KTH ROYAL INSTITUTE
OF TECHNOLOGY

School of Electrical Engineering and
Computer Science

2D or not 2D: Assessing environmental awareness and cognitive load in mixed reality data visualizations

Visualizing gas leak data visualizations for an industrial surveying setting

Julia Dang Linh Schwuchow

Degree Project in ICT Innovation
Second Cycle, 30 Credits
Course Code: DA258X
August 2023

Examiner:
Prof. Dr. Li Haibo, KTH

Supervisors:
Dr. Christopher Peters, KTH
Dawid Ziobro, ABB
Dr. Mariët Theune, University of Twente

Company: ABB
School of Electrical Engineering
and Computer Science
KTH Royale Institute of Technology
Brinellvägen 8
114 28 Stockholm
Sweden

Acknowledgements

I would like to express my deepest gratitude to my KTH university supervisor Prof. Dr. Christopher Peters who helped me in understanding how good research is done and in abstracting what I do in the company to what value it can bring to the science community. I am also deeply indebted to my ABB company supervisor Dawid Ziobro who welcomed me caringly into the team and supported me during the whole master thesis journey, providing me with any support I needed although being occupied with many other things. I could not have undertaken this journey without the amazing ABB research team (Dr. Yemao Man, Emmanuel Brorsson, Petra Bjorndal and Reem Khan) who all provided me with valuable tips and guidance, from clarifying questions about the statistical analysis, reviewing GDPR-related issues to teaching me about a qualitative analysis tool. Many thanks also go to the other master thesis students Thomas Jinton, Eliz Fikret, Elias Carlson and Jens Junkers with whom I could discuss my progress and share my experiences as a thesis student. Additionally, words cannot express my gratitude to Tobias Lindborg who became my ABB technical support pillar by helping me with the integration of the HoloLens 2 into the ABB ecosystem and by troubleshooting any technical issues that arouse. I would like to express my deepest appreciation to the ABB business stakeholders Dane Silva, Julia Gao, Zach Plante and Mathieu St-Jean who took an interest in this research and shaped the final outcome. Many thanks should go to my second university supervisor Dr. Mariët Theune from the University of Twente who supported me in getting approval from the university ethics committee so that I could perform the intended research activities. I would also be remiss in not mentioning my partner Henri Schmidt who provided me with emotional support throughout my journey and helped me with my initial development environment setup. Lastly, I would like to extend my sincere thanks to all the employees of the ABB Corporate Research Center who participated in this research and shared their genuine insights with me.

Abstract

Increasing the environmental awareness of invisible data, such as a city's air pollution or a household's energy consumption, has been an ongoing information visualization challenge that has entered the fields of immersive technologies like augmented (AR) and mixed reality (MR). In order to understand how form and dimensional affect this understanding for MR environments, this research investigated what the difference between statistical, iconographic 2D and physical 3D visualizations is with regards to minimizing cognitive load and fostering environmental awareness.

By following a HCD process, two MR prototypes were developed for the HoloLens 2 which depicted invisible gas leak data. They were evaluated through a survey and usability study with 14 experts and non-experts, using SAGAT, NASA-TLX and other questionnaires.

The results suggest that there is no significant difference between statistical, iconographic 2D and physical 3D visualizations in terms of environmental awareness and cognitive load. However, an analysis of the effect size and qualitative post-interview data imply that the physical 3D visualization led to a higher environmental awareness in all its three dimensions, while an analysis of the cognitive load sub-dimensions indicate that the statistical, iconographic 2D visualization was better suited for lowering cognitive load.

Keywords: *environmental awareness, cognitive load, information visualization, mixed reality, sustainable HCI*

Sammanfattning

Att öka medvetenheten om osynliga data, såsom luftföroreningar i en stad eller hushållens energiförbrukning, har varit en pågående utmaning inom informationsvisualisering som har nått fram till områden inom immersiva teknologier som förstärkt verklighet (AR) och blandad verklighet (MR). För att förstå hur form och dimension påverkar denna förståelse i MR-miljöer, undersökte denna forskning skillnaden mellan statistiska, ikonografiska 2D- och fysiska 3D-visualiseringar när det gäller att minska kognitiv belastning och främja miljömedvetenhet.

Genom att följa en HCD-process utvecklades två MR-prototyper för HoloLens 2 som skildrade osynliga gasläckagedata. De utvärderades genom en enkät- och användbarhetsstudie med 14 experter och icke-experter, där SAGAT, NASA-TLX och andra enkäter användes.

