. Saidi and C. Gauvin, “Towards real-time thermal stress prediction systems for workers,” *Journal of Thermal Biology*, vol. 113, p. 103405, 2023.

[148] G. K. Mani, Y. Nimura, and K. Tsuchiya, “Advanced artificial electronic skin based ph sensing system for heatstroke detection,” *ACS Sensors*, vol. 5, no. 4, pp. 911–916, 2020. PMID: 32157870.

[149] X. T. Zheng, Z. Yang, L. Sutarlie, M. Thangaveloo, Y. Yu, N. A. B. M. Salleh, J. S. Chin, Z. Xiong, D. L. Becker, X. J. Loh, B. C. K. Tee, and X. Su, “Battery-free and ai-enabled multiplexed sensor patches for wound monitoring,” *Science Advances*, vol. 9, no. 24, p. eadg6670, 2023.

[150] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” 2020.

[151] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *Journal of Big Data*, vol. 3, no. 1, pp. 1–40, 2016.

[152] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft COCO: Common objects in context,” in *Proc. of the European Conference on Computer Vision* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), (Cham), pp. 740–755, Springer International Publishing, 2014.

[153] Dolby Professional, “Dolby atmos for cinema playback.” https://professional.dolby.com/cinema/dolby-atmos, 2023. [Online; accessed 02-June-2023].

[154] A. T. Zijlstra, *Using the HoloLens' Spatial Sound System to aid the Visually Impaired when Navigating Indoors*. PhD thesis, Faculty of Science and Engineering, 2017.

[155] G. Turini, S. Condino, P. D. Parchi, R. M. Viglialoro, N. Piolanti, M. Gesi, M. Ferrari, and V. Ferrari, "A Microsoft HoloLens mixed reality surgical simulator for patient-specific hip arthroplasty training," in *Augmented Reality, Virtual Reality, and Computer Graphics* (L. T. De Paolis and P. Bourdot, eds.), (Cham), pp. 201–210, Springer International Publishing, 2018.

[156] M. Reinecke, "Experimental cognition for qualification of voice warning systems in aircraft," *Deutsche Gesellschaft fuer Luft- und Raumfahrt, Jahrestagung, Munich, West Germany*, 1976.

[157] F. Chen, G. Qvint, and J. Jarlengrip, "Listen! there are other road users close to you: Improve the traffic awareness of truck drivers," in *Proc. of the International Conference on Universal Access in Human-Computer Interaction: Ambient Interaction*, (Berlin, Heidelberg), p. 323–329, Springer-Verlag, 2007.

[158] T. A. Dingus, S. Jahns, A. Horowitz, and R. Knipling, "Human factors design issues for crash avoidance systems," *Human Factors in Intelligent Transportation Systems*, pp. 55–93, 1998.

[159] Z. Yang, J. Shi, Y. Zhang, D. Wang, H. Li, C. Wu, Y. Zhang, and J. Wan, "Head-up display graphic warning system facilitates simulated driving performance," *International Journal of Human–Computer Interaction*, vol. 35, no. 9, pp. 796–803, 2019.

[160] A. Dettmann and A. C. Bullinger, "Spatially distributed visual, auditory and multimodal warning signals–a comparison," *Proc. of the Human Factors and Ergonomics Society Europe*, pp. 185–199, 2017.

[161] M. Rauterberg, "About the importance of auditory alarms during the operation of a plant simulator," *Interacting with Computers*, vol. 10, pp. 31–44, 03 1998.

[162] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, vol. 52, pp. 139–183, North-Holland, 1988.

[163] J. Brooke, "SUS: A quick and dirty usability scale," *Usability Evaluation in Industry*, vol. 189, no. 194, pp. 4–7, 1996.

[164] Y. Liu, N. R. Stiles, and M. Meister, "Augmented reality powers a cognitive assistant for the blind," *eLife*, vol. 7, p. e37841, nov 2018.

