Tampere University

Harri Halonen

INTERACTION DESIGN PRINCIPLES FOR INDUSTRIAL XR

Faculty of Information Technology and Communication Sciences
M. Sc. Thesis
May 2021

ABSTRACT

Harri Halonen: Interaction Design Principles for Industrial XR
M.Sc. Thesis
Tampere University
Master's Degree Programme in Human-Technology Interaction
May 2021

Convenient access to task-relevant information in a robust and unobtrusive manner allows an industrial worker to perform their duties with high efficiency, saving valuable time at a workforce scale. Wearable augmented reality display devices are compelling alternatives to handheld devices as they free up user's hands and allow easy consumption of information.

This thesis takes a grounded theory approach to investigate interaction designer's perceptions of industrial worker's needs and accompanying interaction requirements regarding the use of XR technologies in industrial environments. The purpose of the qualitative study is to increase understanding by presenting crucial insights and practices leading to the swifter adoption of XR technologies to industrial field work and recommendations of interaction techniques most suitable to industrial environments. Results consist of semi-structured expert interviews to understand the domain-specific opportunities and limitations.

Discovered grounded theory calls for a usable and robust hands-free and touchless freehand operation of a readable display supported by an eyes-free output option. When producing practical XR solutions, the most impactful interaction techniques deliver the right information at the right time, ensuring sufficient safety, comfort, and efficiency of task performance on a reasonable cost-benefit ratio. Interaction designer's role is to support and empower the industrial professional, with low tolerance towards nonfunctioning tools, to focus on the real task at hand.

Key words and terms: XR, Augmented reality, Interaction design, Industrial field work.

PREFACE

I am grateful to the industry professionals who agreed to take part in the interviews. I want to thank my supervisor, professor Markku Turunen for guidance and support throughout the thesis process. I also want to thank the computer sciences head of study services Heli Rikala. Without all of you, the completion of this thesis would not have been possible.

Thank you to my dad for always believing me to get the job done, my friends for doubting if I would, Kida for distracting me, Niko for showing up whenever I most needed it, and Miia for showing me how to commit hard to a goal. Special thank you to my mother for the care, circumstances in life, and motivation that even a learner like me can and should pursue higher learning; thank you for everything you have had to endure for it.

Tampere, 18.5.2021

Contents

1 Introduction

An industrial organization aims for operational efficiency and safety for their or their customer's procedures. The efficiency of operations is primarily determined by the procedures and competency of the workforce, but it also can be affected by the technological state of operations brought about by the Industry 4.0 transformation. The fourth industrial revolution of automatization of conventional manufacturing brings forward vast amounts of computing and information sources available to the industrial worker in an attempt to unlock an increase in productivity. Outside of industrial facilities, companies have been actively seeking methods to digitalize their mobile maintenance workforce. Accessing helpful knowledge to an individual professional's duties requires them to use a computer or a handheld device, but these are not always readily at hand. Augmented reality head-worn displays like Microsoft HoloLens offer the potential to tie information into a location for more ubiquitous access [Microsoft, 2019a]. Simultaneously, heavy and specialized computing devices used by industrial field maintenance personnel are giving away to smartphones and lighter head-worn displays offered by brands like RealWear and Vuzix [RealWear, 2020; Vuzix, 2020].

However, computing solutions always carry entry barriers for enterprises as prototypes might get stuck at a proof of concept stage for a multitude of reasons and end up not being deployed to the workforce's benefit. The researcher's motivation to learn about the practical use of the technologies mentioned above was due to their portrayal in the technology press. New devices come with complexity both in the back end and front end, but one of the most crucial issues is still end-user adoption. User experience goals have been proposed to guide the design of industrial systems [Kaasinen et al., 2015], but interaction techniques have not been discussed as much.

Industrial field workers' jobs are physically demanding both in assembly and as maintenance technicians. Field work requires both moving and standing for prolonged periods in manufacturing floors that may be hot, dusty, and noisy. Maintenance technicians are needed both day and night to attend to machinery in plants, factories, or similar settings but also equipment like elevators in semipublic spaces. In industrial plants, they work in long eight or twelve-hour shifts, keeping equipment in good working order and swiftly fixing problems to meet production goals. To perform their job, they need a suitable amount of information about it and means to report their progress and actions taken.

The utilization of interaction techniques is a balance between the needs of two stakeholders. Keeping in mind the industrial companies need to ensure business continuity and the employee's needs, there are still few unanswered questions. What kind of interaction techniques is the user comfortable using but can also function efficiently in challenging industrial environments? What is the appropriate relationship between immersive and less

immersive features? What kind of aspects of the aforementioned technology solutions should be taken note of in general to meet the goals for practical use?

The topic has so far been addressed through studies that generally elucidate the success factors of industrial augmented reality implementations [Masood and Egger, 2019], studies based on theoretical knowledge of interaction design attributes from a maintenance technician-centered perspective [del Amo et al., 2018] or empirical studies of human factors and ergonomics of a particular concept [Aromaa et al., 2018]. However, no direct research data covering the perceptions of suitable interaction methods in the industrial context from research and development professionals working for industrial companies was found.

This study aims to learn about the reasoning behind interaction design decisions of interaction techniques utilized in modern industrial conditions and develop understanding of their use in industrial field conditions. The success factors of the interaction techniques are sought in this study through semi-structured interviews with several individuals who have recently been a part of the development of applications aimed to be used in field environments and have observed the industrial workers using previous applications. Typically, the interface solutions used in industrial work refer to smartphones, but recently a variety of head-worn displays have become commercially available. Therefore, interviews also seek whether and how head-worn or hands-free displays add value to the user and what kind of attitudes industrial corporations have towards utilizing such display solutions. Another interview aim is to understand the suitability characteristics of more emerging input technologies for industrial environments and how they meet the needs of employees working in the field. Analyzed interview findings will guide a supplementary literature review into the emerging input interface research in the form of a theory.

Keeping the focus on interaction, this thesis will touch on but not cover user experience, usability, graphical user interface, or technical challenges related to interface devices. Instead, this thesis approaches interaction from the direction that interaction techniques suitable for industrial field work could be reused for other, potentially less demanding, use scenarios within the industrial sector.

The structure of this thesis tries to keep the grounded theory approach within the faculty guidelines and is as follows. Chapter 2 defines the fundamental concepts and terminology regarding XR and discusses immersive display styles and object interaction. Chapter 3 goes through the study's methodology, including grounded theory and the interview as a data collection method, and Chapter 4 presents the interview findings in categories related to the main theme. Chapter 5 discusses the findings, their implications, limitations of the study, and Chapter 6 summarizes the conducted study.

2 Theoretical Framework

This minor literature review examines basic terminology in the research area. Section 2.1 discusses the language surrounding the concept of XR. Section 2.2 presents immersive computing image generation solutions and briefly reviews the technology of head-mounted displays. Section 2.3 introduces object interaction and mid-air gesturing.

2.1 Terminology of real-virtual continuum

Virtual Environment (VE) is a three-dimensional (3D) computer-generated environment. Viewing VE is possible with a flat computer panel, but it can also be presented to a user with a closed-view head-mounted display (HMD). While wearing it powered-on, the user cannot see the natural world around them as its vision has been entirely replaced with the VE. If this egocentric viewport into VE moves in 3D space, called egomotion, the user feels they are in the center of space or immersed in a Virtual Reality (VR).

VR can be extended beyond vision to other senses like hearing, skin pressure sensitivity, vibrations, temperature, and proprioception. VR already disrupts the user's normal sight, but sensation can be added by moving the user's viewpoint without a connection to the real-world movement of the head. The user's perceived point of view of VE can be altered. For example, switch between a first-person perspective where the user can interact with a steering wheel, a third-person view of the vehicle driving on the road, or a bird's eye view where the user can see their avatar's current location on a map. While a person is wearing a VR HMD, it can have disruptive effects on the user's spatial judgments of distance, orientation, and movement of objects in the unseen real world.

In comparison, augmented reality (AR) can be a lot less disruptive. It is characterized by the combination of real and virtual information. The first AR implementations can be considered to have been stage illusions accomplished with light, mirrors, and panes of glass [Wikipedia, 2021a]. The principles of this illusion have stayed the same throughout the years. AR display enables the user to observe the natural world, but with virtual objects composited or superimposed on it [Azuma, 1997].

All immersive environments can be thought of as belonging to a spectrum rather than being their own separate entities. The reality-virtuality continuum is extended from reality to virtual environments via the intermediate states of augmented reality and augmented virtuality [Milgram et al., 1995]. The reality in the continuum is a mediatized representation of natural reality, which the method of capture neither augments nor diminishes. Mixed reality (MR) applications can exist sliding along the reality-virtuality continuum by mixing real and virtual together in proportion [Mann et al., 2018]. X Reality (XR), similar to MR but less strictly defined, is an umbrella term for all realities. X is representative of a mathematical variable used to define a continuum [Mann et al., 2018]. Portrayed this way, X can stand for all technology combinations of virtuality and reality, even flat digital experiences.

Both virtual reality and augmented reality can be described to be mixed reality. Microsoft Corporation's marketing demonstrates this by branding both the AR system HoloLens and a series of VR HMDs, as Windows Mixed Reality [Microsoft, 2019a]. The use of MR in a marketing term has led to the term hinting at it being a better form of AR where virtual objects integrate, occlude, and interact with passive and dynamic elements of the real environment surrounding the user [Unity, 2020]. This thesis accepts this terminology in part. Using the term augmented to refer to a view where digital objects are placed on top of ordinary reality on a separate layer and generally do not have visual anchoring in the real world, like often seen with AR HMDs called smart glasses. To avoid the potentially messy terminology, the term MR will not be used to describe devices like Microsoft HoloLens 1 and 2, capable of using time-of-flight sensors and artificial intelligence to perform real-time semantic spatial mapping to provide accurate hologram or reality interaction. This thesis will instead refer to them and those like them by their brand name HoloLens. XR technology or immersive technology will be used as an umbrella term for all VR, AV, AR, and flat technologies extended beyond the traditional handheld and desktop computing devices.

2.2 Immersive image-generation solutions

An AR system obtains raw data of the real world via sensors processing it to extract necessary information in order to be combined with computer-generated data, such as illuminated 3D objects, returning the result to reality for inspection on display. Consequently, AR requires the three main subsystems: a tracking and sensing, scene generator, and display device [Azuma, 1997]. For the observer, there are two available methods of combining real and virtual, the optical see-through and video see-through technologies [Rolland et al., 1995]. The optical see-through technique allows the user's eyesight to view the natural environment directly through a partially transmissive material. This optical combiner reflects some of the light guided to it from the image source to it into the user's eye. Reducing the amount of light from the natural environment in the process, but only to the degree that delivering a black image is ineffective in a bright environment [Azuma, 1997]. Video see-through technology uses a camera to capture the environment and computing resources to add virtual information into the mix before presenting the result to the user, leading to a slight delay with the see-through video method compared to the see-through optical implementation [Azuma, 1997]. This delay can buy time to simulate photorealistic virtual objects with color contrast, depicted in Figure 1.

Figure 1. Marketing image for Varjo XR-1 illustrating the difference between video see-through and optical see-through methods [Varjo, 2020].

In Figure 1, on the left side, the car appears relatively complete, total opacity, and photorealistic, simulating even a cast shadow on the environment, while on the right-side virtual content appears translucent and experienced only in a boxy viewport leaving the image incomplete. The view on the left can be thought of as being completely digital, as the presented video represents reality limited by the intrinsic qualities of the capturing device and the screen. Figure 2 by Bimber and Raskar [2006] depicts the diverse options available for an observer to experience an AR image.

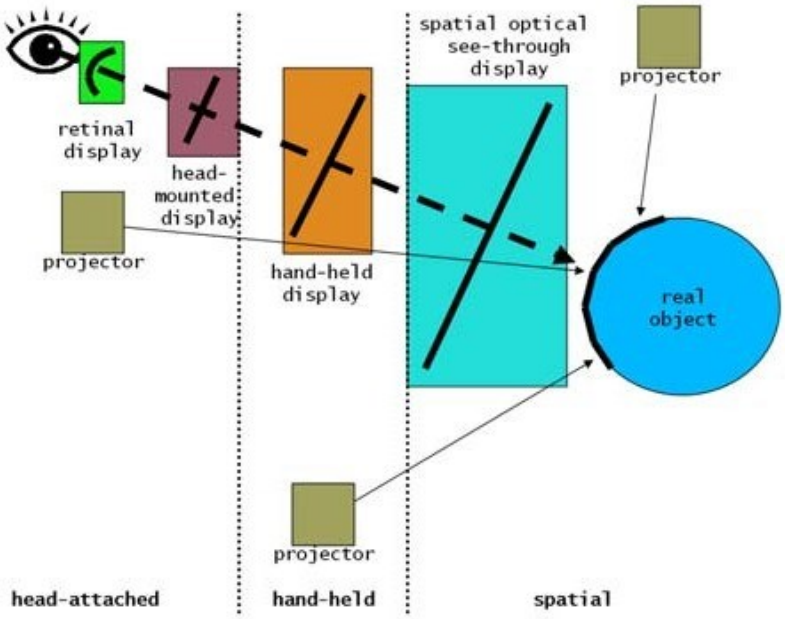

Figure 2. Different ways of image generation for AR displays [Bimber and Raskar, 2006].

The image generation for AR displays from the Figure 2 can be interpreted as follows:

1. Head-attached displays
    (a) Head-mounted optical or video see-through
    (b) Retinal curved optical see-through
2. Handheld displays
    (a) Video see-through
    (b) Optical see-through
3. Spatial displays
    (a) Projection, with fixed, handheld, or head-worn projector, directly onto a physical object or to an optical see-through material partially reflecting light
    (c) Optical see-through panel
    (d) Video see-through panel

A few of the characteristics that visual displays have in common are the field of regard (FOR) and view (FOV), refresh rate, and spatial resolution. The FOR is the visual angle measurement of the user's physical space. FOV is the maximum visual angle user observes. Spatial resolution is the dots per inch quality measurement given by the pixel size. Depending on the use case of the observer, the AR display device form factor can matter a great deal. The various image generation methods affect how much information can be displayed to the observer and how widely distributed the data is to the FOR. Head-attached displays move along with the user's line of sight and leave a hand free from holding a handheld AR display.

When serving an image to both eyes, head-attached displays can be biocular by serving two identical images or binocular with its own image for each eye. In HMDs, only with the optimum alignment of the display elements, the transmitted image can be visible on a comfortable focus. On closed-view binocular video pass-through HMD's this is done by adjusting the distance of magnifying and shaping binocular lenses between the eyes and the near-eye displays. Optical see-through HMDs require adjusting the angle and distance of the optical combiner from the eye. However, with binocular HMDs, the whole region of the environment within users' FOV is perceived clearly in the same focus and depth of field or focal planes, seen previously only in television cartoon shows.

In comparison, in natural environments, eyes use accommodation to discriminate objects at different distances while sweeping across the scene can perceive an area that is much larger than the foveal acuity in the center of the vision [Cutting and Vishton, 1995]. When a user is looking into the distance in the natural environment and a virtual object is suddenly added into the user's binocular HMD viewport to a closer depth. Consequently, the eyes need to shift focus fast between the background environment and the virtual object presented, blurring the other in the process, leading to a short period of double

sight called accommodation-convergence conflict, causing viewer discomfort and fatigue in prolonged use [Shibata et al., 2011].

Monocular head-attached displays provide only one image for a single eye and do not cause accommodation-convergence conflict. Commercial implementations of see-through monocular waveguide optics head-attached displays are often marketed as smart glasses [Vuzix, 2020]. A head-attached monocular display can also be a see-through video display. RealWear HMT-1 claims to be "the world's first hands-free Android tablet class wearable computer for industrial workers" [Realwear, 2020]. There can also be no display at all. Virtual retinal display technology uses a micro-electro-mechanical system module. A scanner produces a low-power laser light beam to be reflected by an oscillating micromirror in points to a curved holographic optical element embedded in the lens surface of an otherwise ordinary pair of glasses [Bimber and Raskar, 2006]. From it, the light is redirected to the lens of the human eye to be combined onto the human retina surface.

2.3 XR input techniques

At the center of interaction with a computer is a human performing a task. In the natural environment, the skills and knowledge to perform both ordinary and demanding tasks are gathered along months or even years of practice and used to transform complex interaction into a natural and intuitive action. A human user's multisensory perceptions lead them to choose an action to perform. When executed, this creates feedback for the human to evaluate and plan the next action. Here a user interface (UI) is the mediator between the user and the computer system on which actions are executed as commands and feedback generated to be returned via the UI in an interaction loop. Wikipedia defines an interaction technique to be "a combination of hardware and software elements that provides a way for computer users to accomplish a single task" [Wikipedia, 2019].

Three major categories for object interaction techniques in VE are direct user interaction, physical controls, and virtual controls [Mine, 1995]. For a computer to understand an interaction technique's intended meaning, it requires these control schemes. Direct user interaction can be thought of as tracked real-world action, like head or hand movement, bound directly into the virtual elements. With physical controls, interaction happens in real environment physical objects with a surface level binding to virtual elements. Virtual controls are indirect control by virtual UI elements, like widgets and panels, separate and independent of virtual elements they are in control.

Interaction for immersive technologies draws a lot from VE interaction. In VE, the environment is, at least in significant parts, the graphical user interface. Object interaction on graphical user interfaces of XR displays relies on similar interaction modes. Outside of movement control, the primary interaction modes upon object interaction relies on, in VE are selection, manipulation, and scaling [Mine, 1995]. Object interaction task is composed of at least two phases, selection and manipulation [Bowman and Hodges, 1997].

Figure 3 by Bowman et al. [1999] presents a comprehensive breakdown of a selection and manipulation task in VE fitted with, at the time, known interaction techniques for them in categories.

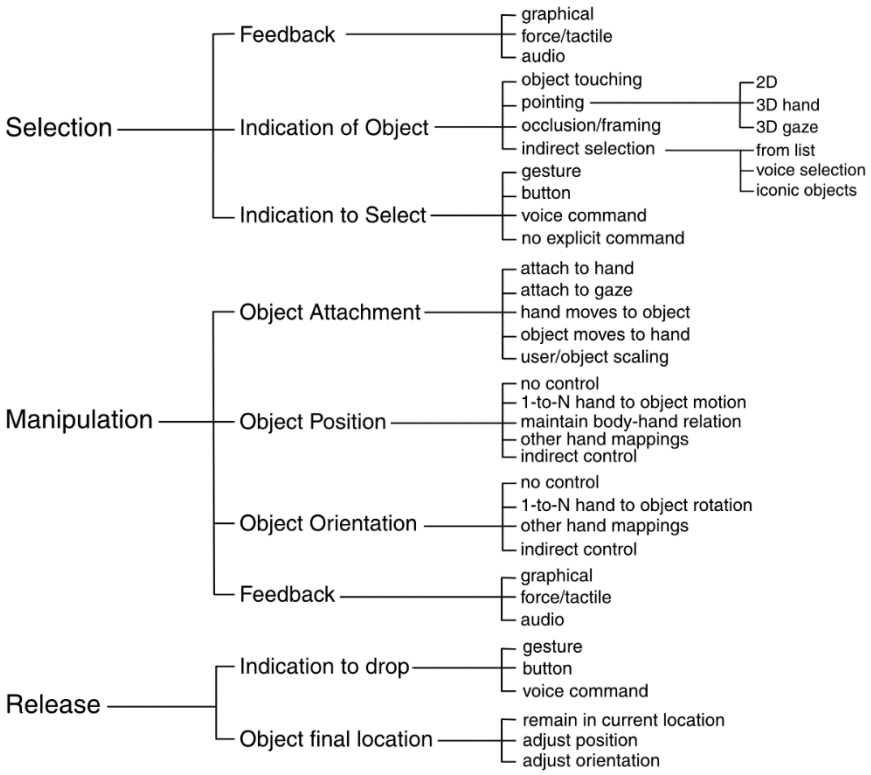

Figure 3. Taxonomy of selection/manipulation techniques for VE [Bowman et al., 1999].

Figure 3 shows multiple modalities that a user can use to perform a selection/manipulation task if an appropriate input device is available. An excellent example of a single platform supporting all Indication to Select –techniques in Figure 3 is the Microsoft HoloLens. The supported selection commit methods for HoloLens products are the "air tap" mid-air hand gesture, a voice command, pressing a controller button, or gaze and dwell, where the user keeps looking at the target for a while to select it [Microsoft, 2019b].

What seems like direct interaction to the user can, in actuality, be a combination of direct interaction and virtual controls. In the "direct manipulation" model on Microsoft HoloLens 1 and 2, 3D objects are handled through affordances of not the object but a virtual bounding box around the object [Microsoft, 2019c]. These virtual controls allow the virtual object to be scaled and manipulated in six degrees of freedom (DoFs) by moving it through space and rotating around its center [Microsoft, 2019c]. DoF of an interaction technique describes the axis on which the selection/manipulation tool can be moved, turned, and rotated around. DoF required for accomplishing a task can signal the interaction technique or input device needed. For example, 6-DoF tracked HMD allows to de-

termine whether a user has moved and rotated their head as tracking of translational motion as well as rotational motion is possible. Before optical hand tracking solutions like Leap Motion were commercially available, VR controllers popularized pointing selection and three-dimensional object manipulation through 6-DoF handheld controllers. The number of DoFs of an interaction technique can be thought of as allowing expressivity or potentially adding complexity to the task performance.