Resultaten antyder att det inte finns någon signifikant skillnad mellan statistiska, ikonografiska 2D- och fysiska 3D-visualiseringar när det gäller miljömedvetenhet och kognitiv belastning. Dock antyder en analys av effektstorleken och kvalitativa postintervjudata att den fysiska 3D-visualiseringen ledde till en högre miljömedvetenhet i alla sina tre dimensioner, medan en analys av underdimensionerna för kognitiv belastning indikerar att den statistiska, ikonografiska 2D-visualiseringen var bättre lämpad för att minska kognitiv belastning.

Keywords: *miljömedvetenhet, kognitiv belastning, informationsvisualisering, mixad verklighet, hållbar MDI*

Contents

List of acronyms

AR	augmented reality
AV	augmented virtuality
CO_2	carbon dioxide
CLT	cognitive load theory
EA	environmental attitudes
EAI	environmental attitudes inventory
HCI	human-computer interaction
IEA	International Energy Agency
MR	mixed reality
NASA-TLX	NASA-Task Load Index
NEP	new ecological paradigm
SA	situation awareness
SAGAT	Situation Awareness Global Assessment Technique
SART	Situational Awareness Rating Technique
SDG	Sustainable Development Goal
SHCI	sustainable human-computer interaction
SID	sustainable interaction design
SPAM	Situation-Present Assessment Method
SWAT	Subjective Workload Assessment Technique
UEQ	User Experience Questionnaire
UCD	user-centered design
VR	virtual reality
GDTA	goal-directed task analysis

Chapter 1

Introduction

1.1 Motivation

According to the International Energy Agency (IEA), reducing methane emissions from the oil and gas industry is among the most cost-effective and impactful measures to combat climate change due to the fact that methane has a much shorter, atmospheric lifespan than carbon dioxide (CO_2) and therefore leads to faster environmental protection [1]. A major source of these emissions are involuntary gas leaks in pipelines which might be caused by corrosion, weld defects or third-party damage [2] and require further prioritization of which leaks should be repaired in which order. To find gas leaks in pipelines, manual detection is performed by human surveyors who search for leakages along pipelines with the help of handheld detection devices [3]. Their work is attributed with a high cognitive load as they have to be aware of different sensual cues that can signal the proximity and gravity of a leakage as well as their direct surroundings which affect their movements and leak assessment. Furthermore, as surveyors greatly contribute in reducing harmful greenhouse emissions, gaining environmental awareness is essential which means that they can understand the environmental situation and the impact of gas leaks on the environment. Yet, although becoming aware of the environmental impact is important for the surveyor's work, a significant challenge lies in the fact that the causes of climate change, i.e. the greenhouse gas emissions, are invisible [4] and thus not perceivable by the human eye.

Visualizing invisible and intangible matter has been attempted by previous research to increase environmental awareness and sustainable behavior. For instance, electricity visualizations were created to better understand one's own energy consumption and to derive motivations for sustainable behavior [5], [6]. Likewise, efforts have been made to visualize a city's intangible air pollution [7], [8] or the invisible wind direction and speed through immersive technologies such as augmented reality (AR) and mixed reality (MR) [9], [10].

To visualize this data, different forms exist. While a majority of visualizations are statistical in the form of line graphs, bar or pie charts, less research has been done on physical or ambient visualizations which are integrated into the physical environment of a user [6]. A few of these attempts included an AR application which visualized air pollution through the number of physical particles in the air [7] (see figure 1.2) and the instantiation of colored particles and arrows in 3D space to indicate the wind direction [9] (see figure 1.2).

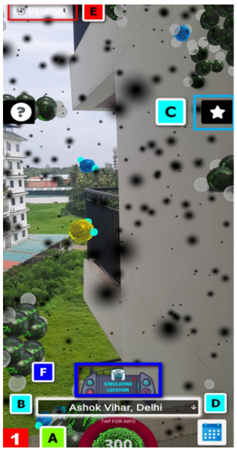
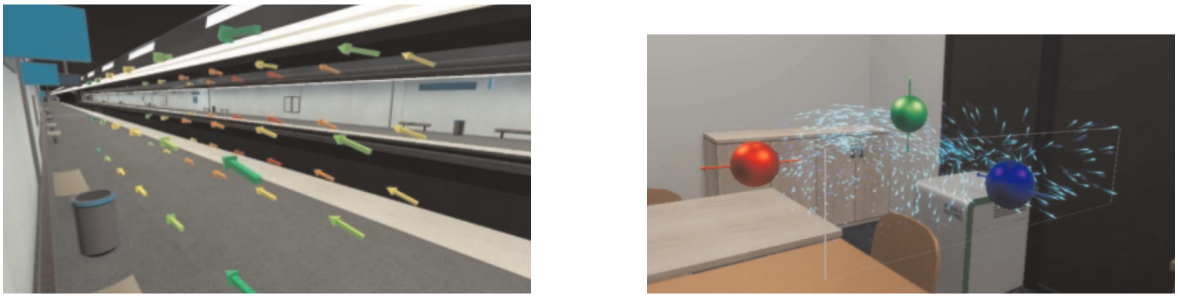