[165] T. Hildebrandt, T. Hermann, and S. Rinderle-Ma, "Continuous sonification enhances adequacy of interactions in peripheral process monitoring," *International Journal of Human-Computer Studies*, vol. 95, pp. 54–65, 2016.

[166] J. P. Bliss, R. D. Gilson, and J. E. Deaton, "Human probability matching behaviour in response to alarms of varying reliability," *Ergonomics*, vol. 38, no. 11, pp. 2300–2312, 1995. PMID: 7498189.

[167] X. Li, *Development of an augmented-reality worker guidance in assembly by using Microsoft HoloLens*. PhD thesis, Wien, 2019.

[168] E. Dubrofsky, "Homography estimation," *University of British Columbia (Vancouver)*, vol. 5, 2009.

[169] S.-H. Baek, P. Rathnayaka, and S.-Y. Park, "Calibration of a stereo radiation detection camera using planar homography," *Journal of Sensors*, vol. 2016, 2016.

[170] T. Zhang, L. Hu, L. Li, and D. Navarro-Alarcon, "Towards a multispectral RGB-IR-UV-D vision system — seeing the invisible in 3D," in *Proc. of the IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1723–1728, 2021.

[171] M. Bhattarai, A. R. Jensen-Curtis, and M. Martínez-Ramón, "An embedded deep learning system for augmented reality in firefighting applications," in *Proc. of the IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1224–1230, 2020.

[172] "Pylepton repository in Github website." https://github.com/groupgets/pylepton, 2023. [Online; accessed 02-June-2023].

[173] "OpenCV website." https://opencv.org/, 2023. [Online; accessed 02-June-2023].

[174] H. Bahri, D. Krčmařík, and J. Kočí, "Accurate object detection system on HoloLens using YOLO algorithm," in *Proc. of the International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*, pp. 219–224, 2019.

[175] D. Lloyd-Smith and K. Mendelssohn, "Tolerance limits to radiant heat," *British Medical Journal*, vol. 1, no. 4559, p. 975, 1948.

[176] J. F. Knight and C. Baber, "Neck muscle activity and perceived pain and discomfort due to variations of head load and posture," *Aviation, Space, and Environmental Medicine*, vol. 75, no. 2, pp. 123–131, 2004.

[177] J. F. Knight and C. Baber, "Effect of head-mounted displays on posture," *Human Factors*, vol. 49, no. 5, pp. 797–807, 2007. PMID: 17915598.

[178] P. Phupattanasilp and S.-R. Tong, "Augmented reality in the integrative internet of things (ar-iot): Application for precision farming," *Sustainability*, vol. 11, no. 9, 2019.

[179] S. Davis, K. Nesbitt, and E. Nalivaiko, "A systematic review of cybersickness," in *Proc. of the Conference on Interactive Entertainment*, IE2014, (New York, NY, USA), p. 1–9, Association for Computing Machinery, 2014.

Appendices

BLANK

Appendix A

Standardized questionaries

A.1 NASA-TLX Test Template

The NASA TLX is a widely recognized and validated subjective assessment tool developed by the National Aeronautics and Space Administration (NASA) for evaluating workload in various domains.

The NASA TLX questionnaire consists of six subscales, each considering different aspects of workload. Participants are asked to rate the perceived workload with each subscale on a scale of 0 to 100. A rating of 0 indicates no workload, while a rating of 100 indicates a high workload. The six subscales of the NASA TLX are:

Mental Demand: This subscale assesses the perceived cognitive demands of the task, including mental effort, concentration, and information processing requirements.

Physical Demand: The physical demand subscale measures the perceived physical effort, exertion, and activity level required during task performance.

Temporal Demand: This subscale evaluates the perceived time pressure or urgency associated with completing the task within a specific timeframe.