While the taxonomy of Figure 3 is for selection/manipulation of a 3D object in VE, it should be kept in mind that smart glasses style devices are starting to inhabit the same device space as binocular HMDs (see section 2.2) and are being referred to as MR or XR devices (see section 2.1). Thus, the modalities involved with object interaction are served through the same input devices. Tung et al. [2015] classified interaction approaches of smart glasses into handheld, touch, and non-touch. Lee and Hui [2018] extended Tung et al. [2015] classification with a survey to interaction methods of smart glasses. Figure 4 depicts a classification of their surveyed approaches.

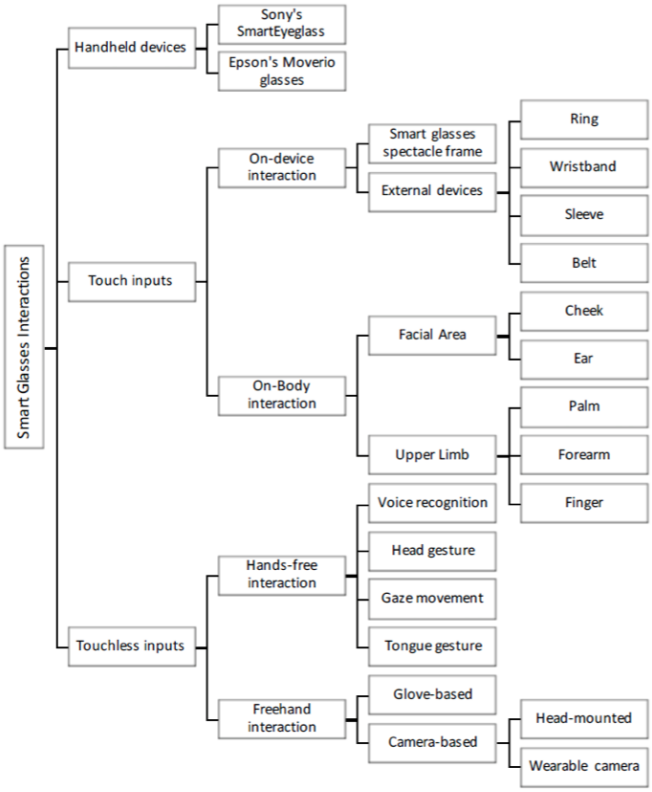

Figure 4. Classification of interaction approaches for smart glasses [Lee and Hui, 2018].

A handheld controller or device is a convenient platform and tangible abstraction layer of interaction in service to the UI shared by all applications it controls. It offers developers a method to map direct actions onto toggles as they see fit, but the amount is limited by controller design after system functions and cardinal directions control. A controller does not need to be a handheld one. Touch inputs can be built as buttons or touch

controls onto a display device body or a peripheral controller worn on the user's body. For example, Vuzix smart glasses have a touch-sensitive spectacle frame, where a tap or a swipe gesture can be performed on [Vuzix, 2020]. More experimental, touch input options rely on sensing devices picking up user's touch on their own bodies [Lu et al., 2020]. The tactile sensation felt on their finger or hand when operating touch inputs reassures the user of their action with immediate feedback. This leads them to be often activated eyes-free, but as the tactile sensation or accompanying sound is easily missed, precise control demands the UI to respond fast with a visual change as the user is expecting it [Apple, 2020a]. Interestingly enough, for feedback, there are suggestive findings challenging the notion that spatial interaction requires a screen, in that user's short-term visual memory can partly substitute screen presented feedback [Gustafson, 2012]. However, if the quick response graphical user interface change is not provided, users will likely assume the action has no effect on the app [Apple, 2020a].

This questions how useful can touch inputs from Figure 4 be in industrial environments. There is a possibility that raising a hand to locate the head-worn touch interface with tactile sense alone repetitively makes prolonged use difficult due to muscle fatigue [Hincapié-Ramos et al., 2014]. If touch inputs are placed on more ergonomic areas, it is still possible that in an industrial environment, a touch sensing area might be inoperable with gloves on or having an excess amount of dust or grease on the user's hands [Aromaa et al., 2016]. Touching a controller with a hand halts the user from interacting directly with real-world objects, undermining the core benefit of using HMDs in a work context. If a handheld controller is suitable for a specific industrial use case, might the user perform the required task with a smartphone instead.

The touchless inputs of Figure 4 containing voice, gaze, and mid-air gesture recognition are also used in VE selection/manipulation (Figure 3) to give direct instructions, or commands, to a computer system to be executed as an action. In multimodal interaction, direct commands can be combined with other selection/manipulation techniques. For example, moving an object by first indicating it by pointing to it with a hand, indicating its selection with a voice command, then indicating a new place for it, confirm it by voice command, and the object in question jumps to the pointed location. Beyond using the available modalities in the system for direct manipulations of the graphical user interface, underlying meanings, or semiotics, can be sought for input in human speech, gaze, and gestures with applications of computational semiotics [Wikipedia, 2021b]. For example, the meaning of the words uttered or probing signs of interest from the user's gaze and facial gestures for the system to guess how to assist the user. Human hand gestures can have meaning imbued to them beyond direct manipulation of objects or pointing, for example, giving another person the "thumbs up" gesture.

Interfaces meet users' expectations better by recognizing the type of user's action [Aigner et al., 2012]. Classification to ergotic, semiotics, and epistemic by Cadoz [1994] was based on their function in the real world. Ergotic gestures manipulate real-world objects in the environment, semiotics gestures carry meaning, and epistemic gestures are used to discover the environment through tactile experience [Luciani, 2007]. Ergotic is otherwise very similar to the idea of direct manipulation in VE, but direct manipulation lacks the required energy exchanged between human and the object [Luciani, 2007]. What is understood to be a symbolic mid-air gesture with the HoloLens platform is a semiotics gesture but so is the direct manipulation input model. The Bloom-gesture, where fingers are held together and opened like a flower bloom, of HoloLens 1 was a symbolic gesture that contains the meaning of activating the menu/task manager [Microsoft, 2019c]. Direct manipulation with hands, the primary input model of HoloLens 2, related gestures are semiotics gestures and only carry meaning observed by the system and cannot change the object's shape.

In human-to-human interaction, semiotics gestures are used to communicate meaningful information to the interactive dialogue and are useful as they expand the conveyance of the information from voice onto a second band through the gesturing modality. Semiotics gestures have different meanings for people; they have learned in their lives, and when they are seen, they are interpreted at the moment. Additionally, computer dictated symbolic gestures are tricky for humans to learn and perform. Bloom-gesture in HoloLens was replaced with a holographic button inside the user's wrist [Microsoft, 2019c]. In the new version, the user shows their wrist and presses the icon with the other hand's finger to open the menu is an analog from a handheld controller with a system button placed. HoloLens 2 recognize up to 25 points of articulation per hand through the wrist and fingers, but for optical hand tracking, hands must not overlap each other when forming hand poses [Microsoft, 2019c]. The wrist button takes this limitation and turns it into a strength in the system.

For gesture recognition, mid-air gestures can be sorted into static and dynamic categories [Hummels and Stappers, 1998]. Static gestures are postures void of temporal movement, and dynamic gestures involve hands and fingers moving in two or three dimensions. Gesture languages, use to communicate commands to the machine, are either created by designers with predefined meanings taught to users or elicited by observing users' behavior [Aigner et al., 2012]. A specific meaning of a gesture is often understood consciously or subconsciously in human-to-human interaction. These are hard to elicit via machine capture and labeling without introducing constraints [Aigner et al., 2012]. Attempting to bridge these approaches by first designing a language and then using it to help elicit gestures belonging to it from a one-way gesture-only human-to-human interaction instruction giving task, Aigner et al. [2012] classified the following eight types of

gesture sets, shown in Figure 5 and individually presented next. These gestures function more as replacing language rather than augmenting language.

Pointing / Deictic: Represented by the index finger, many fingers, or a flat palm, used to isolate an object, to specify a direction, or to learn more information.

Semaphoric-Static: Are carried out with one or both hands without movement directed to the observer. Meaning is derived from social symbols or etiquette understood consciously or subconsciously, for example, thumbs-up for encouragement or a forward-facing palm for stopping action for encouragement or stopping action.

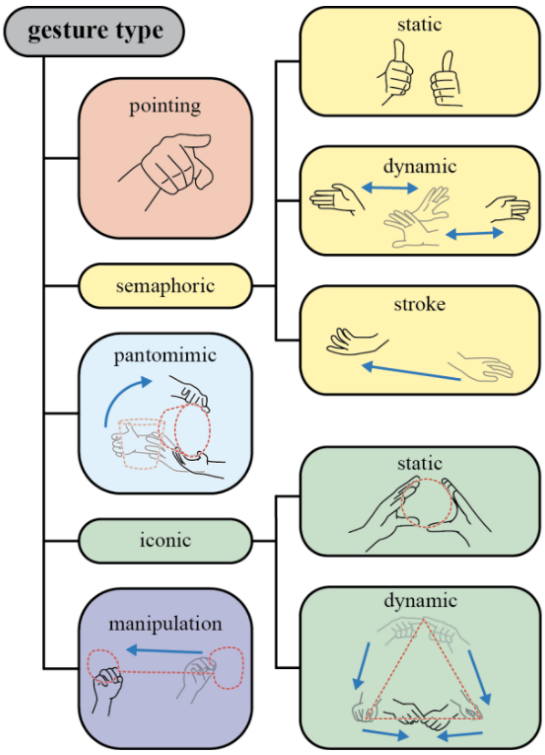

Figure 5. Gesture Classification Scheme [Aigner et al., 2012]

Semaphoric-Dynamic: Temporal aspect through continual movement added to the static version, like conveying rotation or continuous sideways flick to indicate "no". Beat gestures emphasizing communication are omitted as they lack semantic content but may instead be a unique gesture type [Bernard et al., 2015].

Semaphoric-Stroke: Like dynamic version, but constrained to a single dedicated stroke, for example, a single sideways hand flick for moving to next or previous page.

Pantomimic: Acting to convey a narrative line without speech, like a gesticulated idea of the line "then we made a hard left turn".

Iconic-Static: Shape, approximated by fingers, representing an icon or an image, with no social or cultural significance.

Iconic-Dynamic: Added temporal aspect. Movement from both hands to form the outlines of an icon, such as a triangle.

Manipulation: Previous types have an explicit initiation and completion state within a time window, whereas manipulation is performed in a continuous manner while commands are executed. For example, no delay exists between object selection and its updated location once the hand is moved, the object being attached to the hand.

Groenewald et al. [2016] findings from a systematic literature review on mid-air hand gestures show that the main types of gestures created for research prototypes were performed using Semaphoric-Stroke, Semaphoric-Dynamic, and Pointing for a selection

task. For navigation, Semaphoric-Dynamic and Semaphoric-Stroke were used. Finally, manipulation was performed with Manipulation actions. Most of which were performed using uni-manual gestures, and less than 20% being bi-manual, mostly Semaphoric-Dynamic, Manipulation, and Semaphoric-Stroke [Groenewald et al., 2016].

3 Methods

This chapter presents an empirical part of the study. Section 3.1 presents the objectives of the thesis and particular research issues with justification, and section 3.2 presents the course of the study. Grounded theory as a research method introduces section 3.3. Section 3.4 presents the implementation of an interview study, and finally in section 3.5 goes through the analysis of the collected interview material.

3.1 Research problem

This study aims to understand the reasoning behind interaction design for XR solutions in industrial environments and increase understanding of essential aspects for industrial field environment work use of XR solutions. The answers are mainly sought by looking at the designers' perspective, but results are also balanced in proportion by industrial workers' and industrial companies' interests, such as safety, efficiency, feasibility, and technology acceptance.

Research questions are as follows:

1. What kind of interaction techniques and practices suit the industrial environment the best?
2. What aspects make XR solutions easy to adopt to industrial field work?

It is essential to have information about designers' and developers' perceptions of interaction techniques and practices because they have the first-hand experience of XR projects done previously in industrial companies. Designers and developers often make a decisive decision whether to implement or not an interaction technique. The first question seeks more information on the reasoning behind these decisions and if some commonly used techniques do not fit the industrial environment.

The purpose of the second question is to find the factors that are important, useful, and make XR solutions easier to incorporate into everyday field work, in its sometimes challenging conditions. It is sometimes organized with a massive and global workforce and no time to train to use the XR solutions, so solutions should be intuitive and fast to learn.

3.2 Research design

After discussing areas of interest for the research from either VR or AR with the advisor, he mentioned the possibility of him reaching out to Finnish industrial sector contact for potential thesis subjects. After agreeing to the outlook to industrial interaction, there was an opportunity to narrow the scope to either or. After initial familiarization of the rather sciolistic and hype-driven portrayals of AR and VR in the popular technology press and

discovering that they have things in common regarding interaction, everything under XR nomination was chosen as the research topic. In the delineation of the research subject and interviews as a data collection method, it became clear that all participants should be working in or with close familiarity with the industrial sector. Considering the study's subject, the number of potential interview candidates would be pretty limited, and the number of people realistically considering participating would be even narrower. The expected small sample size supported qualitative study over quantitative.

The researcher began the research with complete unfamiliarity beyond casual knowledge of terminology towards the subject matter of AR and VR and the industrial sector. Thus, a minor literature review was conducted pre-interviews to fortify the researcher's knowledge of terminology and discover if any previous theories exist in industrial XR interaction. Following the grounded theory approach, the main literature review was decided to be conducted after the semi-structured interviews in order to tie findings together with preceding research [Hoda et al., 2011]. An approach that fits together with the research method grounded theory. The next chapter 3.3. introduces grounded theory in more detail and the following chapter 3.4 the interview process. The material collected by an interview study was analyzed using the grounded theory method. The analysis of the interview material will be described in chapter 3.5.

3.3 Grounded theory methodology

Associated most often with qualitative studies, grounded theory is a methodological approach concerned with investigating and defining a research area in an attempt to develop new understanding [Glaser and Strauss, 1967; Corbin and Strauss, 1990]. The researcher neither develops nor tests a hypothesis, but a theory arises from the analysis of the material. However, it is good to remember that in a small sampling qualitative study, the observations cannot naturally be generalized to cover the entire interviewee reference group.

The methods of compromise of an iterative process of accumulating data, parsing the data by coding, comparing, and classifying the material, reviewing associated literature, and reporting findings [Hoda et al., 2011]. Data collection methods utilized in grounded theory include but are not limited to interviews, observations, scientific papers, video recordings, and books [Corbin and Strauss, 1990]. In fact, the data collection is ongoing until deemed to be unnecessary by the researcher. Qualitative material analysis is performed at every stage of the research process [Corbin and Strauss, 1990]. Theoretical saturation is met, and the researcher can halt accumulating material when it no longer adds any fresh information or ideas to the created categories created through the coding process [Hoda et al., 2011]. From patterns discerned from the material, the end result of the process is a theory with a basis in data-driven inductive reasoning instead of deductive verification of previous theories [Hoda et al., 2011].

Since the early stages of the project, it was found that research data related to the topic is scarce, and it is why a data-driven method like grounded theory was chosen for this thesis. For a preliminary minor literature review, the researcher sought the terminology and concepts introduced in chapter 2. Beyond this, time was spent looking through and making notes of press coverage related to AR, MR, VR, and XR with connections to the industrial sector. Some of these were in the form of articles and some videos. Especially videos were enlightening for getting a holistic picture of how the technology works functioned as a preparation for the interviews. Along the process, notes were made by roughly grouping material to Microsoft Word documents and storing potential research material to Zotero for further inspection. During this time in November of 2019 and January of 2020, the researcher traveled to XR industry meetups for observation and participant recruitment. In the meetups, pictures were taken, paper and iPad Notes app notes were made, audio recordings were created, and some presentations were video captured with the iPad during events to keep observation notes.

The next addition to the research material was the semi-structured interviews. Due to following the research method, the researcher proceeded to the interview phase with the research area of industrial XR interaction without fully established research questions to narrow the scope of the research [Hoda et al., 2011]. During March 2020, ongoing pandemic protocols in Finland recommended avoiding travel in public transport and for all who can work from home to do so, leading to schedule interview meetings as video conferencing. Semi-structured interviews were conducted with open-ended questions to leave the participants to explain their concerns and how they proceed to address them [Hoda et al., 2011]. During the interview process and analysis process on ideas, perceptions, and relationships in a brain dump manner, tacit knowledge was recorded whenever the occasion arose to Word–documents and audio recordings. The interview process is outlined in the next section, 3.4, and their analysis in 3.5.

Grounded theory favors collecting materials from various sources and supplementing the data collection with observations [Hoda et al., 2011]. While end-users attitudes towards the interaction techniques are not directly studied, the secondhand experiences of interviewed designers involved with user studies and familiarity of the use setting can convey general attitudes and descriptions. It was fruitful having the expert interviewee as an observer possessing insight into previously discovered issues with prototypes or solutions their non-disclosure agreement allows them to talk around off.

A major literature review was conducted after all the interviews were completed. Supplementing the small sample size of the interviews, the researcher immersed fully into attempting to find previous research related closely to the subject and additional findings from the interviews. When theoretical saturation was met, criteria were formulated to identify the core categories of the research material and proceed to present the theory

grounded on the data collected. This was done by comparing the theory with the data collected, identifying cases that did not fit the theory, and trace the causes to complete the incomplete categories to support the writing of the thesis. Additional review of the research on experimental input devices and their associated display outputs and styles of interaction techniques to identify their characteristics and potential suitability for industrial environments was conducted. At the same time, previous research pursuits on a similar or close to similar subject for patterns that the researcher might have deemphasized or overlooked during interview analysis were reviewed. During the thesis process, the used Zotero database collection grew to consist of 1321 manually added titles, which means the researcher saw least seen them on a title and abstract or glance manner, and those are only the ones that were decided to be added. It is noteworthy to point out that grounded theory is difficult to follow perfectly as a novice researcher, and in this thesis, it might be the case that the attempt has fallen short of it.

3.4 Interview process

This chapter will go through in detail the matters related to the implementation of the interview study process. First in section 3.4.1 presents the participant recruitment process, and section 3.4.2 practical implementation of the interview meetings. The interview content is examined briefly in section 3.4.3. The interview script is attached to the end of this document as Appendix 1.

3.4.1 Recruitment

Part of the research was attending four networking events relating to the XR field in Finland between November 2019 and January 2020. These were: VR Aamu at Kampusklubi in Tampere, Match Up 2019 at Helsinki XR Center and X Reality Day at Slush 2019 in Helsinki, and Academic Mindtrek in January of 2020 in Tampere. The trips allowed the researcher to glimpse into the work done in the field and talk to the people working on them. Initial interest among the people in the XR community was high, and the recruiting effort on these events was positive, yielding 17 new contacts and potential recruits. Most industrial companies present at the meetings presented VR and handheld AR implementations related to either tradeshow demos showcasing manufactured offerings or work safety guidance rather than industrial work-related applications. As a result of attending these events, mainly tacit knowledge was gathered of XR and learned what kind of professionals the recruiting efforts should focus on sampling.

The interviewee sampling focused on participants who were either working on or who had recent experience of industrial field work XR projects, rather than those whose knowledge of such projects was too removed to recall any problems. The ideal participant to be sampled had experience from AR and VR implementations, on top of traditional 2D applications instead of only VR applications. Broadly speaking, this included senior-level

designers, developers, and researchers with experience in recent implementations that a nondisclosure agreement did not impede from participating. Participants' ages were not seen as significant.

The benefit of attending networking events was that a couple of contacts from these events supported the recruiting effort by reaching out to their contacts via social media with the interview request, adding to the search effort for potential recruits on LinkedIn. There is a risk of oversampling professionals who are active and engaged with XR community networking and have mainly interest and engaged with non-industrial projects when using social media as a recruitment method. For a research project focusing on industrial environments, interviewing people who are engaged with industrial applications in a work context is a priority. Nevertheless, this should not disregard the use of social media as a recruitment method. In the end, social media recruitment did not yield any interview meetings for this study, and all respondents except one were recruited through direct email interview requests.

As a result of the participant selection process, ten senior and principal-level professionals working in Finland, all with eight or more years of professional experience, were contacted. The pre-selection of the participants was carried out via background research of the designers, websites, and networking that was encountered during the initial research process. Five of the ten responded to the email invitation, and an interview meeting was set up with them. Three of them resulted from face-to-face recruitment from networking events, two from cold emailing authors discovered from previous Finnish research papers involving industrial companies interested in XR. Additionally, one participant meeting was set up by a friend of mine. An hour was agreed upon as the interview meeting length.

All six participants were experts with several years of AR-MR-VR projects, five having worked on a project related to industrial XR, and one interviewee was a domain expert on binocular HMDs with close ties to industrial design. Consequently, they can be recognized as experts related to XR in the industrial domain. The roles in which these people profile themselves were a research director, a software development engineer, and four lead user experience designers. Interviews involved four industrial companies, all with substantial global industrial manufacturing and maintenance businesses, and one consultant and one high-end HMD manufacturer representative. Even though there were only six interviewees, their input was valuable as domain experts had all been serving as an interaction designer or decision-maker for implemented interaction at an XR software project.

3.4.2 Implementation

Before the interviews took place, an interview script with different stages and an interview recording permission query for use as a reminder were compiled. The interviews followed the interview script in Appendix 1, which was especially useful during the first

appointments. At the beginning of each interview, the participant was informed of their rights to the study, the interviewee availability timeframe of one hour was confirmed, interview sections and purpose of the study were explained in a general manner, and the recording permit was collected verbally from the participant after which the recording was turned on, and the recording consent queried again for record purposes. The study's objectives were explained in more detail at the end of the meeting if the need arose.