Figure 1.1: AR app which displays physical particles to visualize an area's air pollution (from Mathews et al. [7]).



(a): Arrows in 3D space representing the wind flow.

(b): Partial representation of the wind flow through dynamic particles.

Figure 1.2: VR wind arrow visualizations (from Christmann et al. [9]).

Apart from the form, there is also the dimensionality of a visualization. While different forms or representations of invisible data have been compared in AR or MR [9], little is known on how the dimensionality of a visualization can affect the understanding of it, except that MR and its three-dimensional perspective is better suited and less cognitively demanding for spatial data [11], [12].

Furthermore, while many environmental visualizations focus on the visualization of energy consumption data [6], [13], less is done on other harmful, environmental and natural resources management aspects such as emissions or water as it would be, for instance, required for a smart city management [14]. Last but not least, although previous research compared different visualizations in terms of their effectiveness and understandability to increase the awareness about one's domestic energy consumption [15], no such comparison attempt has yet been made in MR and the concept of environmental awareness, which was introduced by social sciences research [16], has not yet been evaluated in the scientific domain of information visualizations.

Considering these research gaps, the overall aim of this research will be to investigate and understand how invisible and intangible information such as gas leak data should be visualized in MR to increase the environmental awareness of surveyors in an industrial setting while taking into consideration the cognitive load of these visualizations. For this exploration, two forms and dimensions will be compared in MR: A statistical, iconographic 2D and a physical 3D visualization.

So far, most sustainable human-computer interaction (SHCI) research has focused on increasing awareness in individuals or increasingly also of employees in an office context [17]. However, while the individual in the role of the *consumer* can have an essential impact on the environment, so does the individual in the role of an *employee* in an industrial working context where their daily work can have a direct or indirect, small or large effect on environmental sustainability. Increasing their awareness of the current environmental situation and their work's impact might therefore lead to a more thoughtful and safe relationship with the environment.

Apart from a contribution to the SHCI community, this research will also promote the vision of the Industry 5.0 of a human-centered, sustainable and resilient industrial future [18]. *Human-centered* means that the technology should fit the human worker instead of the other way around to ensure their rights for privacy, autonomy and human dignity [18]. The idea of an augmented surveyor, who is a smart and skilled operator supported by technology to perform their tasks efficiently through monitoring and analyzing data [19], has been successfully implemented in other related, industrial contexts where MR was used for inspection tasks by visualizing spatial data and increasing an inspector's performance [20]. *Sustainable* implies that sustainability becomes part of the business model and resources are used in an efficient way [18]. This entails that the surveyor is aware of all factors surrounding gas leaks that are relevant for sustainability and communicates them to business decision-makers who can forward these information to governments so that accurate monitoring of the environment is possible. Finally, a *resilient* industry should be robust towards crises [18], ensuring that super-emitters can be identified and documented efficiently to prevent more significant environmental harm.

To conclude, this research seeks to compare two visualization techniques (statistical, iconographic 2D, physical 3D) in an attempt to find out how which visualization concept is better suited to foster environmental awareness in a cognitively efficient manner for different types of invisible gas leak data.

The outcome of this research will be insights into statistical, iconographic 2D and physical 3D visualizations in MR and how they perform in terms of environmental awareness and cognitive load. Furthermore, on the basis of this analysis, further design recommendations for visualizations in MR and suggestions for future work directions will be provided.

1.2 Research question

This research was done in collaboration with the corporate UX research department from ABB in Västerås, Sweden, which explored AR / MR technology in the context of gas leak surveying. Within this broad area of interest, this research focused on how the gas leak data, as a type of invisible data, should be visualized in MR to foster environmental awareness and limit cognitive load. More concretely, it sought to discover how the dimensionality (2D vs. 3D) and form (statistical, iconographic vs. physical) could affect both concepts. While 3D visualizations were said to be suitable for spatial data [11], [12], little is known on how the dimensionality could affect other types of data such as localized, invisible gas leak data which embody a sustainability issue. On the one hand, a physical representation which relates to the physical world [13] was chosen for the 3D visualization since the 3D space might allow a more realistic representation of the environmental data. On the other hand, statistical and iconographic forms were chosen for the 2D visualizations as statistical representations served as the most common visualization type of invisible energy consumption data [6] and as icons were a common way to effectively communicate qualities or system objects [21].