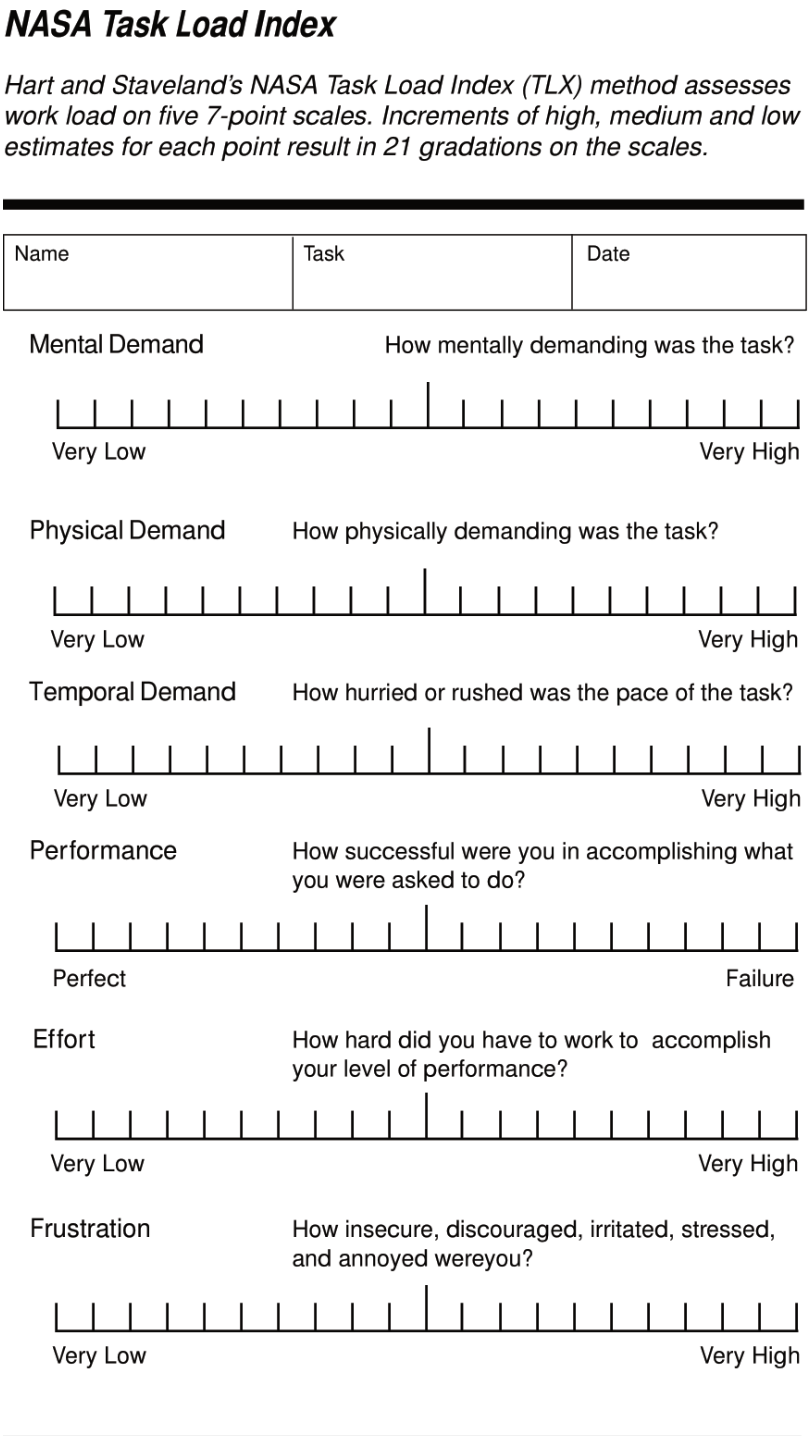
Performance: The performance subscale captures the individual's perception of how successful they were in accomplishing the task goals, including accuracy, efficiency, and overall task performance.

Effort: This subscale assesses the perceived level of effort, exertion, and work intensity expended to complete the task successfully.

Frustration: The frustration subscale measures the level of annoyance, irritation, or dissatisfaction experienced during task performance.

To calculate the overall workload score, ratings obtained from participants in each subscale are combined using a weighting procedure. The TLX uses a pairwise comparison method where participants are asked to compare the relative importance or weight of each subscale against each other. These weight factors, obtained from the pairwise comparisons, are then multiplied by the ratings for each subscale and summed to calculate the overall workload score. This score provides an indication of the perceived mental demands experienced during the task.