Interview meetings were held from 17 April to 4 May 2020. Only the researcher and the participant were present at all research meetings. Interviews were conducted in Finnish over the internet on a Zoom or Teams videoconference, only one with audio-only, and recorded using built-in recording tools. The first research meeting served as a pilot, after which it was decided upon to perform a quick, opensource intelligence -style, search engine familiarization with the expert participant and the company they represented before each interview. The need for more in-depth activation emerged from the pilot meeting, but otherwise, the pilot meeting went smoothly, so the material collected from this meeting was included in the total material. The interview meetings were recorded in their entirety after the recording permit was received. A total of 5 hours and 37 minutes of material was accumulated. The material is stored on an external hard drive until this thesis has been reviewed, after which the recordings are destroyed.

3.4.3 Interview content

The semi-structured interview was centered around the topics of two research questions, with three more distinct themes. The first theme is about previous XR projects done and the general learnings from, second on interaction techniques and their perceived fit to industrial environments, and the third deals with the aspects of XR solutions that make them easy to adopt the field work. Each had its related probing questions, which were asked depending on the informant's answers. The interview script, Appendix 1, was in Finnish, but a rough translation was available if interviewees wanted to be interviewed in English, but none did.

The interview atmosphere was aimed at a relaxed open discussion. The nature of the open-ended questions was such that the participants had a great deal of influence on the topic discussed. Essentially, they talked about the project they could or wanted to discuss. This decision was influenced by their non-disclosure agreements to a great deal, and many talked only about projects publicized, for example, in a company blog or involvement with projects tied to academia in some manner. The topics covered were varied: airliner interior seating assembly, user testing a VR version of a cruise ship's bridge consoles on location, a digital twin of industrial plant automation and related activities, XR solutions for a large global maintenance crew, industrial equipment maintenance on customer premises and one on high-end VR and AV in industrial design and training scenarios. It was made clear that the researcher was willing to withdraw any remarks a participant did

not expressively want to disclose, and this was asked from them at the end of the interview.

After a brief introduction to their professional history meant to relax the participant, the first theme of the interview was centered on singling out a previous project for discussion. Everyone who did not dive directly into a topic was asked what XR projects have been done in the company or they have been part of previously. After which, they would be asked what they were proudest of or had the most to say about and why. These were used to guide the discussion to focus on one project's problem worth solving and gain a more holistic understanding of the company's business in relation to the research topic.

Most significant challenges, risks, most important decisions, barriers to success, and enablers were asked if the participant needed stimulation or had trouble in jogging their memory about the project. If a participant were about to reveal a regret or something they would have done differently in a previous project, they were encouraged to do so. These were key points in turning the interview from an individual project to the second theme to interaction techniques and questions related to their implementation. Participant's experiences, opinions, and even assumptions of what kind of input and output techniques and interaction practices suit the industrial environment the best. After which general misconceptions about industrial interaction they have personally found to be otherwise were asked. Additional questions were based on user needs and the role in which any particular input or output modality functions in the XR solution.

For the third theme use, a general description of an industrial worker's typical workday was asked from which the deeper queries diverged into forks related to safety, personal protective equipment (PPE), ergonomics, and user experience. Use case descriptions also carried over to user needs, motivations, feelings, value to them, and general success criteria of a solution. Participants were asked what a functional XR concept and its interaction techniques implemented would look like in the industrial field work context.

All input and output modalities most commonly found in commercial HMDs were brought up to query an impression about during an interview. Observations were made if a participant was perceived to give a speculative answer to accommodate the interviewee. This kind of weighing done by the researcher is risky, but personal hesitancy on pushing an interviewee with additional queries on a topic they appear to know very little or not wanting to discuss, in favor of appearing sociable or well-mannered. Almost always, after every question, in an attempt to make participants elaborate on their decision making, either; 'why' to enlighten their thinking, the logical path to a decision made, or if and how they validated the idea were asked.

3.5 Data Analysis

Data-driven qualitative content analysis was vital to obtaining insights from the semi-structured interviews. The material was exposed to slight comparison during transcribing and translating, after which constant comparison took the form of open coding known in the grounded theory method.

The approach followed the following structure:

1. Listening to interviews and writing them down word for word.
2. Reading interviews and getting acquainted with the data.
3. Translating the interviews and opening up the meanings.
4. Finding and underlining expressions.
5. Listing of expressions.
6. Open coding expressions to reduce them.
7. Finding similarities and differences in the coded expressions.
8. Combining codes and forming subcategories.
9. Combining subcategories and forming abstract concepts.
10. Combining concepts and forming categories
11. Combining categories to main theme
12. Performing selective coding to the main categories

After listening to the material several times, it was not sensible to transcribe the entire material by hand. Subsequently, to reduce the transcription workload, it was proceeded to try out an improvised auto-transcription method of using the speak-aloud tool of Microsoft Word. A recording was played on a computer aloud via speakers with a word-processing laptop nearby listening for spoken Finnish it can pick up from the microphone. However, Finnish recognition is only an experimental feature, and auto-transcription misheard multiple words leading to unrecognizable sentences, and only part of the transcription was usable, the rest requiring extensive editing of the provided transcription by hand. The video was played on a tablet device with slowed speed, and the Bluetooth headphones touch interface was used for pausing and jumping back and forth. This approach made transcription more accurate but much slower as the researcher was forced to move between applications on different devices. However, it can be argued that the earlier formed auto-transcription speeded up the overall transcription speed compared to the manual approach. It was more effective to edit an inaccurate copy than it is to transcribe the same interview from the start all over again. Each interview was listened to repeatedly, and a one-hour interview took nearby to 6–7 hours to complete transcribe, translating into English taking another 6-7 hours.

The iterative review of the material through the transcription process was the first stage of the analysis. Transcription is an essential process towards the profound comprehension of research material, and complete transcriptions were aimed at keeping all the

answers to the interview questions, even those outside the research topic. Interview subjects were pseudo-anonymized with an index number (e.g., I1). The researcher's own questions and comments were transcribed with a different color and any relevant thoughts or commentary that were marked down during the interview and those that came up during the listening with a different color. Brackets were used on a missing reference word in the sentence when it is not self-evident what the comment refers to. The next step was translating the transcriptions into English, not as direct translations, but translating a statement into a common language, and therefore the statements had to be interpreted in their context, even if the quotation itself does not contain any reference to it. Consequently, this repeated exposure to the data developed a more profound comprehension of the material, and the researcher was already engaged in iterative grounded research. Already at this stage, there was so much repetition of the translations in specific ideas varying only in the point of view that it was possible to identify similar expressions of thought for the next phase.

Once translation and reduced the statements were completed, meaningful sentences were searched for in the material by printing out all six of the interview Word files and going through them with different color highlighters. Nevertheless, the first attempted analysis trying to follow the approach of Miles and Huberman [1994] to reduce, cluster, and abstract my collected qualitative data failed. No headway was made with the approach when attempting to cluster the units of meaning as there was no limit to the number of groups that could be created. Following the grounded theory method, it was decided to go back a step and transfer all the translated data to an Excel sheet to be open coded without the researcher's observation notes. Thus, the material was coded in the purest form possible. Only the key pieces of data were coded with short phrases, avoiding personal reinterpretation of motivation or evidence based on preconceived ideas.

Using constant comparison, the codes were compared first within the single interview and then between interviews and observations. Next, to group codes together as subcategories with comparison, repeated this to form concepts and repeated again to finally form categories. This process was very time-consuming and required constant re-thinking of the material. Each citation has the context it is translated from, and its meaning from in reflection to others. Their interpretation from open codes to subcategories is a process that reflects both of these contexts. Open codes are grouped based on their similarities, and subcategories are distinguished from each other based on their differences. Throughout the process, the relationship was explored of these codes, concepts, and categories by writing notes on ideas, perceptions, and questions.

After each of these individual cycles was completed, the researcher returned back to iterate them. Then longitudinal coding was done for the formulation of the core catego-

ries. The core categories are touchless interaction, occupational safety and personal ergonomics, and empowering worker and technology acceptance. Next, chapter 4 presents the results through the core categories. Chapter 5 presents the grounded theory formed by selective coding.

4 Findings

Some interviewees were more expressive and illustrative than others leading to many quotes in this section originating from the same participants, even if another participant had expressed a similar notion.

The results of the interviews can be summarized this way, emphasized in every interview was the importance of occupational safety when introducing beneficial technologies to industrial work. It must be considered that most interaction techniques can affect user's perception negatively and that the ordinary user is not very accepting of new potentially disruptive technologies. The best-fitting interaction techniques to industrial environments are hands-free touchless interfaces, the least fitting being handheld controllers. Visual information was deemed to be the most important, audio output being a secondary feedback channel that should be customizable and switched off if the user wishes so provided. Users saw tactile and haptic feedback for industrial field work to be low-quality feedback on current HMDs, if present at all.

Table 1 shows a collage of the concepts and categories, the outcome of reducing, coding, clustering, and abstracting the interview material produced. The order in which they appear is not of significance. Category descriptions have been condensed to a main theme representative of the interviews only to have a more of a tree-like structure in the table.

Concepts	Categories	Main theme
Speech recognition is the readiest to make an impact. Mid-air gestures can be a safety concern. Touch input is not suited to challenging conditions. Gaze not viable for object interaction. Multimodal input is required.	Touchless interaction	Interaction designer's perceptions on industrial worker's needs and accompanying requirements.
Unencumbered perception is essential. High mobility needs. Visual information is most important. The technician needs to be able to choose the display they want to use.	Occupational safety & Personal ergonomics	
The technician needs easy access to help. Need for reporting and content creation along with maintenance. Wearables are intimidating. ROI on smart glasses poor compared to a smartphone. XR is not widely deployed to industrial environments. High-end HMDs create value in specific roles.	Empowering worker & Technology acceptance	

Table 1. Concepts, categories, and the main theme produced by the analysis of the interviews.

Introducing HMDs widely to an industrial organization's field work crew was seen as an attempt to create, capture, and distribute knowledge and ensure its availability. It was seen as troublesome due to a high unit cost and the prototype state of most HMDs in comparison to smartphones. Smartphones are currently the most available and familiar device to the user base. However, it was acknowledged that immersive technologies might have the opportunity to produce significant value in certain specific roles. These were dealing with rarer, high urgency, high-cost maintenance technician deployments, or possibly less intensive scenarios in safer and cleaner environments like training in a maintenance simulator rig.

Developing interaction techniques is complex. Research and development of them are rarely done in-house and are mostly left to other commercial ventures. Two of the six interviewees expressing this by:

> *"We trust Microsoft on handling interaction development and have no extensive interface development plans on the works." – (I3)*

> *"Our company is not that involved with modalities, and I cannot comment a lot on voice, other modalities, or fusion of modalities on which other companies experiment on, sometimes with the equipment we delivered. Company decision is to leave it to them, mostly to folks in Silicon Valley." – (I4)*

All participants agreed on touch inputs, handheld devices, and controllers to be less suitable to industrial environments for requiring a hand to operate them. A handheld controller was also seen as a reluctant interaction option for MR and VR equipment in a product demo or usability test as users cannot just drop them away from their hand and is tied around the user's wrist. The handheld controller was described to be covered with strange functioning buttons and an especially programmed artificial high-level abstraction tool for high efficiency on a specific task. A shared notion of participants of handheld controllers was that only experienced users are familiar and competent with controller usage, whereas they remain unfamiliar to ordinary users. Any handheld device, smartphone, tablet, controller, or hand and finger-worn controllers were seen to interfere with a technician's work. Additionally, most of them were mentioned as not suited to be operated while wearing dirty protective gloves.

One interviewee, developing VR and HoloLens supported application, was delighted by Universal Windows Applications supporting 2D and 3D interfaces and the computer to be able to decide upon the version of the application presented based on the output and input peripherals connected. Some situations, social or environmental, were felt to favor certain input and output modalities. Three participants mentioned that multiple means of accomplishing the same on a platform should exist. Two participants mentioned that a voice interface dominated device with Realwear HMT-1 is not an option for them. All

interviewees expressed that multimodality is required due to the need to accommodate different use scenarios and personal preferences.

> *"People have personal preferences, someone doesn't like something, someone speaks poorly or in an accent to speech recognition does not work, someone really does not hear well and therefore does not want to use something, and someone else's reason is something else." – (I5)*

Touchless interaction techniques, speech, mid-air gestures, and gaze were seen to be useful separately but also to be closely intertwined. These techniques had been pre-assessed as owning qualities primarily complementing each other when considered a whole.

> *"Functioning system should rely on multimodal fusion combination of at least mid-air gesturing, gaze, and speech." – (I6)*

4.1 Touchless inputs

To open a discussion on interaction techniques, the interview questions involved with input methods were reliant primarily on discussing the idea of object interaction (see section 2.3), but this was only used to entice thoughts on the matter if the interviewee had a hard time expressing their ideas otherwise. In the latter part of the interview discussing aspects of XR solutions for field work, the interviewees were asked about their wishes regarding the style of interaction on XR solutions. Asking a participant to sort input modalities from the best working to the least working for field work was replied to be difficult because it depends so much on how well individual techniques work. This might open up interaction technique possibilities beyond the ones mentioned in this section. Based on the responses, a fairly coherent picture emerged of what interaction techniques can be used in sometimes challenging industrial environments.

4.1.1 Speech

Participants were most avid on expressing their views of spoken input. Speech recognition was regarded as the most ready to make an impact on industrial work and the most mature technology to enable hands-free working for industrial maintenance. Additionally, it was commented on being the easiest to implement and the simplest to introduce in the work-life context without the resistance other non-touch input technologies might face.

> *"Speech recognition is beginning to be usable. Despite environmental circumstances, the best performing in user studies has been speech recognition." – (I5)*

> *"Speech control, even though a more viable option, has its own challenges. A noisy environment is one of them, and another is how able the user is to give the speech commands. Noise filtering options do exist, so maybe speech control is closer to reality for future use cases." – (I6)*

"It supports hands-free work even if the technician is in a noisy environment if a pair of Bluetooth hearing protectors is used […], especially well if the pair is mic'd up for noise cancellation (muting circuit) and speech command purposes." – (I3)

Speech input is affected by its use environment, environmental noise, and social setting were highlighted as the main challenges. Noise filtering mics and noise-canceling hearing protectors were mentioned to be influential in functional speech queries. Two interviewees mentioned the Realwear HMT-1 as an example of a known HMD where voice control has functioned relatively well in noisy environments. Social acceptability in public or semipublic places was seen as a considerable challenge by all participants and its relationship with an individual's technology acceptance of instructing a machine via speech and the feelings related to being potentially seen doing so in a public social setting.

Compared to hands-free input modalities, the relatively good user acceptance of speech queries was attributed to the assumption that many users are increasingly familiar with voice user interfaces (VUI) from their private lives. Most participants identified speech recognition and VUI as the most straightforward hands-free technology for users to engage with as beginners. According to the interviews well-functioning, VUI guides the user to activate it correctly, indicates that it is listening to the user. Then keeps them on top of what is happening with visual elements of the user interface, actively indicating the state of the system and displaying clear commands so users know what they must attempt to say. VUI should always give feedback on whether it understood the user and visualize the uttered phrase to communicate what the system received or how it interpreted the phrase. This is done to resolve potential misunderstandings and expediting recovery. One interviewee summarized the successful principle of recognize over recall with speech control as:

"Good voice control is such that the user does not have to remember anything." – (I5)

Another intriguing aspect of speech input brought up is the possibility of using free form speech without the need for a special user interface like VUI described above. The interviewee explained that freely spoken input is picked up by speech recognition, and semantic artificial intelligence interface picks up essential key phrases. The interviewee expressed that a maintenance technician can benefit from by freely vocalizing maintenance actions to have them automatically be logged:

"This way, the maintenance technician does not need to know what terminology they would need to use but instead can talk freely. It turned out to be a nice method to fill forms and finds files, depending on the situation." – (I5)

However, two participants mentioned a single modality hands-free interaction technique HMD like RealWear HMT-1 does not support their needs.

> *"We (company) cannot have a device that is merely speech controlled." – (I5)*

Participants recognized that a large user base and varied reasons for the inability to make speech queries that speech recognition can understand was the most significant barrier to success. In their opinion, speech recognition still has wide-ranging issues related to different accents, pronunciations, and languages.

4.1.2 Mid-air gestures

This chapter focuses mainly on optically-recognized mid-air gestures, based mainly on the participant experiences with Microsoft HoloLens. Participants were generally encouraged by how natural to perform and how well mid-air gestures are recognized if there are no glitches and hands kept inside the certain zone the sensors can capture. Interviewees agreed that direct manipulation gestures, closer to real-world actions, are more natural to users than indirect symbolic gestures. Symbolic mid-air gestures can be helpful, but it can be difficult for new users to utilize mid-air gestures in the first place.

> *"Direct manipulation gestures are clearly better for our users. New users are so wide-eyed that pressing a button in AR/VR, while natural, is still enough of a challenge for them. […] symbolic ones are too challenging for our users as they've had problems with the direct ones." – (I3)*

> *"Direct manipulation gestures perform better as our user tests have shown that the number of symbolic gestures must be really limited for users to remember them. Even if instructions to perform gestures were inside of the maintenance manual, symbolics are pretty hard for users." – (I5)*

> *"Symbolic gestures are good if they are simple and easy enough to learn, but they should have foolproof reliability. They should work as reliably as touchscreen buttons and have no exceptions. Also, an in-air button should work all the time, not every third time, depending on the specific angle of the hand or finger."– (I6)*

Symbolic gestures need to be simple in form, functioning with all hand profiles, easy enough to learn, and very limited in number for users to remember. All interviewees were suspicious of users remembering and performing a large set of gestures, agreeing that direct manipulation mid-air gestures generally perform better with novice users to whom pressing a floating button with their hand is already a challenging concept. Participants made it clear that inputs utilizing human movement should mimic the natural world or rely on realism-based mental models.

> *"…if there is a light switch, it should be activated with a finger, like the real one, but if I turn my wrist visible and get a menu where I can turn on/off ALL the lights in the apartment. It is then a new element bringing something on top of ordinary reality and justifying its existence and a good solution. Real object interaction happens with real objects, and if there is something more that exceeds what is actually possible for a human to accomplish in reality, then it is only then when these (symbolic) gestures and other modalities (gaze) can be used."– (I4)*

Mid-air gestures should not exceed human capabilities unless by allowing them specific special abilities with new nonrealistic interaction techniques. Examples of well-executed non-direct manipulation techniques mentioned were HoloLens 2's start menu gesture of turning a wrist over and touching the Windows icon and handray.

> *"… (handray is a) so-called laser pointer, so no extra controller is needed. It comes from the wrist and elbow as an extension, and it becomes like a beam. This is used for operating a faraway object [...] it is sometimes a little awkward to use as it calculates clearly wrong in occasion, but if it gets calibrated correctly, then it works really handily."– (I3)*

All participants found robust functioning of the gesture recognition was found to be a matter of great importance. A couple of participants emphasized that mid-air gestures should be as reliable as a touchscreen to be useful or not be implemented at all. Meaning mid-air gesturing should work independently of a finger or hand size or whether the user was a novice or an experienced user. One interviewee described the impact of close to realism mid-air gesture failing on a user due to poor gesture recognition:

> *"Immediately when the system gives out on the user, unable to recognize the gesture, frustration and uncertainty immediately hit, and the user asks themselves "what can I do" "what is the gesture". – (I6)*

Failure to recognize the user's repeated input attempts runs the risk of the user rejecting the use of the modality out of frustration and deteriorating user perception and focus for a moment. One interviewee regarded the repeated input attempts to be a potential safety risk. Participants speculated that large and arching optically tracked mid-air gestures requiring space can be unsafe in industrial field work if running machinery is nearby while the user focuses on the performance of gestures and forgets their environment. It was common knowledge that low or close-to-body mid-air gestures do not function with HoloLens, as hands need to be prominently extended to be visible for the optical sensor array on the user's brow and kept within a zone. One interviewee mentioned that indoors in an industrial plant can separately configure yellow-black ribbon fence graphic to be

placed in front of machinery the user is not allowed to go near and when crossed HoloLens 2 blinks on and off the display. Another mentioned that gestures could be regarded as safe to perform, providing the technician has a free hand available to use, serviced equipment is powered off, and no usage of heavy machinery would be allowed nearby.

However, the participant also mentioned that technicians might operate in such tight spaces that they might not be able to perform large mid-air gestures, or the lack of light in the environment might negatively influence gesture recognition. Lighting conditions can negatively impact the usability of optically captured mid-air gestures as optical gesture recognition needs optimal lighting to work well. An interviewee noted HoloLens 2 has poor outdoor performance and that both lighting and calibration can be better controlled indoors. One requirement two interviewees expressed was that for field work, mid-air gestures should work with work gloves. Most maintenance tasks require gloves to be worn. One participant knew that at least thin white work gloves of being compatible with HoloLens mid-air gestures. Optically captured mid-air gestures might not be practical for maintenance servicing. Instead, suit training scenarios in a more office-like environment would be better. A combination of visual and audio feedback on mid-air gestures is important as haptics are missing.

> *"Many input techniques are designed for consumer markets. Industrial use is not on top of mind for the developers." – (I6)*

Two interviewees suggested for non-touch close range mid-air gesturing approach to recognize hand hovering movements performed on top of the device, so swiping without need to touch a surface when hands are dirty, greasy, or a touch-sensitive area does not function with gloves on. The interviewee noted that Vuzix smart glasses could recognize close-range mid-air swipes over the camera, and there are similar methods available for handheld devices if placed on a stand or holder. Additionally, one interviewee had experimented with various displays paired with Kinemic Band, a Bluetooth wrist mouse, that enabled pointing and a range of easy to recall and perform symbolic gestures. Users liked the wrist mouse, but the connectivity was not reliable enough in the version they had tested. Small symbolic gestures tracked inside-out with a hand-worn device could deliver safer gesture recognition in the field as mid-air gestures are performed closer to the body. An additional benefit is that the user can blame the device for failing to recognize the symbolic controller gesture instead of themselves not knowing how to perform the needed gestures. An interviewee mentioned a highflying vision of a future where similar band tracking hand activity could match markers of stages in maintenance procedures digital instruction manual to recognize whether a technician might have missed a step.