From this motivation, the following main research question was formulated and divided into two smaller research questions:

Main RQ: *What is the difference between statistical, iconographic 2D and physical 3D visualizations in mixed reality with regards to minimizing cognitive load and fostering environmental awareness?*

RQ1: *What is the environmental awareness of statistical, iconographic 2D and physical 3D visualizations in mixed reality?*

The first research question served the goal of understanding how the dimensionality and form of MR visualizations could affect the environmental awareness of a visualization depicting localized, invisible data. Environmental awareness was chosen as an evaluation criterion as this research aimed at introducing the concept of environmental awareness described by Kollmuss et al. [16] into current information visualization research which currently relies more on knowledge questions to evaluate the awareness level [22]. Environmental awareness was measured in its three dimensions *cognitive*, *affective* and *conative* [16]. While the cognitive understanding of the visualizations was assessed through the Situation Awareness Global Assessment Technique (SAGAT), the affective and conative dimensions were evaluated through adjusted questionnaires based on Fraj & Martinez [23] in a seven-point Likert scale format. The application of SAGAT meant that participants had to answer questions corresponding to the three levels of situation awareness (SA) (perception, understanding, projection) for different leak scenarios while viewing the 2D and 3D visualizations in a HoloLens 2 device or viewing videos of the leak scenarios in a survey. Affective and conative environmental awareness were then evaluated after each shown visualization concept (2D, 3D) through filling in an online questionnaire. A discussion of the RQ1 results can be found in section 4.2.1.

RQ2: *What is the cognitive load of statistical, iconographic 2D and physical 3D visualizations in mixed reality?*

In the second research question, dimensionality and form should be measured in terms of cognitive load which is a common evaluation criterion for immersive data visualizations [11] and for safety-critical domains [24]. To measure cognitive load, the subjective NASA-TLX questionnaire was used in a seven-point Likert scale format which has been commonly employed to evaluate MR systems [25], [26]. It had to be filled in after watching a series of videos depicting gas leak scenarios for each visualization concept (2D, 3D) or physically viewing the leak scenarios through a HoloLens 2 device. The results of RQ2 can be viewed in section 4.2.2.

1.3 Purpose

From the surveyors' perspective, they will benefit from an improved leak surveying tool that can decrease their cognitive load and increase their SA to spot more leaks in a more efficient way and report them more accurately. Furthermore, in the light of the Industry 5.0 which is concerned with a worker's dignity and autonomy [18], exploring how environmental awareness can be increased will help surveyors not only do their work with less cognitive demands but also bring forth the meaning of their work and their impact on decreasing methane emissions in the atmosphere, treating the employee in a more human-centered manner.

Lastly, a few contributions to the science community can be mentioned. First of all, by bringing the concept of environmental awareness from social science research into current SHCI research [16], it might become a new evaluation criteria for environmental information visualizations. Furthermore, the exploration of different dimensions and forms of MR information visualizations and the resulting design recommendations and future work directions can guide future projects of information visualizations in immersive technologies.

1.4 Report organization

The remainder of this report is organized as follows. The report will start in Chapter 2 with a discussion of related work to explain both concepts, present previous MR research and introduce the topic of information visualization with a focus on the dimensionality (2D, 3D) along with the forms used in this research. Then, Chapter 3 will present the methodological approach to answer the main research question together with the activities undertaken in the framework of a user-centered design (UCD) process. The main activities presented in this chapter will be interviews to understand the context of surveyors and the informational requirements, prototyping in Figma and Unity to develop both visualization concepts and an evaluation through a survey and usability study. The results and analysis of the quantitative and qualitative data will then be presented and discussed in Chapter 4. These results will include a thematic analysis of the qualitative data as well as the application of the Wilcoxon-Signed rank test and the calculation of the effect sizes and error rates for the quantitative data. After that, Chapter 4.2.3 will discuss the results in the light of previous related work mentioned in Chapter 2. Last but not least, in Chapter 5, some final conclusions and recommendations for future work will be given.