It is worth noting that the NASA TLX is a subjective assessment tool, relying on participants' self-report measures. The questionnaire is typically administered after task completion to ensure participants have a comprehensive view of their workload experience.



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure A.1: Paper-and-pencil version of the NASA-TLX questionnaire

In this study, the NASA TLX questionnaire was administered to participants, and their responses were used to evaluate and compare the workload experienced across different conditions or tasks. The inclusion of the NASA TLX in the appendices allows readers to gain a deeper understanding of the methodology employed to assess workload and mental demands, ensuring transparency and providing a basis for future researchers interested in utilizing this assessment tool.

The specific questions and rating scales used in the NASA TLX questionnaire, along with detailed explanations of their meanings and interpretations, are provided in the appendices. Additionally, information about the scoring and analysis procedures, including any modifications or adaptations made to the original TLX, are also documented for reference.

By including the NASA TLX questionnaire in the appendices, this research aims to enhance the transparency and reliability of the workload assessment conducted in the study, facilitating further exploration and understanding of workload dynamics in various domains.

A.2 SUS Test Template

The SUS is a widely used and validated tool for assessing the perceived usability of a wide range of systems, including software, hardware, and user interfaces.

The SUS questionnaire consists of a set of ten items, each designed to measure the participant's perception of different aspects related to usability. Participants are asked to rate their agreement with each item on a five-point Likert scale, ranging from "Strongly Disagree" to "Strongly Agree" (see Fig. A.2). The items cover a range of usability dimensions, including ease of use, learnability, efficiency, and overall user satisfaction.

To calculate the SUS score, the ratings for each item are transformed and summed, resulting in a total score ranging from 0 to 100. A higher score indicates a higher level of perceived usability, while a lower score suggests lower usability.

In this study, the SUS questionnaire was administered to participants to gather their perceptions of the usability of the system or technology being evaluated. The participants' responses were then analyzed to assess the overall usability of the system and identify any areas for improvement.

Including the SUS questionnaire in the appendices allows readers to gain insights into the methodology used to evaluate the usability of the system and provides a transparent and standardized measure of perceived usability. It also serves as a valuable reference for future researchers interested in assessing the usability of similar systems or technologies.

The specific questions and rating scales used in the SUS questionnaire, along with detailed explanations of their meanings and interpretations, are provided in the appendices. Additionally, information about the scoring and analysis procedures, including any adaptations or modifications made to the original SUS, are also documented for reference.

By including the SUS questionnaire in the appendices, this research aims to enhance the transparency and reproducibility of the usability assessment conducted in the study,

System Usability Scale Questionnaire	**Strongly Disagree**				**Strongly Agree**
1. I think that I would like to use this product frequently.	1	2	3	4	5
2. I found the product unnecessarily complex.	1	2	3	4	5
3. I thought the product was easy to use.	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this product.	1	2	3	4	5
5. I found the various functions in the product were well integrated.	1	2	3	4	5
6. I thought there was too much inconsistency in this product.	1	2	3	4	5
7. I imagine that most people would learn to use this product very quickly.	1	2	3	4	5
8. I found the product very awkward to use.	1	2	3	4	5
9. I felt very confident using the product.	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this product.	1	2	3	4	5

Figure A.2: Paper-and-pencil version of the SUS questionnaire

facilitating further exploration and understanding of usability factors in the context of the investigated system or technology.

A.3 Questionnaires for Warning Types Evaluation

In this appendix, we provide questionnaires used to evaluate the three warning types (auditory, visual, and audiovisual) and general AR-based warning systems. The questions for each warning type were derived from both the System Usability Scale (SUS) and NASA Task Load Index (TLX). The questionnaire for the general AR-based warning system was designed to assess participants' preferences, desire to use augmented reality (AR), perceptions of cumbersomeness related to AR glasses, the need to learn new things to use AR, and the need for technical support when using AR.