4.1.3 Gaze

While the technology readiness of gaze technologies is improving, all interviewees made the point that they have not especially made use of gaze or planned on it in any way but are still waiting for more integrated out-of-the-box solutions. While none had yet designed user input relying on gaze, three participants expressed hope that gaze could be soon used to enhance experiences as a side-channel input. Gaze can enhance the experience by providing added value for the users as a cue of user interest to bring up added information.

> *"For example, if the user is inside a building and gazing on a valve and information is conveyed of what that type of valve is it..." – (I4)*

A human being does not perceive the whole scene in front of them all with one glance, but constant active gaze movements are utilized to move the point of focus for the eyes and the accurate area. One interviewee mentioned that in some high-end closed-view HMDs, gaze data is used to optimize the virtual experience by guiding the computing power on the thing user is focusing on at the moment for a sharper end result. After configuring gaze works reliably within limits and is easy for novices to perform, gaze control rarely can be useful in maintenance scenarios.

> *"Gaze control is not always applicable, it is challenging in the sense that technician must keep an eye on what is going on with the equipment, like flickers LED's that communicate changes, so maintenance technician knows where the process is going, and they cannot relinquish their gaze from that LED panel." – (I5)*

All participants regarded gaze not to be a viable primary interaction technique but saw it more as a fallback method when other input methods are unavailable. Designers can not directly assume that gaze means intention. Mental models from the real world limit the usefulness of gaze and dwell as a selection method as nothing in real life is chosen, nor manipulated, with gaze alone. Gaze can be the last option for object interaction in situations where there is no other method of accomplishing the task.

> *"A fighter pilot, for example, is subjected to terrible G forces, which means that they don't turn their head a lot. Instead, they use gaze." – (I4)*

4.2 Occupational safety and ergonomics

When asked about the most important matter aspect for adopting XR technologies to field work, all respondents replied that above anything else, it is the safety of the user. All technology adoption has to deal with a safe introduction to the occupational environment. An unencumbered perception was identified as being the single most crucial factor contributing to occupational safety based on the interviews. All participants touched on the

topic that, while especially AR HMDs are disruptive to user perception as they are meant to deliver information straight into the user's line of sight, all work practices, including technologies, need to be evaluated for their potential impact on safety. One of these points of an inspection being the impact on the worker's perception. Controlled industrial environments are specially built to reduce user perception harming traits, such as standardized assembly or factory environment and work-related audio alerts. It was mentioned by an interviewee that, for HMDs supporting technicians, device audio alerts could be sacrificed as they are easily lost in the cacophony of alarms in an industrial plant, or they may disrupt technicians' communication. In the field work conditions, however, participants noted that distracting AR HMD visuals might be a safety risk due to deteriorated awareness in an unfamiliar environment, but that all output and input technologies have the potential to be a distraction.

> *"I have wondered that when a user is wearing smart glasses and looking at the environment through them, is the user so focused on the incoming information that the user is looking at the environment but does not "actually" see it? So, how intuitive is it? When information comes in, and you forget your environment and takes a misstep [...] suddenly everything is a lot less safe. With all these technologies, the question is how safely can they be used there in the environment." – (I6)*

Technicians have high mobility needs, and HoloLens 1 and 2 were deemed not to be practical for maintenance technicians due to their physical dimensions and weight. Technicians have a hard time carrying large equipment to work on cramped spaces with some tricky access entryway like hatch located on top of a ladder. Instead, they were HoloLens was seen to be useful for use cases demanding less mobility. By an expert working more closely with industrial plant technicians, the compatibility of HoloLens 2 with Trimble XR10 helmet with the possibility of lifting the visor up was seen as an adequate solution for the environment. It was clear from the responses that observing field work helps to recognize user mobility needs.

> *"It (HoloLens) is too clumsy and heavy of a device for real-life field operations, and it forces user field of vision. Better to be used in training sessions happening at the office. It is a nice device and would perform well under the circumstances." – (I5)*

One critical insight related to safety that was shared between four interviewees familiar with the nature of industrial field work was the importance of keeping the maintenance technician's hands available for performing work tasks.

> *"Occupational safety is key, so successful device implementation keeps technician's hands free." – (I6)*

Without this ability, if not otherwise prepared, a technician would have to choose during servicing which hand to detach from the task at hand for them to reach for the smartphone in their pocket and potentially fumble in the process. After which, a smartphone is always needed to be kept manually visible and touched, potentially by removing dust, oil, and grease-covered safety gloves. Another example mentioned was that a technician might also get into a hazardous situation if the device is accidentally dropped, and they need to reach for it by placing their arm into a dangerous space to get a hold of it.

All the interviewees emphasized that all XR solutions, their input and output methods, meant to be used in industrial field work conditions, must have convenient compatibility with personal protective equipment (PPE) to be usable. Making convenient compatibility more of a challenge is that PPE limits maintenance technician mobility, so technicians have the habit of only equipping themselves with the required PPE when the situation demands it. Participants made clear that field workers only use gloves, safety glasses, or hearing protection only when engaging with a task requiring them. Head protection is always required when entering industrial environments, but maintenance technicians often switch between a bump cap, a lighter version head protection, and a helmet with a chinstrap when it is required.

> *"Realwear HMT-1 style device serves better on the field and others like it or lighter integrated or ones that meet safety goggle standards. Essential considerations are the lightness of the device and that it does not limit user's field of vision at the field […] and is compatible with PPE and prescription eyeglasses." – (I5)*

Many participants reported helmet installation and bump cap installation of smart glasses to be a slow and challenging task for a maintenance technician. Additionally, they would have to detach and attach the device and repeat the adjustments in the middle of the maintenance visit if they switched head protection. One interviewee mentioned that adjusting the smart glasses before the device was usable to achieve a readable display could take over five minutes. After attaching smart glasses onto their helmet, they need to adjust the device's angle on multiple dimensions so that text became clear enough to read from the see-through monocular display. This raised wishes for cheaper head protection integrated safety glasses or AR HMDs without restrictions to the user's FOV or the need for adjusting the optical combiner. In addition, most wished for the ideal field work HMD to be lightweight enough for comfortable all-day usage with all-day battery life and safe cable management. Current binocular HMDs were described as; not sitting comfortably on the face for being front weighted, this causing neck sore and shoulder pain when looking down and to the sides. Participants mentioned that the current remedy

for extending battery life by attaching a bigger battery to the back of the device for balance to be an ergonomic issue or a chord connected to a secondary battery carried with the user to be a potential safety issue. Solutions mentioned to be a hassle were recharging the device during lunch break and switching devices for recharging.

Interviewees agreed that an always visible or easily visible display is the most usable solution. This does not necessarily refer to AR HMDs. The position of the virtual screen created by AR HMD affects readability. Interviewees felt optical stereoscopic HMDs could be a better form factor due to the larger FOV they offer than monocular options. Nevertheless, due to the device itself limiting the FOV by its frame and suboptimal brightness of the display in a well-lit environment leading to poor readability, monocular pass-through displays were held to be better for maintenance. Some considered small monocular displays on smart glasses not to deliver a good reading experience as it places information on a patch on the corner of the vision, but that it still is better than having no display at all or having to hold the display by hand.

> *"It is very important to get the FOV wide enough. Humans have wide a field of vision more than 180 °. Edges of the view are important because movement is perceived there, e.g., an imminent lion attack and react to it. Therefore, the narrower the field of vision in a headset limits away certain use cases, as such a tunnel screen vision limits its use of applications. The bigger the FOV, the better, but peripheral vision does not need to be nowhere near as accurate as the center, on the sides framerate is more important, that way user can detect motion." – (I4)*

> *"Brightness, ambient lighting, the twilight or lighting somehow affect how visible the information and readable the text is. It is also about which glasses you use; with some glasses, it is easy to read text, and with others, it is quite difficult. – (I6)*

Handheld display solution was described as awkward and unergonomic compared to wearable display as a technician's head direction moves between a task and the display when feeding parameters checked from the display. However, it was speculated by an interviewee that the above-mentioned adjustment or readability problems with smart glasses might be enough to lead the user to reject them in favor of a smartphone propped up on a holder. The only thing that matters to the participants is that the display is safe to use, readable, and operationally reliable. Interviewees mentioned that there would be multiple styles of displays in use for XR applications in any given company at one point in time in the future. An interviewee stressed that applications developed need to support different display styles, so the users may choose between various HMD and stationary screen options. Otherwise, technicians should be able to choose any display they want to

use as its style affects its readability, and it may be a matter of personal preference. need is significant enough to warrant an approach of any means necessary.

4.3 Empowering worker and technology acceptance

An unexpected but significant category that emerged from the data analysis was the number of expressions linking to an individual's technology acceptance. Interviewees had a lot to say about the topics behind technology acceptance, as well as the value and motivational factors behind the perceived reasons of why and how XR solutions should be implemented to the field work.

4.3.1 Empowering worker

While maintenance technician is servicing equipment, a problem might arise to which the required knowledge to solve it might not be easily accessible. The digitalization of older model manuals, which only exist in paper form and might have been lost, is slow. If they have been digitalized technician's smartphone is stored away in a pocket not easily accessible.

The topmost need of a maintenance technician mid-maintenance is easy to access to help with any device available. Help, in this case, means quality information in an accessible form. XR technology can help technicians by delivering documentation, instructional manuals, and analytics of what needs to be fixed and what to do for predictive maintenance on the location from a connection to cloud services. After these, the information needs of a technician have been satisfied, can users view be extended with semantic sensor information from an AR HMD infused with data from the serviced equipment and other IoT data sources on-site. The goal is to raise situational awareness to a higher level on what is worth doing there on the location and how to do it. Digitalization should empower technician to feel like a professional that can accomplish a lot on their own. need is significant enough to warrant an approach of any means necessary.

> *"Maintenance technician feels like they are a hero because they can solve those problems."- (I5)*

> *"Maintenance technicians focus area is to maintain industrial equipment. They are not technology freaks who enjoy using wearable technology. It is their work that they enjoy performing. Their sense of accomplishment comes from getting their task done and taken care of and move on forward to the next one."- (I6)*

It is already a technological leap just to be able to deliver information to smart glasses form factor. A handheld device or paper manual and a headlamp are currently available

options to an ordinary maintenance technician. Digitalization brings efficiency to repetitive tasks like delegating less critical work of informing the next parameters to be fed to equipment under servicing to the smart glasses.

> *"User would interact with the measurement tool by inserting physical cursor, like a pen, to obstruct the beam of the laser to test which slot is specifically the right one. Smart glasses would display visual feedback of this action by displaying the possible offset for a correction or lighting up green if the correct one was pointed at. The worker would interact hands-free with the glasses via voice command to acknowledge that the row had been installed and readiness to proceed to the next task."- (I1)*

Interviewees highlighted that for a company with a large maintenance crew and a lot of equipment to maintain, the benefits of one maintenance technician performing their duties faster and spending less time on one job scales exponentially when others are helped to do the same.

The know-how of technicians needs to be on a high level, beyond just getting the job done. Specialized knowledge of a large maintenance crew can be put to use by collecting tacit knowledge from more experienced technicians. Described by interviewees, XR solutions have the possibility to raise less experienced technician's competency level faster than with on-the-job trial and error. Already smartphones enable technicians to produce learning material on the job. According to a participant, this material is not only more economical than commercially produced, but higher quality as well, and the company only moderates the resources. Care must be taken on what work and when technician should be guided and given instructions on as a trained maintenance crew member considers step-by-step instructions to be foolish:

> *"Videos of AR maintenance contain an awful amount of great looking 3D animations, most of which show a tool like a screwdriver with guidance to *now open this screw*. Seeing this, a trained maintenance crew member can only comment that: 'We know how to open a screw. We do not need help with that.'.". – (I5)*

> *"Doing all maintenance work with AR step by step guidance is not cost-effective if 90% of tasks are so simple that with a training of three years or less one can do them." – (I6)*

Interviewees aptly described that AR HMDs lack problems that "pull" the technology into being beneficial to their use, instead of the "push" scenarios where AR HMDs are deployed as a technology demo. According to interviewees, when dealing with unfamiliar components, maintenance technicians value just-in-time learning much more over step-by-step instructions. Rather than step-by-step instruction, participants highlighted easy to

browse digital instructional manuals with complementary visual supports, like standardized symbols, line drawings with highlights and animations, to be the best way to learn ad-hoc on the job. Various devices and scaled screen ratios should support these digital manuals so that content flows to multiple views in a usable manner. Crowdsourcing checking for mistakes to technicians and tracking the style of visual aids most consumed could help the moderation and guide content creation. A participant noted that added support to inspect 3D models or 3D animations in AR, reserved for ideas awkward to communicate otherwise, could slowly introduce technicians to AR without it being mandatory. Another participant mentioned creating a speech-to-text maintenance assistant chatbot, responding text-to-speech from AI knowledgebase. They believe even a more senior technician could accept asking how to do a specific thing without being imposed to go step-by-step through instructions. Only fewer, costly for all parties involved, high complex tasks require fast and low threshold support from manual and beyond. Remote support is only used for top-of-the-hill problems, but pandemic travel restrictions have increased the demand for remote support over video calls with AR features as flying in a global maintenance specialist is not an option.

Another issue XR is sought to ease is reporting during maintenance. According to a participant, the mandatory reporting for maintenance technicians has increased almost five-fold over the past decade. Already optional method of reporting by smartphone has shortened the time spent reporting, and the quality of maintenance journals has improved as technicians prefer reporting during maintenance or right after the session from the field. They believe time spent reading maintenance journals before maintenance visits and writing a new entry should be cut as time spent reporting instead of maintenance affects job satisfaction negatively.

> *"If 2-4 hours of an 8-hour workday is spent on reporting, it is a clear indicator that something has failed." –(I6)*

When work is diverted from their core task to reporting, employee experience can become encumbered. In the interviewee's opinion, reporting should be created automatically or made so effortless and integrated into the workflow that a technician does not need to think about reporting consciously. They envision integrating reporting into the maintenance workflow with automatic image capture and categorizing or speech recognition used to log maintenance actions being performed.

> *"[…] The method does not matter. The most important thing is that everything happens then and there […], and reporting would not require any extra work afterward. The maintenance technician can continue to the next assignment or at the end of the day go home." –(I6)*

4.3.2 Technology acceptance

A user having a piece of technology visibly worn on the body can confuse observers to falsely perceive someone as using technology when in reality, they are not. Participants saw a critical point regarding user's technology acceptance to be factors external from the medium. These factors can act as mental barriers blocking novice from experimenting with the technology. Participants noted that HMDs are intimidating to most new users and have technicians talked about looking silly or amateurish when using wearables.

> *"XR is a bit of a weird thing for everyone in such a professional work context […] novice users, for example, have the specific desire not to move while wearing an HMD. They think that because they don't know what is waiting for them one meter away, stepping forward expecting to bump into someone looks awkward and perhaps risking them socially by making them look silly. – I4*

> *"Some of the technicians even think that they don't want to be seen using smartphones on the client's location as the client may think ill of them. Stuff like these come up where the technology itself does drive towards these behaviors, but it is just the way how people start to think about what others might think of them." – (I6)*

One interviewee described that counteraction against wearable technology emerges from the feeling of requiring to put on wearable technology to accomplish things instead of the capability just existing on an everyday object they use. Participants saw convenience being a significant contributor to the user experience of an interface solution. Benefits help users accept undergoing unpleasant things. High variance in the technological competency of individuals could be seen as a signal that mass adoption is delayed or is not coming at all. Younger workers were seen to be more open towards immersive technologies, enticement of new technology, making early adopters more willing to experiment with experimental interaction techniques. In contrast, the rising age of users was seen to correlate with a low level of technology acceptance towards wearables. Humane usage of technology allows only willing participation. For deploying HMDs in mass to the industrial field work, this implies consumer-side solutions need to become more commonplace for the mass audience of ordinary users before it is possible.

People have different attitudes towards technology. Facilitating multimodal input channels and various display options is valuable for providing alternative methods of accomplishing a task and a respectful approach towards diverse personal preferences of users and accommodating situational needs. Users might not want to use some input technologies, they might perceive them as too complex, possess an impediment inhibiting their use, or the proximity of other people can make the user embarrassed.

"Technicians are reserved. It must be kept in mind that we want to help the maintenance technician. Our center of focus is on the user. Essentially, if we produce (with the technology) a feeling of uncertainty to the user, it scares them away. The value created by the technology in question is diminished. The level of technology acceptance is individual – we can support it, but we can't force it to be faster." – (I5)

The implication here is that the mass of ordinary users is somewhat reserved towards technology and the interaction designer's job is to ease them and help them focus on their job at a high level. There are no good solutions yet for external mental barriers, but increasing user's familiarity with the technology used gives them the confidence to use it. Participants mentioned that utilizing familiar UI patterns from 2D environments helps users cope with binocular HMDs and support technology acceptance of immersive technologies. Participant elaborated that; users are generally unaware of the potential of XR technologies. Even design professionals have a high threshold of switching from familiar 2D tools to working natively in 3D as learning new tools and interactions is felt to be a bit too much to take on. Augmented Virtuality helps timid users to overcome fears related to closed-view HMDs. Additionally, in AV and AR settings, mid-air gestures can be embarrassing to perform. The social consideration threshold is lowered when a user, while wearing an AR HMD, is training with a physical object, and other people with devices are present to confirm the status of what is real. need is significant enough to warrant an approach of any means necessary.

4.3.3 Use of smart glasses

Many participants expressed worries about several manufacturers entering and exiting the Interviews revealed through personal experience or perceived state of the technology a lot of opinions, attitudes, and even company policies regarding the use of wearables and what kind of aspects hinder their adoption to the field work. Interviewees described the current generation of glasses-style wearable devices to be highly intriguing but not yet acceptable quality to be used by maintenance technicians on the field. The reasons behind this notion varied, but interviewees made negative remarks about smart glasses battery life, weight, durability, price, the field of view, connectivity, voice recognition, camera quality, and the time it takes to adjust a readable viewing angle. Updates to technology since last using it were slight improvements, like durability, battery endurance, and HD imaging, and as such to be minuscule at best with no cost-benefit improvement insight. Several participants wished them to be as light and as easy to put on a regular pair of glasses but at the very least comfortable enough to wear for a long 8-hour work shift.

Smart glasses are not yet everyday glasses with enhanced capabilities. Interviewees interested in the maintenance technician's point of view compared together display solu-

tions available for the technician to use. All devised that the issue of most HMDs, surprisingly of smart glasses, is that they are not a self-contained, or standalone, device. Non-standalone displays require a sometimes tricky and time-consuming setup. An extreme example is VR HMD that needs to travel with a powerful PC, cables, and beacons. Smart-glasses form factor device does not have the capability of mobile network connectivity. A participant explained that a technician going to a maintenance visit needs to load the required materials onto the device before the visit if the Wi-Fi network is not available at the worksite or create their own wireless hotspot for the needed connection. An interviewee noted that smart glasses likely require a smartphone with mobile connectivity to share a network and possibly its Bluetooth connection for sharing information between other peripheral devices. Three participants mentioned a user having to configure a Bluetooth connection to a second device using only a tiny touchpad and buttons on the device's frame, as they were the only means to manage the connection. Low-powered smart glasses had issues with wireless connections that can be seen to deteriorate the user experience significantly.

> *"The two technical challenges of smart glasses related to the connectivity to the Bluetooth laser and voice recognition, both of which we tried to overcome with the aid of the manufacturer. […] Smartphone connectivity was much more reliable with the measurement laser." – (I1)*

One participant stated that the 2016 model smartglasses they had developed on top off were still "a bit of a toy". Others referred to the same generation devices as being in the "prototype" stage. Hands-free operation is the goal of smart glasses, but often users must rely on buttons to operate them as issues with the rudimentary speech recognition, and the poor low light performance camera could not satisfy the interaction needs. This meant that to deliver a functional experience to a user, the application development team had to think of an alternate method of completing the task as the functionality of the smart glasses could not be relied upon entirely. Meaningful is that the smart glasses interviewees had worked on ran Android as their operating system. Thus, an application developed for the smart glasses could be made to support smartphone use as well. need is significant enough to warrant an approach of any means necessary.

> *"Even if we would put glasses form factor into the users' hands, still for a long time the smartphone would be the actual device terminal of the masses. It is not insight that they could replace the smartphones with glasses - Instead, glasses could be used in some special situations." – (I5)*

One interviewee estimated that more than 30000 technicians on their staff have a smartphone at their disposal. Smartphones are still the most familiar and convenient interface for technicians. Touch interface works fine for a pre-maintenance checkup and

post-maintenance reporting. So, in non-hurried situations that support taking a break to look for information, a reliable smartphone functions well and serves the technician fine. It is only midway through the maintenance that they are met with hindrances of having hands with greasy gloves and eyes on the job when a need for information emerges, and they have an awkward time reaching for a pocket often containing their personal handheld device. For mass deployment to the field work, however, smart glasses were seen to add only a little or no value at all over the smartphone. need is significant enough to warrant an approach of any means necessary.