Chapter 2

Related Work

2.1 Related Work

2.1.1 Sustainable HCI

A new direction in human-computer interaction (HCI) was proposed by Eli Blevis [27] in 2007, who introduced the concept of sustainable interaction design (SID) as ”an act of choosing among or informing choices of future ways of being”, emphasizing the importance of designing for environmental sustainability. SHCI incorporates two focus areas: Sustainable computing and computing for sustainability. The former is concerned with the question of how the environmental footprint of computing itself could be reduced, while the latter investigates computing in a broader sense of how it could serve sustainability in different societal sectors. While the former included research which was more understandable and tractable, the latter has remained rather complex as it involves economic, political, cultural and biophysical dimensions which can exceed the expertise of a computer researcher [28].

This problem was emphasized by Prost & Mattheiss [17] who criticized that SHCI was too much focused on the individual, while ignoring larger social and cultural elements which can hinder an individual's sustainable behavior, leaving the user ”helpless” despite raising environmental awareness. In a similar note, Scuri et al. [29] proposed that more focus should be given to the economic lens of the triple bottom line of sustainability (environmental, social, economic) such as the idea of flora and fauna as essential stakeholders in the production process of new materials.

Previous research [28] also argued that SHCI would have its greatest impact when addressing societal challenges and when leading a process of *disintermediation* which means that the true ecological and human cost of produced goods should be made available through modern information systems instead of being hidden through them. According to the authors, this could be achieved through three key concepts of decentralization, decommercialization and decomplexification. While this research was not meant to decentralize or decommercialize the gas sector, it still tried to unravel and visualize the environmental aspects resulting from gas leaks, decomplexifying the surveying context.

These presented examples showcase how the topic of sustainability has entered HCI practices in the last decade. However, sustainability has not only been increasingly discussed in research, but it has also become a major goal for societal areas such as the industrial working context. According to the European Commission [18], the so-called *Industry 5.0* stands for the "power of the industry to achieve societal goals beyond jobs and growth to become a resilient provider of prosperity, by making production respect the boundaries of our planet and by placing the wellbeing of the industry worker at the centre of the production process". While the Industry 4.0 was more concerned with efficiency and shareholder value, its new iteration focuses on human-centered interaction with technology, resilience towards crises and sustainability which should be better embraced at an organizational level to increase resource efficiency.

Therefore, this research will contribute both to ongoing SHCI research as well as the vision of the Industry 5.0 by exploring environmental awareness in an industrial setting through different information visualizations in MR. Furthermore, it will specifically address the 12th Sustainable Development Goal (SDG) of responsible consumption and production [30].

2.1.2 Environmental Awareness

Environmental awareness is a multidimensional concept with different existing conceptualizations which can be broadly defined as "knowing of the impact of human behavior on the environment". It can be seen as a condition precedent of pro-environmental behavior which is behavior that seeks to minimize the environmental harm done through one's actions, and embodies three dimensions of awareness which are *cognitive*, *affective* and *conative* [16] (see Figure 2.1). These three dimensions correspond to the tripartite classification of mental activities coined in the eighteenth century by the German faculty of psychology [31]. They can be understood as followed.

Cognitive Dimension

The cognitive component relates to knowledge stored in memory along with cognitive processes [32] which, through the environmental lens, can be narrowed down to an understanding of environmental issues, their causes and impact on the natural world. One cognitive construct which corresponds to this dimension is the concept

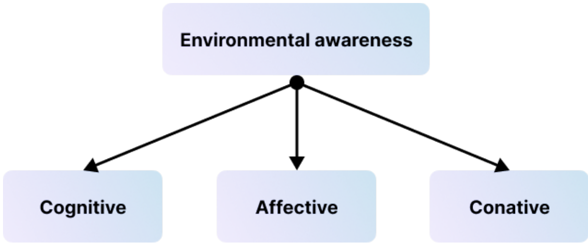


Figure 2.1: Environmental awareness has three dimensions: i) cognitive, ii) affective and iii) conative. Cognition relates to environmental knowledge, affection to emotional feelings towards environmental issues and conation to a willingness to act sustainably (based on Kollmuss & Agyeman [16]).

of SA coined by Mica R. Endsley in the late 1980s. According to Endsley, SA is defined as "the perception of elements of the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" and is constituted of three levels of increasing complexity [33]. Level 1 of SA is the simple perception of static or dynamic cues in the environment through the five senses. Once these cues have been perceived, level 2 of SA is working on synthesizing all cues based on their meaning in regards to one's goals to fully comprehend the current situation. Finally, level 3 of SA is the projection of the future status. It means that predictions can be made of how the elements in the environment will behave and affect the future [34].