The questionnaires consisted of a combination of Likert-scale questions and ranking questions to gather participants' opinions and preferences regarding the different warning types.

Questionnaire for Warning System Experiments

Date:_______

Time:_______

Subject #____

Gender: ☐ Male ☐ Female ☐ Other

Age:____

Occupation:__________

Have you ever used an AR/VR headset?

☐ Never ☐ Once or twice ☐ Often ☐ I already own one of them

Have you ever worked with industrial machines?

☐ Never ☐ Once or twice ☐ Often ☐ I already own one of them

Date:_______
Time:_______
Subject #____

a) Auditory warning

1) I felt very confident using the auditory warning system (Confidence).

(not confident) □1 □2 □3 □4 □5 (very confident)

2) I found the auditory warning intuitive (Intuitiveness).

(not intuitive) □1 □2 □3 □4 □5 (very intuitive)

3) Tasks with an auditory warning system were mentally demanding (Mental demand).

(very low) □1 □2 □3 □4 □5 (very high)

4) I find the system beneficial for the working process (Beneficiality).

(disadvantageous) □1 □2 □3 □4 □5 (advantageous)

5) I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system
(Frustration).

(very low) □1 □2 □3 □4 □5 (very high)

Date:_______
Time:_______
Subject #____

b) Visual warning

1) I felt very confident using the visual warning system (Confidence).

(not confident) ☐1 ☐2 ☐3 ☐4 ☐5 (very confident)

2) I found the visual warning intuitive (Intuitiveness).

(not intuitive) ☐1 ☐2 ☐3 ☐4 ☐5 (very intuitive)

3) The visual warning task was mentally demanding (Mental demand).

(very low) ☐1 ☐2 ☐3 ☐4 ☐5 (very high)

4) I find the system beneficial for the working process (Beneficiality).

(disadvantageous) ☐1 ☐2 ☐3 ☐4 ☐5 (advantageous)

5) I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).

(very low) ☐1 ☐2 ☐3 ☐4 ☐5 (very high)

Date:_______
Time:_______
Subject #____

c) Auditory-visual warning

1) I felt very confident using the auditory-visual warning system (Confidence).

(not confident) ☐1 ☐2 ☐3 ☐4 ☐5 (very confident)

2) I found the auditory-visual warnings intuitive (Intuitiveness).

(not intuitive) ☐1 ☐2 ☐3 ☐4 ☐5 (very intuitive)

3) The auditory-visual warning task was mentally demanding (Mental demand).

(very low) ☐1 ☐2 ☐3 ☐4 ☐5 (very high)

4) I find the system beneficial for the working process (Beneficiality).

(disadvantageous) ☐1 ☐2 ☐3 ☐4 ☐5 (advantageous)

5) I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).

(very low) ☐1 ☐2 ☐3 ☐4 ☐5 (very high)

Date:_______
Time:_______
Subject #____

General questions

1) Which warning do you prefer? Sort them in order of preference (Preference).
a. Auditory Warning b. Visual Warning
c. Auditory-Visual Warning

(most preferred) 1.__ 2.__ 3.__ (least preferred)

2) I would like to use the AR smart glasses for safety in the working environment (Desire to use AR).

(strongly disagree) ☐1 ☐2 ☐3 ☐4 ☐5 (strongly agree)

3) The smart glasses are cumbersome to use. (Cumbersomeness)

(convenient) ☐1☐2 ☐3 ☐4 ☐5 (bulky)

4) I needed to learn a lot of things before I could get going with this system (Learning new things).

(strongly disagree) ☐1 ☐2 ☐3 ☐4 ☐5 (strongly agree)

5) I think that I would need the support of a technical person to be able to use this system (Technical support).