> *"Consideration must be given to that the user wants to avoid using technology in this kind of environment. Even if the user knows the device (smart glasses) functions well in the conditions, it's human instinct to be gentle with a device especially, if they know it's priced in the range of 1000-2000€. Ergo, the device has to be cheap or rugged enough the technicians don't have to care about its condition too much if it gets dirty or greased." – (I6)*

> *"For glasses form factor to be worthwhile, they should be robust and cheap enough to be economical. Users can get pretty far for these benefits by placing the smartphone on to a suitable place – The interaction would happen with it." – (I5)*

Interviewees described work environments as usually containing dust, oil, grease, and occasionally hot or humid conditions. Technology is delicate, and users might try to shield the device from the elements, so a practical mass use industrial XR device should be rugged enough to make the user feel confident and cheap enough to be economical. One interviewee suggested a magnetic smartphone holder, previously having tested a sports band holder on the arm during maintenance. Before smartphone is recognized as an unfit tool for a user's job, smart glasses should not be expected to be embraced by users.

> *"There are two different scenarios, "specific" and the other "everyday" which differ a lot from each other. Specific scenarios are special cases where a few users for whom the glasses are really important, work well and use them so much that they become familiar with the usage. This is quite different from the "everyday" scenario, where we are talking about the mass of thousands of maintenance workers who we would like to get to use this technology." – (I5)*

Specific and everyday use scenarios differ a lot. The interviewee explained that, even if two companies are both established market leaders, the priorities of a company performing less critical maintenance service on a larger scale differ a lot from the needs of an industrial plant provider, with high-priority customers paying hundreds of millions of euro.

> *"Stakes are lower for an elevator being offline two days. It is only going to annoy residents of one building; costly fast response is not needed often. Industrial plant risk of going down costs millions; making fast response of workers extremely important." - I5*

The cost of AR HMDs is high. Even the smallest company has several thousand maintenance technicians working globally. Thus, an interviewee predicted that corporations' long-term, wide acceptance of AR HMDs is unlikely for everyday industrial field work. Still, even in a company with a large workforce, expensive AR HMDs are meaningful and familiar to a few users dealing with specific scenarios. These specialist jobs can be worth the money spent on expensive devices if the devices truly benefit and the need is significant enough to warrant an approach of any means necessary.

Many participants expressed worries about several manufacturers entering and exiting the smart glasses market in a relatively short period. One interviewee was suspicious of the low-price manufacturers' business practices regarding not delivering security updates to older models favoring low product lifecycle instead of guiding the customer instead of purchasing a newer generation device. However, investment in time savings on maintenance has quantifiable benefits from a significant workforce. All participants expressed that the smart glasses technology advances very fast and hope to be able to use them in the future.

5 Discussion

This thesis aimed to identify interaction techniques best suitable for industrial environments and the aspects of XR solutions aiding easy adoption to field work. Analysis of the findings supports the theory that for a feasible industrial environment XR solution and its interaction techniques, an affordable and robust hands-free operation, with a freehand touchless alternative input method, of a readable display and an opportunity to receive eyes-free output, in a usable and safe manner is required.

Based on the findings of the thesis, it would appear that the needs of industrial workers are strongly intertwined around knowledge. Mainly to the convenient availability of it when help is required, and its creation as mandatory reporting is a part of their duties. The best style of display solution for industrial field work is visible when needed, readable, and operationally reliable. For an industrial maintenance technician, mid-maintenance, this is either rugged and affordable AR HMD which does not distract their job with poor readability or usability. Another option is to place their smartphone on a suitable location before maintaining equipment and then interact with the visible display without touching the device itself. Handheld displays and devices are not an option for maintenance technicians as mid-maintenance they need to have hands available for both safety and efficiency reasons.

Multimodal interaction techniques should be available for the user to pick from according to situation or preference. Hands and eyes need to be available for work, and hands-free input techniques are favored. The most mature hands-free modality for use is voice, but it was seen to have limited use cases. The use of gaze as an input technique has potential for implicit use and as a supportive modality rather than a primary modality for selection/manipulation tasks. One hand used touchless input methods are more suitable than on-device and on-body input methods, as users might be wearing work gloves or have dirty hands. All interaction techniques are judged by their safety and reliability in maintenance, so any input or output technique might have potential if evidence exists of it. Therefore, all introduced wearable technology must be ergonomic and compatible with personal protective equipment and not endanger them by obstructing the user's mobility, line of sight or divert their awareness. In fact, all technology introduced to industrial work should act as an enabler only and allow users to stay focused on their job duties.

Beyond the above introduction of the grounded theory, the interpretation of the results will be dived in more depth in section 5.1. Supplementary major literature review guided by the grounded theory is presented in section 5.2. Section 5.3 will deliver the implications, and in section 5.4, the limitations of the research and opportunities for further work are examined.

5.1 Interpretation of the interview findings

Each participant interviewed had multiple principles on how to engage in the design of industrial XR interfaces. This chapter tries to interpret the findings and the driving principles of interaction-related decisions needed to make. Some of the principles relate to dealing with the technology acceptance of the users, how to cope with safety and environmental restrictions, the proper use of technologies on offer, and what kind of tasks XR technologies should support.

Unexpected was that in total, the greatest number of perceptions in the interviews indicated challenges related to users' technological competency and acceptance regarding new technologies. However, it would seem that most of these were in actuality related to usability issues on the hardware elements of the current generation of user interfaces. The effect of these usability issues on the designers was that the perceived potential of XR-related display and input solutions to create value is low. This thesis did not collect perceptions of the usability graphical user interfaces and the most prominent usability problems in the findings related to interaction. These problems were with the readability of HMDs, handheld devices, and the recognition of touchless input attempts.

An industrial worker should perceive the XR solution as a useful tool, a positive enabler, and a reducer of busy work. According to Kaasinen et al. [2015], setting user experience goals can smooth this thesis' issues related to introducing new technologies to the industrial usage context by drawing attention to the positive experiences and minimizing reservations towards negative experiences.

5.1.1 Designing for focus

The findings of this thesis indicate that industrial field workers need to be able to focus on performing their job, and the technology adopted to the context should be an enabler of it. XR technologies implemented in the context should minimize or be such that no additional burden to the user's information-processing is created. These burdens can be to the user's perception, cognition, and/or physical ergonomics. In industrial worksites all relate to occupational safety. Based on results, a designer of XR technologies for the industrial field work context is to note what the user is working on and design for them to accomplish things more efficiently and not interfere with their goals. A consideration that is in line with Aromaa et al. [2016] suggestions of first considering the goal and tasks of a user when creating systems with wearables or AR for industrial maintenance technicians.

It is not entirely the case, but it can be at times that maintenance technicians need a method to activate their display and perform the required selection/manipulation. Notable is that the XR technologies exist parallel to the real world, and as such, the design of the interaction techniques should take note of it. When an information need emerges during servicing the equipment, and the industrial maintenance technician's focus is on the main

task of performing the maintenance, there are three options for them. They can take a small moment to activate their display and to check something from their assistive display or input some information, interacting with it before returning their focus back to the maintenance. They can have the necessary information they need already open and glance at it or give input occasionally mid-servicing. Thirdly, they have an intelligent system tracking their actions and behavior by recognizing them in the background and, as a result of it, creates an output. The second option allows the user to focus on maintenance as their focused interaction and having their peripheral attention be bothered occasionally by the secondary task of the XR interface. This method of interaction where the second task is running alongside the main task can be called peripheral interaction, or reflexive interaction if the interface is closer to the third option of implicit interaction in which the user's awareness of the system is not expected [Matthies et al., 2019]. The theory presented here keeps the first option as a possibility but tries to shift the interaction to less focus requiring a second option, and in the future, with more advances in computing, the third option is a possibility.

To design for the support and empowerment of the end-user means supporting the user only where they require assistance. This viewpoint is very different from commercial applications and interaction patterns, where outside of the utility category software is made to compete of the user's attention, and immersive interaction techniques are encouraged. Technology suitable for field operations focused on getting things done for the single user with high mobility needs with a robust solution that stays out of their way when required. Consequently, the XR tool should stay in the background and respond when needed. If true, as a result, an XR solution for maintenance journals would get relatively little use with seasoned professionals with much know-how under their belts unless they use the tools to create content or aids during maintenance transfer to the junior members of staff, or better yet all use the tools available to conduct maintenance reporting. Kaasinen et al. [2018] found that regardless of challenges posed; maintenance technicians would accept using the wearable technology during the maintenance operations if it would contribute to a reduction in maintenance reporting after servicing.

5.1.2 Readable display

Similarly to the findings, visibility of information was also recognized by Masood and Egger [2019] to be the highest relevance success factor for industrial AR system implementation. Readable visual output was seen as the only required output, and applications with an audio need to have an option to it be on-off switchable when the user desires it. It can be argued that industrial maintenance technicians should be able to choose their own display solution as personal preference, and for example, prescription glasses can affect readability. It was found that, with conscious preparation ahead of

maintenance and planning for the ergonomic viewing of it, a smartphone can also be always visible and readable during industrial work.

For smart glasses, the above line-of-sight placed text reduces comprehension significantly instead of normal line-of-sight and below it [Rzayev et al., 2018]. Center of the vision performing better in speed, comfort, and learnability than bottom-right positioned text for head-worn displays [Lin et al., 2021]. A worker on an industrial shop floor, head-worn optical see-through display with a minimum of 30◦ horizontal FOV has been suggested for a desirable level of user experience due to delays with video-based display technologies [Syberfeldt et al., 2017]. However, according to findings for maintenance technicians, the optical see-through display does not perform well as they suffer to produce a bright enough display output to work in well-lit environments or outside. Due to this, many optical see-through devices restrict incoming light from the sides of the HMD with a bulkier design, but in a worksite, peripheral vision is needed to catch potential hazards in the environment. This would indicate that a monocular pass-through display, like Realwear HMT-1, placed on or right under users' line of sight, has better suitability for industrial maintenance over optical see-through displays. However, if a comfortably worn and bright enough binocular HMD that does not suffer from accommodation-convergence issues and meets safety goggle standards comes along, the previous point would be required to be analyzed again.

Display-related safety concern was how eyes-free could an HMD solution be when there is potentially an always-on and present screen within the FOV. Aromaa et al. [2018] noted that following procedural steps from a display in the view frustum might weaken maintenance technicians' situational awareness. The issue of accommodation-convergence must be resolved for the visual interface of a binocular HMD not to suddenly steal the worker's focus and attention in a manner leading to a risk of a workplace accident. Based on the case, it might be an option to hide the UI while the user is not performing maintenance standing still as the user needs to keep their focus on the actual environment they are located. Additionally, a computer vision-based system could be used to improve safety by mitigating hazards and proactively informing workers of potentially dangerous situations [Kim et al., 2017].

5.1.3 Robust technology use

Findings showed that maintenance professionals with a reserved attitude towards new technologies have a low frustration tolerance of low-performance tools. Consequently, all the interaction techniques implemented in industrial field work need to be reliable in their performance to deliver a good user experience. Besides this study, user interface usability is an important success of industrial AR system implementation recognized by Masood and Egger's [2019] findings. Additionally, Aromaa et al. [2016] found that maintenance

technicians prefer fewer devices required to interact with, as shifting devices while interacting with the system was regarded as problematic. Selecting an input method, including the proper feedback methods, reduces the total amount of devices required and the complexity of the system. The safe use of interaction techniques in industrial environments requires a user not to be distracted by usability issues related to output or input technologies. User's hands need to be kept free of handheld devices to have full use of them if the need arises. As a result, handheld devices can be directly discarded as an unfit input option for a maintenance technician mid-maintenance. This does not mean the implemented XR solution should not support touch inputs that are often activated eyes-free. A low cognitive load is essential in eyes-free interactions because it relates to safety [Yi et al., 2012]. If a touch input device can be operated with one hand for a short period of time in a reliable and usable manner and handle dust and grease that might rub off from the user's fingers or gloves, it can be regarded as a useful alternative or complementary interaction method. Findings insist that no unimodal input solutions, like speech only, can be implemented for industrial field maintenance. This is due to the need to accommodate the user's personal preferences, both personal inhibitions and those relating to social setting, and situational challenges related to environment and task, like having both hands and eyes engaged on the job.

A usable XR solution needs to support multimodal input to enable several different paths for task completion. Capturing mid-air gestures, speech, and gaze input modalities supported by HoloLens 2 was seen as a model solution for the multimodal support of input techniques. However, the technology utilized needs to be compatible with existing safety procedures and equipment. Additionally, all wearable technology introduced to the industrial environments needs to be compatible with the personal protective equipment and clothing technicians are required to wear when the environment or safety procedures dictate it and mobility demands of the job performed. HoloLens 2 physical dimensions of weight and size, limited FOV, short battery life, poor high and low light performance, and price were seen as deterrents of outfitting workers operating under unknown working conditions. Safe and productive use cases for HoloLens 2 are in familiar and optimally lit environments, like industrial plant control rooms and training simulations where its capability for the optical mid-air hand gesture capture requiring the user's hand to be extended in front of them can be used safely. Similar restrictions to HoloLens products were seen to apply also to lighter and relatively more affordable smart glasses. Nevertheless, two other significant technical factors made the current generation of smart glasses unsuitable for industrial field work. These were: the lack of input modalities they support natively and poor connectivity to external data sources and potential external input devices.

Devices used in industrial environments need a durable build quality to handle physical damage and the environmental conditions present at the worksite. Industrial maintenance technicians often work in the customer company's facilities, and the environments can be challenging, including noise, high temperatures and humidity, confined spaces, high places, dusty and greasy surfaces, and poor lighting [Aromaa et al., 2016; Kaasinen et al., 2018]. The rugged appearance of technology means that the device must handle the physical conditions and signal to users that they can use the device without a worry of damaging it in normal use. Maintenance technicians can become concerned over the expensive technology they are using, something Aromaa et al. [2018] noted. The appeared lack of robustness might even lead to avoiding using the technology, an argument for technology that should tolerate the harsh industry conditions by being robust enough and cheap enough to be practical or at least appear to the users in a way no to evoke the aforementioned reaction.

It is not only about the risk of technical failure slowing down the deployment of smart glasses to industrial field work mass use. A large portion might decide that they do not want to use the smart glasses solution. Familiarity from private life with interaction techniques, especially input modalities outside of touch interfaces, was deemed to affect the technology acceptance of new interaction techniques. Individual's own mental models from consumer technology are important and the reason for seeing voice-activated user interfaces as the most prominent eyes and hands-free technology to use. The user's age was found to be a factor the designers assume to affect both the competency of use and the willingness to experiment with devices that create value but are lacking in usability. Wearing wearable technology can make technicians self-conscious about looking silly or amateurish, implying that technology should be integrated into PPE, vests, or other worn gear. An HMD could pose as a helmet visor for it to not stand out from regular equipment.

A short-term wide corporate acceptance and utilization of smart glasses is not expected in light of physical dimensions and weight of the device, FOV limitations, and easy readability of text not been taken a significant and serviceable leap yet. The most dominant obstacle standing in the way of equipping a large workforce with AR HMDs is the price of an individual unit. However, for specific maintenance jobs with a more considerable monetary sum attached to their completion, costlier AR HMDs were seen as an affordable option if they would bring such benefits to make them worth the price. As a result, the business case of the industrial company dictates the point where an industrial company might start introducing head-worn display solutions to their employees.

Currently, the smartphone is the go-to device of industrial maintenance technicians as it is always with them for professional and personal use. A smartphone is a robust provenly piece of technology that works as a hub for peripherals and provides internet connections even to smart glasses. Focusing on information design that supports various

terminal devices can circumvent designing for only a wearable display. Creating the same app for smartphones and smart glasses is a good idea as no one should be forced to use an HMD if they do not want to. Carefully designed content for small displays, like smart glasses, works in different display devices, including smartphones and smartwatches on top of other wearable displays [Siltanen and Heinonen, 2020].

To state that consumer-side solutions and experiences need to become more commonplace for users to gain technology acceptance is too simplistic. Interview findings showed that a technology gaining popularity is a double-edged sword. An example is that even now some technicians forgo using a smartphone as technician perceives that the client may think ill of them for checking social media or handling personal matters. In able to move both user acceptance and industrial XR product development forward, it is much easier to have a palette of interface devices available for the workers and instruct the users to take advantage of all or whatever is familiar or comfortable to them.

5.2 Literature review of suitable input methodologies

Collection for more material to fortify the grounded theory discovered and to answer the first research question beyond the interview findings. It was decided to conduct a literature review of input techniques. This review was guided by the interpretations of the findings made in section 5.1 and is divided into two parts. The first part delves into hands-free input techniques and the second into freehand input techniques.

Findings showed that industrial XR smartphone systems need to extend the input range of a touchscreen through non-touch interaction techniques. Some participants expressed that their companies do not experiment with the development of interaction techniques but rather trust more established and startup companies to develop the input solutions for their use. While the industrial sector has a lot of pull within the XR marketspace to entice developers to benefit from it before as a generally recognized solution as the smartphone's touchscreen exists, there is still room to innovate and expedite the technology maturation. In part, this chapter tries to show that there is room for subcontractors and industrial companies themselves to try and build with the technologies already available. Portal-ble by Qian et al. [2019] shows that it is possible to build a novel touchless interaction technique system with a purse-sized portable computation and battery unit. The unit, comprised of Intel CS325 computetick and a 22,000 mAh battery, acted as a wireless data relay to bypass the technical issue of Leap Motion not directly supporting smartphones [Qian et al., 2019]. It could be ergonomically feasible to fit a unit with similar dimensions to an industrial worker's vest or belt. Alternatively, it can be left stationary together with a smartphone stand. While an expensive prototyping solution, it allows high-mobility experiments to be done in industrial field work environments. Because of this possibility, not all interaction technique solution examples presented are yet portable,

without a connection to a desktop computer, or cable-free. However, interaction techniques need to be such that they explicitly do not interfere with work.

5.2.1 Hands-free input

Sufficiently said that if the user's hands are occupied by holding a device, it is more difficult for them to perform other tasks, like industrial maintenance, while operating a controller. Making hands-free modalities appealing to service technicians who often have both hands on the job.

Spoken input was the most mature interaction technique and the one with the least resistance to user adoption, compared to other input modality technologies might face, if familiar from outside work-life on handheld devices. Easy to implement real-time natural speech recognition technology is becoming more commercially available [Speechly 2020].

On HoloLens, a user can use shortcuts by reading any button's name aloud to activate it in a "see it, say it"-manner or converse with a digital agent to whom to delegate tasks [Microsoft, 2020a]. Utilizing HoloLens speech recognition, a participant had built a chatbot capable of assisting on maintenance by returning information from an artificial intelligence-built knowledgebase, with a text query created from user speech and Amazon Web Services Polly integration able to read the available information in return. Thus, spoken input works well for easily recalled, and verbalized tasks, including dictation or selecting categorical items, but not for object interaction performed continuously, like scaling an object or changing audio volume level. Two participants did note that spoken input alone does not fit their needs as they cannot have a device that is merely speech controlled due to environmental and social settings inhibiting its usefulness. HoloLens 2 is an example of how narrow the voice recognition support is based on its documentation; it supports speech commands and dictation for English dialects native to four countries, two French dialects, German, Italian, Spanish, Japanese, and simplified Chinese [Microsoft, 2020b]. This example and this thesis findings suggest that a non-native language user may meet voice recognition issues when giving speech commands.

For an individual wary of causing disturbance and obtrusion, voice input may be less appealing than other input approaches [Kollee et al., 2014; Shanhe Yi et al., 2016]. Of Tung et al. [2015] participants, only 2% of smart glasses users in a public setting chose to use speech input. For those uncomfortable speaking instructions out loud, an exciting development could be the use of silent speech commands. Under real-world scenarios, EchoWhisper recognized the content of silent speech up to 45 words, with an error rate of 8.33%, by capturing mouth and tongue movements without vocalization using the micro-Doppler effect via leveraging the smartphone's two microphones and speaker [Gao et al., 2020].

Eye gaze and head movement used as inputs are not yet supported by most AR HMDs outside of HoloLens 2. Based on findings, gaze control is not suitable for explicit selection and manipulation tasks. It is not robust or swift to use on its own. An industrial maintenance technician must keep an eye on the visual information the maintained equipment could be providing them, like LED flickering. Additionally, participants were emphatic that gaze should only be a last option for selection/manipulation tasks. Microsoft HoloLens documentation confirms this as the "eye-gaze and commit" input model is recommended as a third choice for HoloLens 2 [Microsoft, 2020c]. However, participants had not yet leveraged gaze explicitly or implicitly.

Gaze can be used to extend user's abilities. Explicitly, gaze can be used to point and select erroneous words by dwelling on the word for 0.8 seconds [Sengupta et al., 2020]. Sengupta et al. [2020] suggested that gaze and dwell could assist text entry and editing in head-mounted multimodal displays. Implicit gaze interaction could support contextual knowledge sharing in an industrial context, like tacit knowledge transfer on industrial maintenance jobs. Users of ubiGaze can attach noticeable messages onto real-world objects and retrieve the messages from the objects with the aid of gaze direction [Bâce et al., 2016]. For example, a maintenance technician could leave a note for a technician visiting the equipment on the next maintenance visit. Gaze sharing can support remote collaboration to refer to objects and guide the task performer's attention based on the remote expert's attention [Higuch et al., 2016]. Gaze could also aid computer vision to label objects in cluttered environments by detecting objects of interest from real-time visual gaze data analysis together with convolutional neural network object detection algorithms on egocentric video [Silva Machado et al., 2019].

Gaze has an opportunity to shine as a pointing and selection input modality in combination with other modalities. In virtual object manipulation, gaze can be used for a pointing object selection together with mid-air gestures to perform 6-DoF object manipulation [Slambekova et al., 2012]. Combination of gaze and touch text entry for smart glasses outperformed eye-only and touch-only typing, achieving 11.04 words per minute [Ahn and Lee, 2019]. While head pointing is slower than eye-gaze input alone for target selection, it can, in refined combination with gaze, achieve robust and accurate target selection for AR applications [Kytö et al., 2018]. For peripheral or reflexive interfaces, vibrotactile feedback could be impactful and subtle feedback channel. According to Rantala et al. [2020], it is at least as good as auditory and visual feedback for gaze in task performance and user satisfaction.