In the context of surveying, the cognitive dimension encompasses the knowledge of all the environmental aspects that are related to gas leaks in pipelines. In this case, the main environmental aspect would be fugitive methane emissions which greatly contribute to climate change [35]. As being environmentally aware does not simply involve the perception of methane emissions but also the insight into what these emissions might mean for the global carbon footprint, all three SA levels are of importance for environmental awareness and should be evaluated in this research.

Cognitive Dimension - Evaluation

One common objective method to measure situation awareness is the SAGAT which was initially developed to test aircraft designs. It entails that pilots fly a mission scenario in a man-in-the-loop simulation until the simulation is halted at a random point in time and the screen goes blank. The pilot then has to answer a series of knowledge questions until they can continue the simulation. This process is repeated several times for different scenarios which rely on important SA information and more secondary SA information [33]. Another objective method is the Situation-Present Assessment Method (SPAM) which differs from SAGAT by not hiding the information during the assessment and using response latency as a measure instead of correct answers. Rather than relying on a memory component, it assumes that sometimes it is enough to know where something is instead of what it is [36]. In comparison to SAGAT, it was however found to have a lower sensitivity [37]. Apart from the objective methods, there is also a subjective measurement called the Situational Awareness Rating Technique (SART) which is a 10-dimensional questionnaire [38] that is, however, deemed a less effective method than SAGAT [39]. The same holds true for task performance metrics such as response time and error counts which assume that behavior occurs in a particular state of SA and do only correlate with it [37]. Newer methods are physiological measurements of eye movements which were used in one study to assess the SA of airline pilots [40]. However, due to the lack of previous evaluations and much scope for interpretation power, direct, objective measures should be preferred over subjective and indirect physiological methods [37].

Based on this recommendation, this research employed SAGAT as a direct, objective measure for SA which was found to be highly sensitive, reliable and predictive of SA across many domains and experimental settings [37].

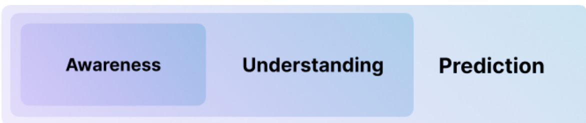


Figure 2.2: Situation awareness entails three levels: i) perception of environmental cues, ii) understanding of the cues and iii) projection of this understanding onto the future (based on Endsley [33]).

Cognitive Dimension - AR / MR

In a systematic literature review, researchers found out that a majority (81%) of the papers which motivated their work within AR by improving SA did not use a specific SA evaluation technique but instead applied performance metrics. Moreover, the majority (78%) did not focus on studying the presentation of information but instead evaluated the applicability of using AR for different situations which led to the conclusion that the way information should be designed for increasing a user's SA, especially for AR headsets, remains an open question [41]. While little is known on how the visualization in AR affects a viewer's SA, multiple studies have shown that AR / MR can increase SA in comparison to traditional technologies. For instance, in one study, SAGAT was used to evaluate wearable augmented reality displays in a marine transportation context which improved the operator's track-keeping performance and SA in comparison to conventional displays [42]. As immersion is important for contexts such as the military or surgery that require high SA, AR / MR can solve the problem of holding another external device (e.g., tablet) which distracts the user from the main task [41]. Moreover, in another study using SAGAT, patient vital signs were visualized in AR and evaluated for clinicians to increase the awareness of important patient alarms which currently only consisted in auditory alarms. The comparison suggested that AR visualizations could lead to faster reaction times, less errors in missing alarms and higher SA. All these examples indicate that AR / MR has a positive effect on SA for critical domains.

To summarize, this research will fill the gap of existing AR / MR research regarding SA by employing SAGAT as a traditional SA evaluation method and exploring how SA can be enhanced through the dimensionality and form of a visualization.

Affective Dimension

The affective component is concerned with emotional reactions, feelings and expectations towards environmental issues [43] and comprises the emotional judgment about one's own impact on the environment [32]. While affect refers to ”emotional responses and feelings engendered by an attitude object”, evaluation or attitude refers to ”thoughts, beliefs, and judgments about an attitude object” [44]. Attitudes that are based on the affective component are more reliable predictors of pro-environmental behavior as they are less complex and therefore less prone to situational factors such as e.g., the price of a product [32].

While all three dimensions of cognition, affect and conation can be viewed independently, there is still some influence that can be found between them. For instance, gaining more knowledge about environmental issues can shape one's feelings and attitudes towards these issues [23], or in the opposite way, an emotional non-investment can be caused by the lack of environmental knowledge [16].