(strongly disagree) ☐1 ☐2 ☐3 ☐4 ☐5 (strongly agree)

BLANK

Appendix B

Results Experiment 2

B.1 Quantitative results of Experiment 2

	Participant	1	2	3	4	5	6	7	8	9	10
	Gender	m	m	m	f	f	f	f	f	m	m
	Age	34	26	23	19	20	27	28	31	30	20
a) Auditory warning											
1	*I felt very confident using the auditory warning system (Confidence).*	4	4	4	5	5	5	5	5	3	5
2	*I found the auditory warning intuitive (Intuitiveness).*	5	4	5	5	5	5	5	5	3	4
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	3	1	3	1	1	1	2	1	3	1
4	*I find the system beneficial for the working process (Beneficiality).*	4	3	5	5	4	5	5	5	3	4
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	2	2	2	1	2	1	1	1	1	1
b) Visual warning											
1	*I felt very confident using the auditory warning system (Confidence).*	5	5	5	5	5	5	1	4	5	5
2	*I found the auditory warning intuitive (Intuitiveness).*	5	4	5	5	4	5	5	4	5	4
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	2	1	1	1	1	1	3	2	1	2
4	*I find the system beneficial for the working process (Beneficiality).*	4	3	5	5	4	5	5	5	5	4
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	1	2	1	1	1	1	1	1	1	1
c) Auditory-visual warning											
1	*I felt very confident using the auditory warning system (Confidence).*	5	5	5	5	5	5	2	5	3	5
2	*I found the auditory warning intuitive (Intuitiveness).*	5	4	5	5	5	5	4	5	5	4
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	2	1	1	1	1	1	4	1	3	1
4	*I find the system beneficial for the working process (Beneficiality).*	4	3	5	5	4	5	4	5	5	5
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	2	3	2	1	1	2	2	1	1	1
General questions											
1	*Which warning do you prefer? Sort them in order of preference (Preference). a. Auditory Warning b. Visual Warning c. Auditory-Visual Warning*										
	1st order preference	b	c	c	c	b	b	a	a	c	c
	2nd order preference	c	b	a	b	c	c	b	c	b	a
	3rd order preference	a	a	b	a	a	a	c	b	a	b
2	*I would like to use the AR smart glasses for safety in the working environment (Desire to use AR)*	2	2	5	5	4	5	4	2	5	4
3	*The smart glasses are cumbersome to use. (Cumbersomeness)*	5	4	3	2	2	1	2	2	3	2
4	*I needed to learn a lot of things before I could get going with this system (Learning new things).*	2	3	1	3	4	1	3	1	2	1
5	*I think that I would need the support of a technical person to be able to use this system (Technical support)*	1	4	2	3	4	3	5	2	3	2

	Participant	11	12	13	14	15	16	17	18	19	20
	Gender	m	m	m	m	m	m	m	f	f	f
	Age	40	33	29	29	31	21	29	21	20	20
a) Auditory warning											
1	*I felt very confident using the auditory warning system (Confidence).*	5	4	5	4	5	5	5	5	5	5
2	*I found the auditory warning intuitive (Intuitiveness).*	4	4	5	4	5	5	5	5	5	5
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	1	4	1	2	4	1	1	2	2	1
4	*I find the system beneficial for the working process (Beneficiality).*	4	5	4	4	5	4	5	4	5	5
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	1	2	1	3	3	2	1	4	2	1
b) Visual warning											
1	*I felt very confident using the auditory warning system (Confidence).*	5	3	5	5	5	5	5	5	5	5
2	*I found the auditory warning intuitive (Intuitiveness).*	5	4	5	5	5	5	5	5	5	5
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	1	4	1	1	4	1	1	1	1	1
4	*I find the system beneficial for the working process (Beneficiality).*	5	5	5	4	4	4	5	5	5	5
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	1	2	1	1	4	1	1	1	1	1
c) Auditory-visual warning											
1	*I felt very confident using the auditory warning system (Confidence).*	5	4	5	5	5	5	5	5	5	5
2	*I found the auditory warning intuitive (Intuitiveness).*	3	4	5	5	5	5	5	5	5	5
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	3	4	1	1	5	1	2	2	1	1
4	*I find the system beneficial for the working process (Beneficiality).*	3	5	5	5	5	5	5	4	5	5
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	1	3	1	3	3	2	1	3	1	1
General questions											
	Which warning do you prefer? Sort them in order of preference (Preference). a. Auditory Warning b. Visual Warning c. Auditory-Visual Warning										
	1st order preference	b	c	c	c	c	b	a	b	c	c
	2nd order preference	a	b	b	b	b	c	c	c	a	a
1	*3rd order preference*	c	a	a	a	a	a	b	a	b	b
2	*I would like to use the AR smart glasses for safety in the working environment (Desire to use AR)*	5	4	4	5	5	5	5	5	5	4
3	*The smart glasses are cumbersome to use. (Cumbersomeness)*	1	3	2	2	4	3	1	2	1	2
4	*I needed to learn a lot of things before I could get going with this system (Learning new things).*	1	5	1	3	1	1	1	3	1	4
5	*I think that I would need the support of a technical person to be able to use this system (Technical support)*	1	4	1	2	5	2	1	2	2	4