One of the earliest uses of head-direction-based pointing as an interaction technique in VE included head-directed navigation and object selection [Mine, 1995]. Hindering the utilization of head movements as a pointing device are the ergonomic restrictions

accompanying user moving their head for long periods. The neck is strained with oscillation or bobs in a set frequency or duration, especially if they are wearing a heavy helmet. Simulating computer mouse point and click interaction, "head-gaze and commit" is a recommended input method of HoloLens 1 and recommended as a third-choice input model of HoloLens 2 [Microsoft, 2020c].

Another restriction is a practical one. Maintenance technician needs to keep their head directed on the job they are working on and cannot use their head for directional input. Head movement text input might not offer a method for maintenance journals but could assist in inputting search queries hands-free. Handheld accelerometer-only text input is possible in lab environments with 5.4 words per minute [Jones et al., 2010].

A more viable implementation area than head-tilt gestures would be to use data from smart glasses' accelerometers and gyroscopes to enable fingerprinting for security. Data from smart glasses could be security-related for head movement by utilizing a sequence of head movements as authentication input robust against imitation attacks. Google Glass can be used to monitor an individual's own patterns of head movement in answer to external audio stimuli for authentication, with an acceptance rate of 95.57% [Li et al., 2016]. GlassGesture by Yi et al. [2016] had 92.28% accuracy and inhibited imitators from successfully masquerading as the authorized user.

Face and tongue gestures present a possibility for more discreet inputs than mid-air gestures and speech but are currently less technically ready to be used within the industrial XR setting. Using face, cheeks, and tongue as captured inputs were unmentioned in the interviews as commercial solutions are not available. Tongue gesturing may be more subtle and requiring less muscle activation, and less exhausting than regular talking because it is plausible to retain the jaw, closed [Li et al., 2019]. Sensing interfaces relying on facial gestures, especially tongue sensors, can be unobtrusive and limited in input bandwidth or obtrusive with sensors that do not only take space but require wires as well.

Facial muscle deformation information can be captured from an inertial measurement unit within wireless earable, and both smile and frown can be identified with high accuracy in a controlled non-conversational setting [Lee et al., 2019]. Another method could be to add a discrete piece of equipment in front of the mouth to capture mouth gestures, like a surgical mask embedded with a mutual-capacitance sensor recognizing five distinct vowel-based non-verbal mouth shapes [Suzuki et al., 2020].

Tracking the absolute position of the tongue and recognizing non-vocalized speech is possible with wired and a difficult to mass-produce array of 124 capacitive touch sensors attached to dental impression mouthpiece device on the roof of the mouth [Li et al., 2019]. Computer-connected TongueBoard recognizes non-vocalized words with 91.01% accuracy for native American English speakers and 77.76% accuracy for non-native speakers

[Li et al., 2019]. TongueBoard is an example of the fidelity of input computers can receive from the tongue, but a special interface requiring a wired connection to a computer does not allow enough mobility to be considered as a usable input technique option for most industrial environments. By limiting the fidelity of input in favor of mobility, other input device options surface. In-ear sensor furnished with a light-emitting diode can recognize binary input of the tongue being held on the roof of the mouth for a second with total accuracy while resting or chewing and near-complete while walking, with a small sample size of five people [Taniguchi et al., 2018].

Gallego Cascón et al. [2019] investigated the comfortable and subtle placement, form, dimension, and obstructiveness of a wireless intraoral input interface, resembling an edible object to achieve equilibrium with providing higher fidelity tongue gestures and social acceptability. ChewIt allows users to perform ten gestures, like rolling the object and molar bite, both simply and discreetly [Gallego Cascón et al., 2019]. Users' felt comfortable holding the object for a minimum of 15 minutes in their mouth while performing their everyday tasks [Gallego Cascón et al., 2019]. While ChewIt is an intrusive piece of technology testing, intraoral interface in a professional setting could yield insight on the value and the relational acceptancy of such an input device.

5.2.2 Freehand input

Humans are skilled in accommodating to different situations whether tools or tasks occupy their hands or not. [Zheng et al., 2015]. Indicating that, while safety conscious, performing input hands-free may not be the inevitable requirement for challenging industrial environments. Still, performing touch interaction might not suit dirty and greasy environments, especially with touchscreens when gloves are worn. Additionally, finger-worn devices or other devices that interfere with work or wearing PPE gloves are not suitable for industrial work.

The interface of HoloLens and Android smart glasses rely on WIMP (Windows, Icons, Menus, Pointers) paradigm, and raising a hand to the device can become encumbered and error-prone from prolonged and fatiguing task completion. Especially optically captured static mid-air gestural interaction requires lengthy dwelling times for recognition [Istance et al., 2008]. Figuring gesture beginning and endpoint is vital, and currently, high accuracy in segmenting mid-air has only been achieved with electromyography [Chen et al., 2016]. Based on findings, good mid-air gestures are dynamic as they should aim to be realistic in the sense of the natural world, and while a minimal set of simple symbolic gestures are recallable, their performance to be recognized is difficult for users. Interview findings in section 4.3 showed results of mid-air gesture recognition with optical imaging technologies to be unreliable during extreme lighting conditions and posing safety and mobility incompatibilities of performing broad arching hand movements compared to closer hover style and wrist mouse capture methods. Desirable characteristics of gesture

recognition technology of note consist of robust initialization and re-initialization, robustness to background clutter, independence of illumination, and computational effectiveness [Moeslund and Nørgaard, 2003]. Next, free hand-based interaction methods robust against varying light conditions will be presented. Most of them non-optical except infrared technologies. Mentionable non-optical hand-tracking technologies are electrostatic field modulation and capacitive sensing, electromagnetic sensing, ultrasonic acoustic soundwave modulation sensing [Rise and Alsos, 2020].

Hand activity recognition can be used to open possibilities for a more implicit style of interaction (see subsection 5.1.1). One interviewee mentioned the possibility of tracking fine-grained hand activities for matching maintenance procedures stages in the digital instruction manual to recognize whether a step might have been missed (see subsection 4.1.2).

Still obtrusive but capable method of inferring hand positions could be worn on a finger and wrist. Thumb ring surface-transducer sent chirps captured by four receivers distributed on the wrist as acoustic resonance traveling different paths within the hand were used to discriminate up to 22 fine-grained hand poses, thumb tap locations of 12 phalanges of the hand, and ten poses from American Sign Language with an accuracy of 93.77% and 95.64%. [Zhang et al., 2018].

In the future, data cloves capturing in real-time could replace the field worker's safety gloves [Brice et al., 2020]. A comfortable wearable data glove made out of silicone composite and textile layers with an embedded stretch-sensing capacitive sensor has been proven to be effective in reconstructing the gestures in real-time with high accuracy by exploiting the spatial layout of the sensor [Glauser et al., 2019]. An off-the-shelf approach would be modifying existing technology for hand activity collection. Laput and Harrison [Laput and Harrison, 2019] collected 25 hand activities from a commodity smart-watch by modifying one to capture coarse hand movement and orientation and bio-acoustic data proving 95.2% average accuracy. Buddhika et al. [2019] collected data of smartwatch photoplethysmogram sensor with refined enough detail to recognize the force of hand action in two levels, as an independent channel of input, on various grasping gestures types. From the user experience side, personal activity trackers for industrial professionals should be reliable, unobtrusive, and discreet even with their encouraging feedback [Heikkilä et al., 2018]. The work of Heikkilä et al. [2018] has listed multiple ethical considerations for introducing personal activity tracking into workplace context.

Touchless near-surface mid-air gestures can be done with a hand or fingers near a sensor and are helpful in situations where contactless use of technology is favored by the user, for example, when the hands are unclean when working. Near-surface and wrist-worn sensor gesture recognition hand gestures are easy for operating a WIMP interface with semaphoric-stroke gestures for directional cues, and a limited amount of static hand poses supported in the confined space. They can extend the input technique capabilities of smartphones and are also compatible with more power-efficient and resource-constrained devices like smart glasses. Waving a hand in front of the user's vision has implications on obstructing the view and social acceptance as 63% of vision-based gesture interaction on smart glasses attempts failed due to difficulties produced by social acceptability and fatigue [Tung et al., 2015]. In comparison, non-head-tracked mid-air gestures strive to implement subtler interaction for greater acceptance in social settings and the ergonomic benefit of preventing arm fatigue present in head-worn capture. Touch-free swipe gestures have the potential to complement spoken instructions to avoid voice-only interfaces repetitious spoken commands [Saad et al., 2018].

Recognizing five iconic-dynamic thumb-tip gestures through sensing variations in thermal signals fingers give out by using an infrared pyroelectric sensor, Pyro achieved 93.9% cross-validation accuracy [Gong et al., 2017]. One-handed ultrasonic doppler sonar, with one transmitter and four receivers, gesture recognition for handheld devices is usable even when the user is wearing gloves, and the touch screen is not functional [Saad et al., 2018]. It reached high classification accuracy of 96% with seven directional semaphoric-stroke gestures used for menu control [Saad et al., 2018]. Vision-based near-surface multi-finger tracking is possible with a standard smartphone mono front camera focusing on hands and fingers interacting above the touchscreen surface by providing the camera a stereo vision with a prism mirror [Yu et al., 2019]. Currently, the algorithms outputting the depth image of the hands from a monocular color image feed cannot be run on commodity smart-phone devices as only a PC deep learning implementation can provide a real-time performance of 30 frames per second for estimating and identifying each finger's 3D position and a gripping posture of the hands [Yu et al., 2019]. Similarly, a fisheye camera attached at the bottom of a touchscreen as an image source can locate a fingertip with 98.0% accuracy and classify contact fingers with an accuracy of 89.7% to thumbs, index fingers, and others [Park et al., 2020].

Goole's Soli, a short-range extremely high-frequency radar-based hand sensing, achieved an 87% average recognition rate on 11 dynamic gestures [Wang et al., 2016]. Soli mmWave radar has been implemented commercially in the Google Pixel 4 smartphone [Google, 2020a]. Hover gestures can be sensed through work apparel with near-field antennas. A sewn pair of loop antennas from conductive thread onto a textile substrate allowed Doppler motion-sensing of 11 touchless unimanual and semaphoric-

stroke and thumb-to-finger iconic-dynamic gestures performed 10 cm above them, as well as one touch gesture, with 92.8% cross-validation accuracy [Wu et al., 2020]. A technician can discretely perform micro mid-air gestures to interact with the display. For example, sliding the thumb on the length of the index finger, with the hand dangling freely beside the body or other safe area close to the body.

Based on this thesis' findings, free-hand gesture movements with near-surface sensors and wrist-worn trackers could be safer in industrial field work than optically captured wide arching mid-air gestures. However, near-surface sensors, compared to head-worn optical capture methods, are aware only of the presence of a hand above them or the motion of the fingers rather than the location of the whole arm in relation to the output of an AR HMD. They have a restricted range of expressivity as direct manipulation input model-style interaction with virtual objects is not possible, and pointing gestures cannot be performed. Methods of pointing gestures for smart classes do exist. Utilizing a monocular camera on cheap and low-power hardware smart glasses mid-air interaction technique, UbiPoint is a markerless fingertip detection system that delivered a better user experience compared to previously available smart glass interaction techniques, and users completed typical tasks 1.82 times faster than when using the original hardware [Lee et al., 2020].

Distant point and click device for large screens was found practical with Myopoint using a combination of forearm mounted electromyography sensor for detecting arm muscle contraction and relaxation and inertial motion sensors for arm movement detection [Haque et al., 2015]. AR HMD interaction can be expanded with the smartwatch pointing technique [Chen et al., 2020]. Chen et al. [2020] found viewport-based pointing to be more efficient and low-fatigue input for target selection for HMDs than world-based pointing. World-based pointing allows the cursor to move independently of the HMD's field-of-view.

A wristband equipped with inertial sensors called the Kinemic Band can be used as a wrist mouse for viewport-based pointing for smart glasses and mobile devices and can recognize 13 symbolic mid-aid gestures and a wrist touch [Amma et al., 2016]. The wrist can also be used as a one-handed continual input device for the same hand display is worn on the arm, and the user can interact without tilting the display and losing sight of it [Gong et al. 2016]. To a same-side smartwatch with an array of proximity sensors on the watchband, WristWhirl can activate four directional marks at an average rate of half a second and four free-form shapes at 1.5 seconds with an accuracy of 93.8% [Gong et al., 2016]. A different style of interaction would be to move a hand against a surface. Eyes-free directional stroke gestures can be performed on any surface using a vision-based motion sensor on a wrist-strap [Yeo et al., 2020].

Hand operable touch input. Alallah et al. [2018] found a ring controller or a smart glasses touchpad preferred over speech and mid-air gesture controls on smart glasses. Eyes-free interaction is facilitated by tactile feedback from touching the controls [Yi et al., 2012]. This eyes-free use and feedback could lower the maintenance technician's cognitive efforts while focused on their job. Tactile feedback providing gesturing and tapping touch input can also happen with non-handheld touch-sensing wearable devices and body surfaces [Tung et al., 2015]. Locating the input area with a hand alone and the tactile feedback from the interaction allows the user to pay heed to the surroundings, reducing distraction and adding to user safety [Fuentes and Bastian, 2009]. This ability could affect the focus of the user by moving the interaction to the background.

Boldu et al. [2018] built and evaluated an interface for one-handed thumb-to-ring touch gestures that does not interfere with physical activity. Among gestures selected to minimize distraction and cognitive load, tap and swipe gestures across the ring were the most comfortable to perform simultaneously with a running exercise suitable for brief and discreet interactions rather than complicated manipulation of interfaces [Boldu et al., 2018]. For the industrial environment, Brice et al. [2020] purpose-built a larger glove operable three-button controller that attaches to the frame of a HoloLens 1. It performs in line with other input modalities of HoloLens 1 in input time, error count, and usability, with 81,75 on system usability scale, when evaluated within a controlled laboratory setting on an industrial task assistive system [Brice et al., 2020].

Interaction can happen with a touch-sensitive surface within arm's reach. Flexible and versatile wearable textile interfaces are created using an off-the-shelf embroidery machine to augment an arbitrary base fabric with a sensing device. They can be textile-based pressure sensors, resistive textiles, or conductive yarns, but together with the fabric textile interfaces, combine the practical and conceptual challenges related to electronics, clothing, and interaction design.

Jacquard is a highly conductive yarn woven into an area in a piece of clothing on allowing a user to perform touch gestures on the continuous sensing area [Poupyrev et al., 2016]. The piece of clothing requires an additional module to be carried within the Jacquard-enabled clothing to enable wireless communication to a user's display device [Google, 2020b]. A touch-sensitive area that supports swipe and tap gestures could be its own device like a sleeve for the forearm [Schneegass and Voit, 2016]. Dampening the impact of solutions based on conductive wire devices in industrial settings is the dust and grease that can rub off to a jacket sleeve, or the fingers might accidentally transfer the dirt.

In comparison, pushbuttons would at least function with gloves on. Goudswaard et al. [2020] used digital embroidery to interweave 3D printed mechanical pushbuttons into fabrics for eyes-free use. The method allows them to fabricate the electrical circuitry and

3D print the mechanical structure directly onto the pre-stretched Lycra layer and integrate the two layers together [Goudswaard et al., 2020]. The textile-based buttons are locatable landmarks by touch. They return tactile feedback when activated and useful for wearable on-body interactions [Goudswaard et al., 2020]. Aigner et al. [2020] embroidered unobtrusive resistive pressure-sensitive sensors, showing very low activation and good dynamic range, onto stabilized fabric.

5.3 Implications

This section discusses this work's implications first to academia in subsection 5.3.1 then to the industry in 5.3.2.

5.3.1 Implications for academia

The findings of this thesis build on the existing evidence of Aromaa et al. [2016], who suggested that careful thought needs to be used to "the goals and tasks of the user, the usage environment, the whole system of tools and devices, and contextual information" when creating industrial maintenance technicians systems with wearables and AR. Del Amo et al. [2018] identified factors and attributes in ISO 9241-11, for Human-Computer Interaction systems, regarding a maintenance technician-centered perspective of AR systems and validated the relevant factors through a questionnaire with six AR-design experts. This thesis clarifies their results by indicating tactile input interaction category not to be entirely suitable for maintenance technicians. Only those tactile interfaces that can be operated onehanded wearing a glove and proof to substances like grease are suitable. Otherwise, this study is in line with the factors found to be relevant by them. Additionally, this thesis findings clarify how decisions are made with the framework they have created to select suitable AR devices.

Regarding smart glasses, thesis findings were in line with Syberfeldt et al. [2017]. Smart glasses have lesser battery life, display size, input interface options compared to smartphones. They identified what kind of aspects of an HMD would suit the needs of a worker on an industrial shop floor by evaluating commercially available devices on five parameters: powering method, weight, the field of view, battery life, and optics. Based on the findings of this thesis, the parameters apply to maintenance technician's devices and are extendable to other industry work roles, but adding connectivity to the parameters should be done as it is the method that helps a user connect to helpful information or aid.

Findings suggest that industrial workers' on-the-job mobility requirements affect the type of readable display that can be kept visible during work. For industrial manufacturing shop workers, head-worn AR devices are the best-suited display style [Syberfeldt et al., 2017]. It appears that industrial field maintenance technician's mobility requirements are less intensive in comparison to assembly workers installing seats to an aircraft or industrial plant technicians. The nature of the work of maintenance technicians is very mobile,

but when the servicing begins, they are stationary or moving in a limited area for the duration of maintenance. Indicating maintenance technicians to be only semi-mobile in their work compared to an assembly worker and a smartphone on a temporary holder could be enough for a display solution for industrial maintenance before an ideal head-worn device is available.

Mobility is also affected by the body locations of wearable input and output devices. Khalaf et al. [2020] analyzed 84 wearable input devices resulting in a prescriptive design framework of devices consisting of four stages to be analyzed: the interactivity type, connected output modalities, mobility allowed, and body location. Their work enables designers to define which devices satisfy the user's situation's distinct requirements and limitations. On Khalaf et al. [2020] framework, the body location of a wearable device should be considered through categories of the upper body, lower body, and freeform. The environment dictates PPE needs for industrial workers, and they can limit mobility or be an ergonomic burden taken on for safety. Body locations where wearables could have overlap with the PPE of industrial workers should be identified. However, industrial field maintenance technicians do maintenance jobs on-site in customer locations that meet the most unfamiliar environment with different needs for PPE and have it affect their mobility even more. For the framework to be more helpful in designing wearables for industrial field workers, the model could have higher fidelity than Khalaf et al. [2020] framework model suggests. The revised framework would need to distinguish upper body considered devices for torso, head, shoulders, arms, wrists, hands, and fingers. Additionally, based on findings, the PPE requirements are not constant for maintenance technicians working on various worksites during the shift, as they strip and put more on depending on the job and the phase.

Masood and Egger [2019] studied the implementation success factors influencing industrial AR by utilizing technology, organization, environment framework-based anonymous questionnaire, through social media, for 84 professionals who had been involved with industrial AR projects. They only collected information on what factors are important, and results were ranked in a relevance graph of quantitative count and qualitative importance. The results here indicate the same points of importance, but the findings collected here contribute to their work by clarifying why these factors are important. This thesis connects their low importance factor of hardware robustness to be a contributor to the high importance success factor of user acceptance. Also, this thesis finds that emphasizing user co-creation by including users in implementation, a medium relevance success factor in the work by Masood and Egger [2019], could yield an avenue to user acceptance. Together, this work and theirs can provide guidance on which aspects the industry and device manufacturers should focus on for practical utilization of immersive technologies.

5.3.2 Implications for industry

XR solutions themselves have a multitude of use-cases for them to demonstrate their value in industrial field operations. Industrial maintenance technicians need help quickly from manuals and remote assistance when required mid-maintenance, more so with less experienced employees. Everyone in maintenance is required to report their work and add to maintenance journals, and this could be easily done along with maintenance. It should also be enabled for the maintenance technicians to pass along their knowledge to peers by creating learning materials easily on the job. Additionally, positioning technology and spatially mapped environments through camera could allow maintenance journals to be on locations. These could even be attached to parts of the equipment for technicians to leave metaverse notes for the next maintainer's visit. Much of the potential for value creation exists in shifting preparatory and post-maintenance actions to just-in-time actions taken.

In the industrial context, the verdict on the most productive input method and display device heavily rely on the nature of the user's job, environment, objective, and the XR device type. The interactivity needs of the proposed industrial application should be mapped out based on the worker's problem worth solving. The problem can be affected with implicit or explicit means depending on how the service is designed to function. If possible, for efficiency and user's technological acceptance, both forms of interaction should be implemented. The user does not need to be always aware of implicit interaction. Thus, it can demonstrate for the user at selected times its value and give even a timid user beneficial reasons for engaging in explicit forms of interaction with new and emerging input technologies.

For industrial scenarios, voice input is considered the most perfected hands-free input technique. At least a microphone is recommended in industrial assembly work for enabling hands-free operation via voice-based interaction [Syberfeldt et al., 2017]. However, the findings of this dissertation point to the fact that while a microphone is also recommended on its own, it does not serve the needs of industrial maintenance technicians interaction needs. Mid-air gesture control and gaze input were found to be among the most considered emerging interaction techniques for industrial applications. While seen as having potential in the future, mid-air gesturing and gaze are not fit for all users or use cases because they lack accuracy and low latency to be practical enough to meet expectations.