For the surveyor, being environmentally aware in the affective dimension would imply existing feelings towards the topic of sustainability and an awareness of their own sustainability impact through their work.

Affective Dimension - Evaluation

Some methods were mentioned in literature to measure environmental awareness in an affective dimension. One method is the new ecological paradigm (NEP) scale which measures the environmental concern of people or the advocacy of a ”pro-ecological” world view and comprises fifteen statements in a Likert scale format [45]. To measure environmental attitudes (EA), Milfont & Duckitt [46] developed a multi-dimensional inventory called the environmental attitudes inventory (EAI) which consists of twelve scales with ten questions per scale. However, as both questionnaires inquire more about environmental knowledge and attitudes and less about feelings towards the environmental impact of somebody's actions, they were less suitable for this research.

A more promising technique from previous research was found to be the ecological attitudes and knowledge scale by Maloney [43] which was used by Fraj & Martinez' work [23] where it was shown that pro-environmental behavior is determined by emotions towards the environment which can better explain pro-environmental behavior than a verbal commitment of acting in a pro-environmental manner [23].

In this research, a revised version of Fraj & Martinez [23] will be used which is geared towards emotions concerning the environmental impact of gas leaks and a surveyor's work (see appendix C.1).

Conative Dimension

When it comes to behavioral intentions and the willingness to contribute to a positive environmental impact through one's actions, the conative dimension should be considered [32]. Conation is also expressed as a form of *verbal commitment* which signifies that there is an intention to act [43]. Although the willigness to act can introduce pro-environmental behavior, it has to be noted though that it does not necessarily lead to a pro-environmental action itself [32].

In the context of gas leak detection, the willigness to act should refer to actions made by the surveyor in their work-related context and not as individuals in a private setting. For instance, conation in the surveying context could mean that the surveyor is willing to find super-emitters as fast as possible to minimize environmental harm or to diligently record all environmental damage caused by a gas leak.

Conative Dimension - Evaluation

Except for Maloney's scale of verbal commitment [43] which was adapted by Fraj & Martinez to explore pro-environmental behavior [23], there is little research on how conation can be measured. As the scale was also used by later research to understand the relationship between affect, conation and pro-environmental behavior [32], it will be adjusted and employed in this research. This entails that the statements will be rewritten to fit the surveyor and gas leak context (see appendix C.2).

Studies on Environmental Awareness

Some research on the concept of environmental awareness was motivated by the goal of increasing the pro-environmental purchase behavior of consumers [23], investigating pro-environmental attitudes of students [47] and understanding the relationship between environmental awareness, corporate social responsibility practices and a firm's reputation [48]. In terms of information visualizations, some research has been done to create awareness around domestic energy consumption behavior [5]. Other research focused on public areas, designing systems to raise environmental awareness for students in educational settings [49], [50] or for a community's engagement with environmental topics [51]. One study also assessed the environmental awareness of road freight transportation drivers to assess the success of green road freight transportation in China [52]. However, compared to consumer awareness, little has been done to research the environmental awareness of employees in an industrial setting [53] and the evaluation of environmental awareness for visualizations has been done differently than intended by this research. For instance, in one paper, students' environmental knowledge and attitudes where compared using the environmental attitude scale before and after interacting with a visualization [49]. In another example, a workshop was done with stakeholders to receive verbal feedback on a prototype of energy consumption data [53].

It can be concluded that so far, to the best of my knowledge, no SHCI research has yet investigated environmental awareness for information visualizations through the measurement of all three dimensions (cognitive, affective, conative). Thus, this research can be seen as a first step to bring the concept of environmental awareness from more psychological and business related research into the research of SHCI.

2.1.3 Cognitive Load

The human cognitive architecture is said to consist of a sensory register, a working memory with limited storage capacity and a long-term memory with unlimited storage capacity [54]. The cognitive load theory (CLT), coined by John Sweller in 1988, assumes that the working memory is limited in keeping novel information which is the reason why any instructional method should avoid overloading the working memory to optimize learning [55]. The concept of cognitive load encompasses three components which are *intrinsic*, *extraneous* and *germane*. Intrinsic cognitive load relates to the complexity of the information that has to be learned. Extraneous cognitive load refers to the learning activities and instructional procedures that are nonoptimal or uneffective, while germane load entails procedures which are indeed effective and lead to the establishment of schemata and knowledge in long-term memory [56]. Newer cognitive load research removed germane load as its own cognitive load category, stating that germane load has a "redistributive function from extraneous to intrinsic aspects of the task rather than imposing a load in its own right" [57] (see figure 2.3).