	Participant	21	22	23	24	25	26	27	28	29	30
	Gender	f	f	f	f	f	f	m	m	m	f
	Age	19	22	19	19	19	23	20	22	21	33
a) Auditory warning											
1	*I felt very confident using the auditory warning system (Confidence).*	5	5	5	4	4	5	5	5	5	5
2	*I found the auditory warning intuitive (Intuitiveness).*	4	4	5	5	3	5	4	5	5	5
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	2	1	1	1	2	1	2	1	1	5
4	*I find the system beneficial for the working process (Beneficiality).*	3	4	5	4	3	5	4	3	5	5
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	5	1	4	1	2	2	1	3	1	1
b) Visual warning											
1	*I felt very confident using the auditory warning system (Confidence).*	5	4	5	5	5	5	5	5	5	5
2	*I found the auditory warning intuitive (Intuitiveness).*	5	3	5	5	4	5	5	5	5	5
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	1	1	2	1	1	1	1	1	1	3
4	*I find the system beneficial for the working process (Beneficiality).*	5	3	5	5	4	5	5	5	4	5
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	1	1	1	1	1	1	1	1	1	2
c) Auditory-visual warning											
1	*I felt very confident using the auditory warning system (Confidence).*	5	5	4	5	2	5	5	5	5	5
2	*I found the auditory warning intuitive (Intuitiveness).*	4	3	5	5	5	5	4	5	5	5
3	*Tasks with an auditory warning system were mentally demanding (Mental demand).*	1	1	3	1	3	1	1	1	1	4
4	*I find the system beneficial for the working process (Beneficiality).*	2	4	5	4	5	5	5	5	5	5
5	*I felt insecure, discouraged, irritated, stressed, and annoyed using the warning system (Frustration).*	5	1	2	1	2	2	1	3	1	2
General questions											
	Which warning do you prefer? Sort them in order of preference (Preference). a. Auditory Warning b. Visual Warning c. Auditory-Visual Warning										
	1st order preference	b	a	c	c	b	c	c	b	b	c
	2nd order preference	c	c	b	b	a	a	b	c	c	b
1	*3rd order preference*	a	b	a	a	c	b	a	a	a	a
2	*I would like to use the AR smart glasses for safety in the working environment (Desire to use AR)*	2	4	4	4	4	4	4	2	3	5
3	*The smart glasses are cumbersome to use. (Cumbersomeness)*	3	2	4	2	3	4	2	2	4	3
4	*I needed to learn a lot of things before I could get going with this system (Learning new things).*	5	2	2	1	3	3	1	1	1	2
5	*I think that I would need the support of a technical person to be able to use this system (Technical support)*	5	2	4	2	4	5	4	1	2	1

B.1.1 Experiment 2 Interview Scripts

Participant 1: *The dark visor of the smart glasses hinders me from acclimating to the working process.*

Participant 2: *I think holograms should be more transparent to not obscure the vision and on-site camera-based eye-tracking could be more beneficial than smart glasses' eye tracking.*