The explicit interaction techniques introduced to industrial environments should support multimodal means of performing the required selection/manipulation tasks. For functioning explicit input, the user needs to have a range of subtasks available for them. Figure 3 of section 2.3 introduced Bowman et al. [1999] selection/manipulation taxonomy for VE, where advice can be sought for the multimodal combinations potentially available. At a minimum, to control readable information, the user needs to perform Indication of

Object and Indication to Select tasks with industrial XR solutions. Most likely, these should be implemented with an indirect selection technique from a list or voice selection from memory if pointing with hand or gaze is not available. The selected object can be manipulated within the viewport available, and its position and orientation can be controlled with indirect control mechanisms like voice command, dynamic or static mid-air gesture, physical touch inputs, or virtual controls.

Pointing to indicate an object input interaction in fieldwork should happen in a viewport mode, meaning the cursor only travels within the confines of the user's display or FOV. Pointing interaction can happen with a wrist-worn pointing device, or when available, another means would be the tracked gaze of the user. If a pointing mechanism is available, by the means mentioned above, the selected object can be attached to the cursor controlled or be moved in relation to it and released with a similar voice, mid-air gesture, or a touch input command. When spatially mapped environments and binocular AR HMDs can be considered, object manipulation in multiple DoF becomes more relevant, and various handheld controller-free means to organize the environment need to be supported. Such interaction is closer to VE interaction, and then industrial environment suitable object interaction techniques can be chosen from Bowman et al. [1999] selection/manipulation taxonomy. It is a possibility that through the described technology combination, world-based pointing could become relevant if the user needs to be able to point to directions outside of their FOV.

Touching the object to indicated directly or to move it, like on a touchscreen, was found to be in this thesis a non-option. An indirectly selected object from a list should be selected with a voice command, mid-air gesture, button, or with implicit mean like dwelling on the object for a short moment. It is recommended against deploying traditional handheld devices in situations where the user handles tools and equipment. On-device controls like buttons and touchpads are not always suitable if safety gloves are worn or hands are dirty. On-device touch inputs are not usable on current AR HMDs devices deemed by the findings to be too delicate and unergonomic for repeated inputs.

For touch to be useful in an indication of selection, textile interfaces could be introduced to work uniforms or tool belts as non-washable elements. Especially mechanical pushbutton made out of two integrated layers of the circuit layer and elastic foam or 3D printed mechanical button designs. Harsh environments give glove-operated textile interfaces a hard time when the dirt is transferred to the interface from the hand, and mechanical buttons break down after multiple cycles of use, inferring that their use needs to be economical. However, the wireless Bluetooth dongle transmitting the information can be used again even if the circuit component of the interface cannot be recycled into new interfaces. The textile interface could be integrated into workwear if they are not washed at all, or to a belt, or be an attachable panel with Velcro.

Another option for touch interfaces would be for the user to engage with a near-surface mid-air gesture sensor with dynamic hand strokes and more fine-grained finger rub gestures to be performed. Near-surface mid-air gesture technology should be implemented safely and ergonomically. This means the interaction should happen close to the body or over an area designated safe, not just in front of the user's face. Google Pixel 4 smartphone already makes use of semaphoric-strokes for effective controlling of WIMP user interfaces. Emulating this solution should yield positive results in whatever lighting conditions exist in the environment it is deployed in.

The interaction loop feedback is provided mainly by graphical means from a bright and readable enough display solution as the usefulness of audio is subject to environmental variables. The current generation of AR HMDs was found to be suitable in less environmentally complex and more safety-controlled industrial settings in training and industrial plants. There, users can perform tasks efficiently with traditional interaction methods, with the option to introduce users with more advanced input techniques. It must be noted that power requirements are extensive for always-on display, connectivity needs of the device and input devices, and can make any wearable solution to be unergonomic. Monocular smart glasses were found to be an unviable option to serve the current interaction needs of maintenance technicians performing maintenance in the field, better served by extending the input technique capabilities of smartphones placed at a suitable place for visibility.

Currently, a maintenance technician rarely requires explicit interaction outside of sudden information need or need to report by vocalization along with maintenance. Subsection 5.1.1 described how implicit and peripheral interaction allows users to stay focused on their main task. An implicit interaction system should be supported in the form of technician’s activity recognition. It could allow for matching maintenance actions against maintenance instructions for task duration analysis and assisting the worker in remedying a missed maintenance step. Additionally, the same means could be used in support of maintenance reporting or tacit knowledge capture. It needs to be judged how, when, or if implicit interaction could create feedback during field operations and presented to the worker. Examples of technology that required hand activity tracking and recognizing are given in subsection 5.2.2 and gaze as more implicit interaction in subsection 5.2.1. Peripheral interaction requiring only short inputs like flipping a maintenance instructions page on a display device could be achieved through hands-free interaction techniques like facial and tongue gestures and micro-interactions. While the hand and fingers of a technician are already grasping an object, micro-interactions, with available fingers, could allow them to execute a secondary task alongside the primary task with low attention required [Wolf et al., 2011]. These interaction techniques would need to be sensed in a manner that would not interfere with the primary task industrial worker is engaging on,

like small thumb-to-finger flicks or those of available fingers. Other options are to use slightly larger hand movements, like sensing the hand as a joystick-like controller or utilizing confined rapid kinetic movements of the arm.

Advanced AR head-worn devices such as HoloLens 2 support a more comprehensive range of user interfaces due to spatial mapping and input techniques due to eye-tracking optically recognized mid-air gestures. Nevertheless, significant value is left to be unlocked in more simplistic XR solutions that allow more mobility. For now, the value of intuitive and straightforward interactions combined with a similarly minded user interface behind a streamlined user experience have been overlooked by the smart glasses manufacturers. These are essential for the technology to reach familiarity with consumers, which will have a carryover to competency and user acceptance of wearable technology in a professional setting. Thus, it is essential to embrace human-centered design already in the initial design stages of industrial application development. Simultaneously, lowering unit costs with device features such as improved readability of displays, better battery life, lower weight, capability for spatial mapping, and better user feedback through UI, including haptics and sound as optional elements, play an essential role in maximizing value for all stakeholders.

5.4 Limitations and further work

Before the interviews, the review of nonscientific articles might have had a disproportionate effect of priming the researcher's mind to sampling experts who had experience with HMDs over those who had worked with handheld variations only. Due to the small sample size of six participants and the Finnish cultural background of participants, the findings of this study might not be widely generalizable. Due to limited access to the industrial organizations, the sampling did not expand to people in field work roles to establish a completely unbiased perspective of the issues. The chosen research target of designers and developers of larger industrial companies is limited scope impacts the reliability of the data collected. Nevertheless, the average participant represents well the ideal interviewee in that their responsibilities in their respected organization were very similar. Additionally, the close alignment of the findings with the industrial AR implementation success factors of Masood and Egger [2019] seems to support the notion that the interviewee sampling was successful.

The researcher only shared a brief and broad opening statement about the research topic in the interview invite e-mail. The interviewees came in “cold” to the interview as it was chosen not to disclose the research questions to the interviewees in advance. The researcher’s lack of experience in conducting interviews made it difficult to draft the interview questions not to be too leading or too cryptic for perceptions to emerge implicitly. Additionally, the experts kept current research and development efforts out of the interviews due to non-disclosure agreements, leading to missing potentially valuable examples

from being shared in full detail. The interview questions were not iterated between sessions after the first pilot interview, as it was reasoned that the same questions should be asked from everyone interviewed. After reading more on the grounded theory, it was realized that this practice might have had a constrained impact on the iteration of data collection.

In retrospect, one thing that could be changed from the interviews is the use of the video conference meeting. Compared to a natural room setting, interviewing via voice chat was more detrimental than beneficial to the interview as video conferencing etiquette was at the time of interviews unfamiliar. Most participants did keep their camera on, but one chose not to, while the researcher's 10-year-old laptop's camera was kept on, making non-verbal cues tricky to analyze for both parties. The researcher tried to be an active listener by nodding and encouraging participants to talk with an occasional verbal confirmation. However, the efforts to reaffirm the participant while they were talking was sometimes perceived as an attempt to interject, leading to shorter answers or broken up thoughts hindering the interview progress. The research could not be supplement with an observation of workplaces or practices due to the COVID-19 pandemic. At the time, the idea of traveling to the cities in person to conduct the interviews while wearing a mask and keeping an appropriate distance was not known to be an option at the time.

Questions regarding the rigor of the interview data analysis may be asked. A quantitative survey portion, producing a matrix of opinions or attitudes on a Likert scale, could have been an enlightening addition to the collected material, but it was not planned upon as the last two meetings were arranged after four participants had already been interviewed. According to grounded theory methodology, coding of data was done, but quantified measure to coding was not implemented as the analysis took so long. Succeeding research may be impaired because it might not be able to build efficiently on the codes produced in this investigation. After collected interview material had been analyzed, the brain dump notes required to support the formulation of grounded theory were disorganized. This was partly due to time between research phases, making the tacit knowledge notes needed to be rediscovered later, bogging down the completion of the thesis as the research felt incomplete well into the process.

Beyond only a few sources, it was challenging to discover related research from academic databases dealing with interaction techniques in industrial field work. It might be that the nature of this subject is such that limiting public access to information by private companies looking to benefit from the market is intended to prevent competitors from gaining a commercial advantage. However, this makes it difficult for well-intentioned people to become familiar with the subject.

Regardless of the limitations, it is important to study this subject matter further. The timeliness of this dissertation makes it very likely that we will see several studies with

closely related subjects in the future. It can be assumed that this thesis works as a good stepping stone for these future studies.

As of now, the findings remain untested, and the reach in providing help to the industry is unknown. An investigation is needed to validate if the design efforts with the presented theory can create good interaction design for practical industrial work. Results from usability studies of a prototype would indicate if the theory holds the potential for future design efforts targeting industrial environments.

More information on the mental models of users engaged with industrial work is required before and after new technologies are introduced to the average workday. Further research could provide a more nuanced understanding of the industrial use of XR novel interaction technologies, especially from the industrial field workers' perspective. These studies could target the robust touchless interaction techniques for smartphones to be used in their work, like testing Google Pixel 4’s hover semaphoric-stroke mid-air gestures. If and what kind of touch-sensitive textile interfaces could be used in challenging industrial environments? Also, could selection and manipulation with a wrist mouse pointing and symbolic gestures be usable during industrial maintenance work?

It was found that end-users have technology acceptance-related attitudes inhibiting them from using smart-glasses technology due to its high unit cost and questionable appearance of durability. A more focused study could be conducted to determine if these attitudes hinder the adoption of smart-glasses form factor to everyday field work. Additionally, more in-depth studies on comparing binocular and monocular devices on the same industrial tasks could yield more information for the future direction of the display device use and their respective readability.

Lastly, studies should consider multimodal fusion in an industrial context by tracking the user's body with wearable devices during industrial field work. There are many scientific challenges related to the increasing amount of data that can be captured from the user during activity recognition. Intertwining these signal sources into an explicit or implicit multimodal fusion input can be of great significance for XR interfaces and wearable computing. One of the technologies allowing maximization of mobility for multimodal fusion could be to use a purselike unit of computingstick and battery.

6 Summary

The sheer number of users working in maintenance or other related hands-on-the-job work in the industrial sector across the globe is massive, and so are the impacts of the benefits that XR technologies can bring to their daily working lives. The purpose of this thesis was to identify interaction techniques best suitable for industrial environments and the aspects of XR solutions aiding easy adoption to field work. While performing the theoretical review of the subject, the author noted that few studies dealt directly with the implications industrial field work poses to the design of interaction techniques on XR platforms.

Therefore, this thesis took a qualitative grounded theory approach, collecting research material with semi-structured interviews and a supplementary literature review. Beyond interaction techniques of XR devices, the interviews focused on XR experts' perceptions of previous projects and the needs and requirements of the industrial professionals they design solutions for. The target group was selected to keep the industrial company business objectives in the mix to balance the moonshot interaction techniques often presented in academic literature rather than developed by in-house designers or hired consultancies for industrial companies. Yet, the supplementary literature review conducted opens the realm of possibilities and future directions through the visionary emerging input technologies.

The results of the study answer the research questions as follows. XR solutions are implemented to perform mandatory reporting or information lookup. Analysis of the findings supports the theory that, in industrial environments, robust interaction with a readable display and eyes-free secondary feedback should happen with usable and safe operation, preferably hands-free or alternatively touchless freehand input method. Currently, only by extending the input capabilities of smartphones that industrial field workers carry with them can this be achieved with a large workforce. The smartphone should be attached to a visible location during maintenance, and interaction with it should happen via several robust input methods so that the user might choose which one they are comfortable with in the situation. However, it is possible to use smaller numbers of more immersive and expensive solutions in industrial plant operations or training environments. Feasible industrial environment XR solution and interaction techniques are dependent on the worker's use case and the business case of the industrial company because the robust solutions implemented need to create value on use to the individual and be economical to deploy in quantity needed.

All interaction techniques are judged by safety and reliability alone in industrial environments; just because the designer is capable, not every solution should be implemented everywhere. The success of XR technologies is related to their ability to exist in parallel to the real world and in relation to it. Any interaction techniques implemented on

the job must act as enablers of focus, reducing friction while delivering value that the user regards to be worth the slight unease of using new, sometimes wearable, technology. Interfaces should respect the priorities of the user by staying out of their way, except when needed, and efficiently help the user, who prioritizes getting through the current job to the next one and at the end of the day going home, without compromising on safety - the number one priority of every industrial company.

It is a sincere hope of the researcher that this research effort will be a catalyst for deeper investigation into suitable industrial XR interaction techniques and helps to ensure, along with traditional interaction heuristics, a more standard quality of industrial XR applications within the industry. One interviewee described that their job's best moments and gratifying feats come from finding a solution to a big problem. Everything begins with wanting to make the working conditions of people better. Designers' should identify what kind of experiences we humans want on XR and reduce friction points to something people actually want to use. It means finding where this tool can fit – the right job for the tool, instead of looking at use cases and targeting it where it "kinda works". If successful, in these jobs, XR solutions will massively reduce costs and improve efficiency. XR is the future of computing, not necessarily all computing but a specific type of computing, and it is going to make all of our lives so much better in the future.

References

Ahn, S. & Lee, G. (2019). Gaze-assisted typing for smart glasses. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology,* (pp. 857–869).

Aigner, R., Pointner, A., Preindl, T., Parzer, P., & Haller, M. (2020). Embroidered resistive pressure sensors: A novel approach for textile interfaces. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* (pp. 1–13).

Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., Zhao, S., & Koh, Jeffrey Tzu Kwan Valino (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for HCI. Retrieved from: *https://www.microsoft.com/en-us/research/publication/understanding-mid-air-hand-gestures-a-study-of-human-preferences-in-usage-of-gesture-types-for-hci/.*

Alallah, F., Neshati, A., Sakamoto, Y., Hasan, K., Lank, E., Bunt, A., & Irani, P. (2018). Performer vs. observer: Whose comfort level should we consider when examining the social acceptability of input modalities for head-worn display? In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology,* (pp. 1–9).

Amma, C., Georgi, M., Lenz, T., & Winnen, F. (2016). Kinemic wave: A mobile free-hand gesture and text-entry system. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems,* (pp. 3639–3642).

Aromaa, S., Aaltonen, I., Kaasinen, E., Elo, J., & Parkkinen, I. (2016). Use of wearable and augmented reality technologies in industrial maintenance work. In *Proceedings of the 20th International Academic Mindtrek Conference,* (pp. 235–242).

Aromaa, S., Väätänen, A., Kaasinen, E., Uimonen, M., & Siltanen, S. (2018). Human factors and ergonomics evaluation of a tablet based augmented reality system in maintenance work. In *Proceedings of the 22nd International Academic Mindtrek Conference,* (pp. 118–125).

Azuma, R.T. (1997). A survey of augmented reality. *Presence: Teleoperators & Virtual Environments,* 6(4), pp. 355–385.

Bâce, M., Leppänen, T., de Gomez, D.G., & Gomez, A.R. (2016). ubiGaze: Ubiquitous augmented reality messaging using gaze gestures. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications,* (pp. 1–5).

Bernard, J.A., Millman, Z.,B., & Mittal, V.A. (2015). Beat and metaphoric gestures are differentially associated with regional cerebellar and cortical volumes. *Human Brain Mapping,* 36(10), pp. 4016–4030.

Bimber, O. & Raskar, R. (2006). Modern approaches to augmented reality. In *ACM SIGGRAPH 2006 Courses* (pp. 1–es).

Boldu, R., Dancu, A., Matthies, D.J.C., Cascón, P.G., Ransir, S., & Nanayakkara, S. (2018). Thumb-in-motion: Evaluating thumb-to-ring microgestures for athletic activity. In *Proceedings of the Symposium on Spatial User Interaction,* (pp. 150–157).

Bowman, D.A. & Hodges, L.F. (1997). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics* (pp. 35–ff.).

Bowman, D.A., Johnson, D.B., & Hodges, L.F. (1999). Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the ACM symposium on Virtual reality software and technology* (pp. 26–33).

Brice, D., Rafferty, K., & McLoone, S. (Aug 31, 2020). AugmenTech: The usability evaluation of an AR system for maintenance in industry. In *International Conference on Augmented Reality, Virtual Reality, and Computer Graphics,* (pp. 284–303).

Buddhika, T., Zhang, H., Chan, S.W.T., Dissanayake, V., Nanayakkara, S., & Zimmermann, R. (2019). fSense: Unlocking the dimension of force for gestural interactions using smartwatch PPG sensor. In *Proceedings of the 10th Augmented Human International Conference 2019,* (pp. 1–5).

Chen, W., Yu, C., Tu, C., Lyu, Z., Tang, J., Ou, S., Fu, Y., & Xue, Z. (2020). A survey on hand pose estimation with wearable sensors and computer-vision-based methods. *Sensors,* 20(4), 1074.

Chen, Y., Su, X., Tian, F., Huang, J., Zhang, X.(., Dai, G., & Wang, H. (2016). Pactolus: A method for mid-air gesture segmentation within EMG. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems,* (pp. 1760–1765).

Corbin, J. & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology,* 13(1), 3.

Cutting, J.E. & Vishton, P.M. (1995). Chapter 3 - perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth*. In: William Epstein and Sheena Rogers (eds.), *Perception of Space and Motion*. Academic Press, pp. 69–117.

del Amo, I.F., Galeotti, E., Palmarini, R., Dini, G., Erkoyuncu, J., & Roy, R. (2018). An innovative user-centred support tool for augmented reality maintenance systems design: A preliminary study. *Procedia CIRP,* 70, pp. 362–367.

Fuentes, C.T. & Bastian, A.J. (2009). Where is your arm? variations in proprioception across space and tasks. *Journal of Neurophysiology,* 103(1), pp. 164–171.

Gallego Cascón, P., Matthies, D.J.C., Muthukumarana, S., & Nanayakkara, S. (2019). ChewIt. an intraoral interface for discreet interactions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* (pp. 1–13).

Gao, Y., Jin, Y., Li, J., Choi, S., & Jin, Z. (2020). EchoWhisper: Exploring an acoustic-based silent speech interface for smartphone users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* 4(3), 80:1-80:27.

Glaser & Strauss, A. L. (1967). *The Discovery of Grounded Theory.* Aldine.

Glauser, O., Wu, S., Panozzo, D., Hilliges, O., & Sorkine-Hornung, O. (2019). Interactive hand pose estimation using a stretch-sensing soft glove. *ACM Transactions on Graphics,* 38(4), 41:1-41:15.

Gong, J., Yang, X., & Irani, P. (2016). WristWhirl: One-handed continuous smartwatch input using wrist gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology,* (pp. 861–872).

Gong, J., Zhang, Y., Zhou, X., & Yang, X. (2017). Pyro: Thumb-tip gesture recognition using pyroelectric infrared sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology,* (pp. 553–563).

Google. (2020a). (Don't) Hold the Phone: New Features Coming to Pixel 4. September 10, 2020, from https://www.blog.google/products/pixel/new-features-pixel4/

Google. (2020b). Jacquard by Google - Technology Retrieved September 10, 2020, from https://atap.google.com/jacquard/technology

Goudswaard, M., Abraham, A., Goveia da Rocha, B., Andersen, K., & Liang, R. (2020). FabriClick: Interweaving pushbuttons into fabrics using 3D printing and digital embroidery. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference,* (pp. 379–393).

Groenewald, C., Anslow, C., Islam, J., Rooney, C., Passmore, P.J., & Wong, B.L.W. (2016). Understanding 3D mid-air hand gestures with interactive surfaces and displays: A systematic literature review. In *Proceedings of the 30th International BCS Human Computer Interaction Conference* (pp. 1–13).

Gustafson, S. (2012). Imaginary interfaces: Touchscreen-like interaction without the screen. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems,* (pp. 927–930).

Haque, F., Nancel, M., & Vogel, D. (2015). Myopoint: Pointing and clicking using forearm mounted electromyography and inertial motion sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,* (pp. 3653–3656).

Heikkilä, P., Honka, A., Mach, S., Schmalfuß, F., Kaasinen, E., & Väänänen, K. (2018). Quantified factory worker - expert evaluation and ethical considerations of wearable self-tracking devices. In *Proceedings of the 22nd International Academic Mindtrek Conference,* (pp. 202–211).

Higuch, K., Yonetani, R., & Sato, Y. (2016). Can eye help you? effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings*

of the 2016 CHI Conference on Human Factors in Computing Systems, (pp. 5180–5190).

Hincapié-Ramos, J.D., Guo, X., & Irani, P. (2014). The consumed endurance workbench: A tool to assess arm fatigue during mid-air interactions. In *Proceedings of the 2014 Companion Publication on Designing Interactive Systems,* (pp. 109–112).

Hoda, R., Noble, J., & Marshall, S. (2011). Grounded theory for geeks. In *Proceedings of the 18th Conference on Pattern Languages of Programs,* (pp. 1–17).

Hummels, C. & Stappers, P.J. (1998). Meaningful gestures for human computer interaction: Beyond hand postures. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition,* (pp. 591–596).

Istance, H., Bates, R., Hyrskykari, A., & Vickers, S. (2008). Snap clutch, a moded approach to solving the midas touch problem. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications,* (pp. 221–228).

Jones, E., Alexander, J., Andreou, A., Irani, P., & Subramanian, S. (2010). GesText: Accelerometer-based gestural text-entry systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* (pp. 2173–2182).

Kaasinen, E., Roto, V., Hakulinen, J., Heimonen, T., Jokinen, J.P.P., Karvonen, H., Keskinen, T., Koskinen, H., Lu, Y., Saariluoma, P., . . . Turunen, M. (2015). Defining user experience goals to guide the design of industrial systems. *Behaviour & Information Technology,* 34(10), pp. 976–991.

Kaasinen, E., Aromaa, S., Väätänen, A., Mäkelä, V., Hakulinen, J., Hella, J., Elo, J., Siltanen, S., . . . Turunen, M. (2018). Mobile Service Technician 4.0: Knowledge-Sharing Solutions for Industrial Field Maintenance. Retrieved from: https://trepo.tuni.fi/bitstream/handle/10024/116275/mobile_service_technician_2018.pdf

Khalaf, A.S., Alharthi, S.A., Hamilton, B., Dolgov, I., Tran, S., & Toups, Z.O. (2020). A framework of input devices to support designing composite wearable computers. In: Masaaki Kurosu (eds.), *Human-computer interaction. multimodal and natural interaction. HCII 2020.* Springer International Publishing, pp. 401–427.

Kim, K., Kim, H., & Kim, H. (2017). Image-based construction hazard avoidance system using augmented reality in wearable device. *Automation in Construction,* 83, pp. 390–403.

Kinemic. 2021. Handsfree control with the Kinemic Band. Retrieved January 10, 2021, from *https://kinemic.com/en/kinemic-band/*

Kollee, B., Kratz, S., & Dunnigan, A. (2014). Exploring gestural interaction in smart spaces using head mounted devices with ego-centric sensing. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction,* (pp. 40–49).

Kytö, M., Ens, B., Piumsomboon, T., Lee, G.A., & Billinghurst, M. (2018). Pinpointing: Precise head- and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* (pp.81:1-81:14).

Laput, G. & Harrison, C. (2019). Sensing fine-grained hand activity with smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* (pp. 1–13).

Lee, L. & Hui, P. (2018). Interaction methods for smart glasses: A survey. *IEEE Access,* 6, pp. 28712-28732.

Lee, L.H., Braud, T., Bijarbooneh, F.H., & Hui, P. (2020). UbiPoint: Towards non-intrusive mid-air interaction for hardware constrained smart glasses. In *Proceedings of the 11th ACM Multimedia Systems Conference* (pp. 190–201).

Lee, S., Min, C., Montanari, A., Mathur, A., Chang, Y., Song, J., & Kawsar, F. (2019). Automatic smile and frown recognition with kinetic earables. In *Proceedings of the 10th Augmented Human International Conference 2019* (pp. 1–4).

Li, S., Ashok, A., Zhang, Y., Xu, C., Lindqvist, J., & Gruteser, M. (2016). Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom),* (pp. 1–9).

Li, Z., Annett, M., Hinckley, K., Singh, K., & Wigdor, D. (2019). Holodoc: Enabling mixed reality workspaces that harness physical and digital content. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* (pp. 1–14).

Lin, G., Haynes, M., Srinivas, S., Kotipalli, P., & Starner, T. (2021). Towards finding the optimum position in the visual field for a head worn display used for task guidance with non-registered graphics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* 5(1), 22:1-22:26.

Lu, Y., Huang, B., Yu, C., Liu, G., & Shi, Y. (2020). Designing and evaluating hand-to-hand gestures with dual commodity wrist-worn devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* 4(1), 20:1-20:27.

Luciani, A. (2007). Ergotic / epistemic/ semiotic action-perception loops. In: Anonymous (eds.), *Enaction and enactive interfaces : A handbook of terms.* Enactive Systems Books, pp. 94–96.

Mann, S., Furness, T., Yuan, Y., Iorio, J., & Wang, Z. (2018). All reality: Virtual, augmented, mixed (X), mediated (X,Y), and multimediated reality. *arXiv Preprint arXiv:1804.08386,*

Masood, T. & Egger, J. (2019). Augmented reality in support of industry 4.0—Implementation challenges and success factors. *Robotics and Computer Integrated Manufacturing,* 58, pp. 181-195.

Matthies, D.J.C., Urban, B., Wolf, K., & Schmidt, A. (2019). Reflexive interaction: Extending the concept of peripheral interaction. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction,* (pp. 266–278).

Microsoft. (2019a). Immersive headset hardware details—Mixed Reality. Retrieved October 10, 2019, from https://docs.microsoft.com/en-us/windows/mixed-reality/discover/immersive-headset-hardware-details

Microsoft. (2019b). Instinctual interactions - Mixed Reality. Retrieved October 10, 2019, from https://docs.microsoft.com/en-us/windows/mixed-reality/design/interaction-fundamentals

Microsoft. (2019c). Direct manipulation with hands—Mixed Reality. Retrieved October 10, 2019, from https://docs.microsoft.com/en-us/windows/mixed-reality/design/direct-manipulation#3d-object-manipulation

Microsoft. (2020a). Voice input—Mixed Reality. Retrieved December 7, 2020, from https://docs.microsoft.com/en-us/windows/mixed-reality/design/voice-input

Microsoft. (2020b). Supported languages for HoloLens 2 | Microsoft Docs. Retrieved December 7, 2020, from https://docs.microsoft.com/en-us/hololens/hololens2-language-support

Microsoft. (2020c). Gaze and commit - Mixed Reality. Retrieved December 15, 2020, from https://docs.microsoft.com/en-us/windows/mixed-reality/design/gaze-and-commit.

Miles & Huberman, A. M. (1994). *Qualitative Data Analysis.* Sage Publ.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. In *Proceedings of SPIE, 2351,*(1), (pp. 282–292).

Mine, M.R. (1995). Virtual environment interaction techniques. *UNC Chapel Hill CS Dept,*

Moeslund, T.B. & Nørgaard, L. (2003). A brief overview of hand gestures used in wearable human computer interfaces. *Computer Vision and Media Technology Lab., Aalborg University, DK, Tech.Rep,*

Park, K., Kim, S., Yoon, Y., Kim, T., & Lee, G. (2020). DeepFisheye: Near-surface multi-finger tracking technology using fisheye camera. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology,* (pp. 1132–1146).

Poupyrev, I., Gong, N., Fukuhara, S., Karagozler, M.E., Schwesig, C., & Robinson, K.E. (2016). Project jacquard: Interactive digital textiles at scale. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* (pp. 4216–4227).

Qian, J., Ma, J., Li, X., Attal, B., Lai, H., Tompkin, J., Hughes, J.F., & Huang, J. (2019). Portal-ble: Intuitive free-hand manipulation in unbounded smartphone-based augmented reality. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology,* (pp. 133–145).

Rantala, J., Majaranta, P., Kangas, J., Isokoski, P., Akkil, D., Špakov, O., & Raisamo, R. (2020). Gaze interaction with vibrotactile feedback: Review and design guidelines. *Human–Computer Interaction,* 35(1), pp. 1-39.

Realwear. (2020). Rugged Android Tablet RealWear HMT 1. Retrieved January 13, 2020, from https://www.realwear.com/products/hmt-1/

Rise & Alsos, O. A. (2020). *The Potential of Gesture-Based Interaction.* Springer.

Rzayev, R., Woźniak, P.W., Dingler, T., & Henze, N. (2018). Reading on smart glasses: The effect of text position, presentation type and walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* (pp. 1–9).

Saad, M., Bleakley, C.J., Nigram, V., & Kettle, P. (2018). Ultrasonic hand gesture recognition for mobile devices. *Journal on Multimodal User Interfaces,* 12(1), pp. 31-39.

Schneegass, S. & Voit, A. (2016). GestureSleeve: Using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers,* (pp. 108–115).

Sengupta, K., Bhattarai, S., Sarcar, S., MacKenzie, I.S., & Staab, S. (2020). Leveraging error correction in voice-based text entry by talk-and-gaze. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* (pp. 1–11).

Shanhe Yi, Zhengrui Qin, Novak, E., Yafeng Yin, & Qun Li. (2016). GlassGesture: Exploring head gesture interface of smart glasses. In *IEEE INFOCOM 2016 - the 35th Annual IEEE International Conference on Computer Communications,* (pp. 1–9).

Shibata, T., Kim, J., Hoffman, D.M., & Banks, M.S. (2011). Visual discomfort with stereo displays: Effects of viewing distance and direction of vergence-accommodation conflict. In *Proceedings of SPIE, 7863,*(1), (pp. 78630P–9).

Siltanen, S. & Heinonen, H. (2020). Scalable and responsive information for industrial maintenance work: Developing XR support on smart glasses for maintenance technicians. In *Proceedings of the 23rd International Conference on Academic Mindtrek,* (pp. 100–109).

Silva Machado, E.M., Carrillo, I., Collado, M., & Chen, L. (2019). Visual attention-based object detection in cluttered environments. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation,* (pp. 133–139).

Slambekova, D., Bailey, R., & Geigel, J. (2012). Gaze and gesture based object manipulation in virtual worlds. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology,* (pp. 203–204).

Speechly. (2020). Speechly | Speech recognition in VR/AR applications. Retrieved September 10, 2020, from https://www.speechly.com/blog/zoan-voice-technology-in-vr-solutions/

Suzuki, Y., Sekimori, K., Yamato, Y., Yamasaki, Y., Shizuki, B., & Takahashi, S. (2020). A mouth gesture interface featuring a mutual-capacitance sensor embedded in a surgical mask. In: Anonymous (eds.), *Human-computer interaction. multimodal and natural interaction.* Springer International Publishing, pp. 154-165.

Taniguchi, K., Kondo, H., Kurosawa, M., & Nishikawa, A. (2018). Earable TEMPO: A novel, hands-free input device that uses the movement of the tongue measured with a wearable ear sensor. *Sensors (Basel, Switzerland),* 18(3), 733.

Tung, Y., Hsu, C., Wang, H., Chyou, S., Lin, J., Wu, P., Valstar, A., & Chen, M.Y. (2015). User-defined game input for smart glasses in public space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,* (pp. 3327–3336).

Unity. (2020). Mixed Reality (AKA 'MR'). Retrieved January 12, 2020, from https://unity3d.com/what-is-xr-glossary

Varjo. (2020). XR-1 – Varjo.com. Retrieved March 23, 2020, from https://varjo.com/products/xr-1/.

Vuzix. (2020). M4000 Augmented Reality (AR) Smart Glasses. Retrieved January 12, 2020, from https://www.vuzix.com/products/m4000-smart-glasses

Wang, S., Song, J., Lien, J., Poupyrev, I., & Hilliges, O. (2016). Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology,* (pp. 851–860).

Wikipedia. (2019). Interaction technique. Retrieved October 10, 2019, from https://en.wikipedia.org/w/index.php?title=Interaction_technique&oldid=919269715

Wikipedia. (2021a). Pepper's Ghost. Retrieved May 1, 2021, from https://en.wikipedia.org/w/index.php?title=Pepper%27s_ghost&oldid=1015042623.

Wikipedia. (2021b). Computational semiotics. Retrieved May 8, 2021, from https://en.wikipedia.org/w/index.php?title=Computational_semiotics&oldid=987275558

Wolf, K., Naumann, A., Rohs, M., & Müller, J. (2011). A taxonomy of microinteractions: Defining microgestures based on ergonomic and scenario-dependent requirements. In: Anonymous (eds.), *Human-computer interaction – INTERACT 2011.* Springer Berlin Heidelberg, pp. 559-575.

Wu, T., Qi, S., Chen, J., Shang, M., Gong, J., Seyed, T., & Yang, X. (2020). Fabriccio: Touchless gestural input on interactive fabrics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* (pp. 1–14).

Yeo, H., Lee, J., Bianchi, A., Samboy, A., Koike, H., Woo, W., & Quigley, A. (2020). WristLens: Enabling single-handed surface gesture interaction for wrist-worn devices using optical motion sensor. In *Proceedings of the Augmented Humans International Conference,* (pp. 1–8).

Yi, B., Cao, X., Fjeld, M., & Zhao, S. (2012). Exploring user motivations for eyes-free interaction on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* (pp. 2789–2792).

Yu, C., Wei, X., Vachher, S., Qin, Y., Liang, C., Weng, Y., Gu, Y., & Shi, Y. (2019). HandSee: Enabling full hand interaction on smartphone with front camera-based stereo vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* (pp. 1–13).

Zhang, C., Xue, Q., Waghmare, A., Meng, R., Jain, S., Han, Y., Li, X., Cunefare, K., Ploetz, T., Starner, T., . . . Abowd, G.D. (2018). FingerPing: Recognizing fine-grained hand poses using active acoustic on-body sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* (pp. 1–10).

Zheng, X.S., Foucault, C., Matos da Silva, P., Dasari, S., Yang, T., & Goose, S. (2015). Eye-wearable technology for machine maintenance: Effects of display position and hands-free operation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,* (pp. 2125–2134).

Appendix 1

Haastatteluskripti

Industrial Interaction Design Principles for XR

Tarkistuslista:

- Esittele itsesi ja johdanto
- Tallennuslupa
- ”Kysyttävää ennen tallennuksen aloittamista?”
- Aloita tallennus
- Varmista tallennuslupa tallenteelle -> Aloita haastattelu
- Noudata kysymysrunkoa parhaasi mukaan.
- Kiitä haastateltavaa haastattelusta
- Kirjaa vaikutelmat muistoon

Johdanto

Hyvää päivää, olen Harri Halonen Tampereen yliopiston opiskelija Ihmisen ja teknologian vuorovaikutuksen maisteriohjelmassa. Tämä haastattelu on osa Tampereen yliopiston pro gradu tutkielmaani. Tapaamisen kesto on noin tunti ja kielenä voi toimia, joko suomi tai englanti.

Kiitos suostumuksestasi osallistua haastatteluun ja jakaa kokemuksiasi pro gradu -tutkielmaani varten.

Tämän haastattelun päätavoitteena on se, että haluan tietää enemmän kokemuksista teollisinten toimivien XR ratkaisujen kehitystyöstä sekä millaiset ihmisen ja tietokoneen väliset vuorovaikutustekniikat ja käytännöt sopivat parhaiten teollisuusympäristöön, erityisesti kenttätöihin. Näin ollen haastattelun aikana kysyn kysymyksiä, jotka liittyvät näihin havaintoihin.

Ennen kuin aloitamme, haluaisin tietää, onko sinulla herännyt mitään kysymyksiä. Seuraavaksi haluaisin kysyä, annatko luvan tämän haastattelun tallentamiseen.

> Haastattelu tallennetaan videolle, jossa mukana äänitetään haastateltavan ja haastattelijan mikrofonien ääni. Mielipiteesi on minulle tärkeä ja sitä käsitellään luottamuksellisesti. Haastattelun tallennetta käytetään vain oman pro gradu tutkielmani tarpeisiin, eikä sitä julkaista osana lopputyössäni. Tallenteet ovat ainoastaan tutkielman tekijän käytössä ja ne tuhotaan opinnäytetyön arvioinnin jälkeen. Anonymiteetin takaamiseksi myös yhteydenotossa käytetyt nimi- ja osoitetiedot hävitetään.
>
> Tutkielmassa ei esiinny tietoja, joiden perusteella haastateltavan voisi tunnistaa. Haastattelun tulokset raportoidaan tavalla, jossa käytän osallistujista koodeja I1, I2 jne. Videotallenteita tai osallistujan henkilötietoja ei luovuteta eteenpäin.
>
> Ilmoitan sinulle selkeästi ennen kuin käynnistän tallennuksen, sekä varmistan tallennusluvan uudelleen välittömästi tallennuksen alettua. Voit halutessasi lopettaa haastattelun missä tahansa vaiheessa.
> Vastaan mielelläni, jos sinulla on jotain kysyttävää.
>
> Suostutko haastattelun tallentamiseen?

Kiitos.

Onko sinulla tässä vaiheessa mitään kysyttävää ennen kuin käynnistän tallennuksen?

Aloitan tallennuksen nyt.

Haastattelu

Kysyn edellisen vielä uudelleen tallennetta varten. Voinko tallentaa haastattelun?
Kiitos. Aloitetaan 😊

Valikoi ja muotoile haastattelukysymykset roolin mukaan! Kuuntele & 5 x Miksi!

Pyydä omasanaisesti kertomaan hieman itsestään sekä roolistaan, jonka jälkeen ohjaa maltilla haastattelu ensimmäiseen varsinaiseen kysymykseen, jos tarve.

Kaikille: Mitä XR projekteja on toteutettu? – Mistä olet ylpein? Miksi?

Ota selvää näistä, jos juttu ei luista:

- Mikä oli liiketoiminnallinen tavoite ja miksi se on merkityksellinen?
- Mitä ovat olleet suurimmat esteet, jotka vaikeuttavat teidän onnistumistanne?
- Minkä yhden asian toivoisit, että olisitte voineet tehdä eri tavalla?

Jos aikaa jää: Mitä ovat olleet riskit, joita otitte? tai tärkein päätös? - Mikä on isoin haaste, joka teitä odottaa tämän johdosta? Kuinka ajattelitte ratkaista sen?

- Industry 4.0 vaikutus organisaationne tai sinun työhösi? -Mitä tapahtuisi, jos ette toimisi trendin perusteella? Mitä teidän täytyisi tehdä tänään pysyäksenne kehityksen mukana? – tiimi, osaaminen ja kulttuuri kolmen vuoden kuluttua?
- Mitä teknologioita tai yrityksiä pidät silmällä? Seuraat somessa? Mikä tekee sinuun vaikutuksen?

Projektistanne

Kaikille: Mitä ratkaisemisen arvoista ongelmaa yritätte ratkaista? Miksi?

- Entä ratkaisunne, kuinka se ratkaisee tämän ongelman? - Mikä erottaa sen muista saman ongelman ratkaisuista?
- Miten olette validoineet konseptin?

Stimuloivia, jos tarve: Mitkä asiat konseptissa vaikuttavat eniten heidän työhönsä? - Mikä motivoi käyttäjiäsi? Mikä on heille todella tärkeää? Mitä käyttäjät tuntevat käyttäessään tuotetta?

- Millaisia näkökohtia on havaittu tärkeäksi – ideat ratkaisevat ongelman tai vähentävät käyttäjän kipukohtia tai parantavat parhaita osia
- Mitkä ovat nykyiset tavoitteesi? – projektin jatkuminen tulevaisuuteen?

Kaikille: Millaiset vuorovaikutustekniikat ja käytännöt sopivat mielestäsi teollisuusympäristöön parhaiten? Miksi?

- Miksi juuri tämä? Kuinka olet validoinut nämä havainnot? Ratkaisun luotettavuus (robustness)
- Onko olemassa joitain yleisesti käytettyjä tekniikoita, jotka eivät sovi hyvin teollisuusympäristöön? väärinkäsitys, jonka olet itse todennut olevan oikein tai väärin?

- *Stimuloivia, jos tarve:* Mitä ovat käyttäjien tarpeet? Mikä on tärkein asia käyttäjän vuorovaikutuksessa?
- Output: AR - VR visual ajatukset (wearable or not)? - Audio ajatukset? (spatiaalinen) immersion merkitys? – Intensiteetti & Realismi
- suora vuorovaikutus - fyysinen ohjain – virtuaalinen
- Input: kosketus (ohjain) – puhe – katse – eleet (erilaiset) - multimodaalisuus

Kaikille: Mitkä näkökohdat tekevät XR-ratkaisuista helpon omaksua kenttätyössä? Miksi?

- Se on mielenkiintoinen ajatus, minkä prosessin läpi olet käynyt tämän päätelmän tekemiseksi?
  Mitkä ovat vähimmäisvaatimukset toimivalle ratkaisulle / tuotteelle (tässä yhteydessä)?
  Mitä onnistunut tuote sisältää? Miksi? Mitä se ei sisällä? Miksi? - Miten estät X: tä tapahtumasta?
  Mitä saavutatte tällä? - Mistä tiedätte, että olette onnistuneet? – onnistumiskriteerit tai mittarit?

- *Stimuloivia, jos tarve:* Turvallisuus – tyypilliset vaarat ja näkymättömät/ennakoimattomat vaarat
- PPE (Personal Protective Equipment) Henkilökohtaisten suojaimien yhteensopivuus
- Ergonomia – Miten ergonomia on huomioitu?
- Käyttäjien huomioiminen – Kuinka käyttäjäkokemusta voidaan räätälöidä? – flow-tila
- Onko sinulla tarvittavat resurssit (taloudellinen, tekniikka, kyky, infrastruktuuri jne.) tavoitteesi saavuttamiseksi?

Lopetus

Kysy lopuksi henkilön taustaa, ammatillisia tottumuksia, vaikutteiden lähteitä - enablers *mallit / järjestelmät, kirjat, ihmiset* tai Mielessä olevat pohdinnat alan tulevaisuudesta?

- Kuinka näet, että lopputyöni voisi auttaa sinua? Mitä tarvitsette minulta, jotta voisin auttaa teitä?
- Mitä kysymystä en kysynyt, mutta minun olisi pitänyt? Tai mitä neuvoja voisit antaa minulle?

Tässä oli viimeinen kysymykseni.
Paljon kiitoksia haastattelusta! 😊 Hyvää työpäivänjatkoa!