Participant 3: *The visual warning doesn't cause urgency, while the auditory warning doesn't give enough information. Merging both warning types could make it both urgent and informative. I think that long-term use of AR goggles can cause frustration.*

Participant 5: *I would rather increase the transparency of the hologram because I have difficulty seeing the machine underneath it clearly.*

Participant 6: *The hologram's size should be reduced, and it should be positioned on a smaller concrete area within the working space. The up and down arrows are not appearing in a timely manner.*

Participant 7: *By increasing the opacity of the cylinder hologram, users would be better able to detect their workplace and orient themselves more effectively.*

Participant 8: *I would rather have a box in front of my eyes instead of arrows, where a green box indicates no hazards and a red box indicates hazards.*

Participant 9: *Speech warning "Look at the hologram" would be more intuitive instead of a beep warning.*

Participant 10: *Completely obstructing the user's FOV when they look away from the hologram could aid in better concentration on the work process.*

Participant 14: *I recommend you add speech warnings in Kazakh from the Kazakh speech corpus and draw directions with several arrows instead of one single arrow.*

Participant 16: *Auditory warning signals can be replaced by more pleasing-to-ear ones.*

Participant 19: *I think it would be better to use more observable shades of color than the ones already used. When the object is more observable, it would be better.*

Participant 26: *Arrows should follow smoothly and warn even when we look at the hologram.*

Participant 30: *Changing the color of the arrow to red would cause urgency and decrease reaction time.*

Appendix C

Source Code

C.1 Script for selecting a warning type

```
%[caption=Script for selecting a warning type]
using System.Collections;
using System.Collections.Generic;
using UnityEngine;

public class SelectWarningSystem : MonoBehaviour
{
    public GameObject cylinder;
    public GameObject arrow;

    // Visual Warning
    public void VisualWarning()
    {
        cylinder.SetActive(false);
        cylinder.SetActive(true);
        arrow.SetActive(false);
        CameraAndRecognizedPos.turnArrow = true;
        CameraAndRecognizedPos.turnAudio = false;
    }

    // Auditory Warning
    public void AuditoryWarning()
    {
        cylinder.SetActive(false);
        cylinder.SetActive(true);
        arrow.SetActive(false);
        CameraAndRecognizedPos.turnArrow = false;
        CameraAndRecognizedPos.turnAudio = true;
    }

    // Auditory Visual Warning
    public void AuditoryVisualWarning()
    {
        cylinder.SetActive(false);
        cylinder.SetActive(true);
        arrow.SetActive(false);
        CameraAndRecognizedPos.turnArrow = true;
```

```
        CameraAndRecognizedPos.turnAudio = true;
    }
}
```

C.2 Script for activating/deactivating auditory and visual warnings

```
%[caption=Script for activating/deactivating beep sound and indication of holographic arrow]
using System.Collections;
using System.Collections.Generic;
using UnityEngine;

public class audio : MonoBehaviour
{
    public AudioSource Beep;
    public GameObject arrow;
    private bool update;
    public GameObject cylinder;
    public Vector3 dir;
    public Vector3 point;
    public Vector3 point2;
    public Vector3 angles;
    //start beep sound
    public void PlayBeep()
    {
        if (CameraAndRecognizedPos.turnAudio) { Beep.Play(); }
        if (CameraAndRecognizedPos.turnArrow)
        {
            arrow.SetActive(true);
        }
        if (cylinder.active)
        {
            update = true;
        }
    }
    private void Update()
    {
        //calculates the orientation of holographic arrow
        if (update == true )
        {
            dir = cylinder.transform.position - Camera.main.transform.position;
            point = Camera.main.transform.forward * dir.magnitude;
            Vector3 dir2 = point - cylinder.transform.position;

            Quaternion rotation = Quaternion.LookRotation(dir2);
```

```
            arrow.transform.rotation = rotation;
        }
        else if (!CameraAndRecognizedPos.turnArrow) {
            arrow.SetActive(false);
        }
    }
    //stop beep sound
    public void StopBeep()
    {
        update = false;
        Beep.Stop();
        arrow.SetActive(false);
    }
}
```