As the formation of knowledge in the form of schemata can extend the working

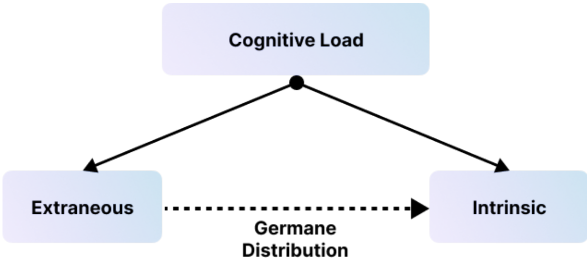


Figure 2.3: Cognitive load has intrinsic load which refers to the complexity of the knowledge, extraneous load which stems from the learning activity and a germane distribution function which transforms extraneous into intrinsic load (based on Sweller et al. [57]).

memory capacity and help process more intellectual tasks such as problem solving, one goal of instructions should be to support the construction of schemata, or germane load [58]. The problem-solving capacity is also dependent on the prior knowledge of the learner as novice learners might need more guidance than experts who can profit from their long-term memory in problem-solving tasks [59]. In general, there is a relation between cognitive workload and SA as cognitive workload affects cognitive capabilities and therefore also influences a user's SA [33].

In this research, cognitive load will be one criterion to assess the visualizations as lowering the cognitive load is important to guarantee that the surveyor has the capacity to take in all the information and as there is some indication that a high cognitive load can negatively affect a user's SA. Since measuring the effectiveness of a visualization through performance metrics such as response time and accuracy does not take into account that the same performance of two designs can be achieved through different mental efforts [60], cognitive load can be an important evaluation criterion to ensure that users can use their working memory efficiently and that they can maintain the mental capacity for other tasks.

Cognitive Load - Evaluation

Measuring the cognitive workload has been common in safety-critical domains such as healthcare [61] or aviation [24]. Two frequently used subjective reporting tools are the NASA-Task Load Index (NASA-TLX) [62] and the Subjective Workload Assessment Technique (SWAT) [63]. While the NASA-TLX is said to be suitable as a general reporting tool of experienced cognitive workload, the SWAT is more useful to investigate individual differences between the participants [64]. Although (indirect) objective measures like eye-tracking, time-on-task and physiological measures have been increasingly used, subjective measurements remain to be the most common form of cognitive load evaluation [11], [65].

Thus, although the HoloLens 2 device allows the collection of articulated eye-tracking data [66], the subjective measure of NASA-TLX will be used in this research due to its prominence in cognitive load research.

Cognitive Load - AR / MR

A systematic analysis of previous research on cognitive load in AR revealed that there is both evidence for a higher and lower cognitive load in AR with more papers postulating a lower cognitive load [11]. In their paper, Buchner et al. [11] pointed out that AR improved cognitive load for procedural knowledge like in assembly, navigation or flying tasks as well as collaborative problem solving due to the split-attention effect which states that information integrated into the task increases performance as no attention has to be split. Moreover, spatial AR such as immersive museum visualizations have been deemed superior in comparison to see-through mobile AR because they allowed users to perform tasks using gestures and movements they are familiar with, rather than forcing them to learn new gestures that were necessary to interact with see-through AR [11] which emphasizes the potential of a MR device such as the HoloLens 2. To measure the cognitive load in MR with a HoloLens, previous research commonly utilized the subjective NASA-TLX. For instance, a neurosurgical MR system was designed to help surgeons locate the right position for catheters in the brain [25]. Or in another study, the NASA-TLX was used to evaluate a network data visualization with a HoloLens to find anomalies in the data [26]. These examples suggest that the NASA-TLX is a suitable method to evaluate visualizations in terms of cognitive load which is the reason why it is chosen as an evaluation method in this research.

2.1.4 Mixed Reality

In its first definition, MR was viewed on a reality-virtuality continuum in which MR would encompass both AR, which augments the real environment through virtual objects and augmented virtuality (AV), which augments a virtual environment through real objects [67]. A newer definition was proposed by Hoenig et al. [68] who defined MR through the aggregation of three conditions: i) it combines physical objects in at least one physical environment and virtual objects in at least one virtual environment, ii) it runs interactively in real-time and iii) it spatially maps physical and virtual objects to each other in order to create interactions between them.

In general, MR technology was mentioned to have some benefits and drawbacks that define its scope of applicability. On the one hand, a significant advantage of MR glasses as optical see-through AR was explained to be that the user is able to see and interact with other humans and the surrounding environment as well as other non-MR tools (e.g., paper, maps) without the need to leave the interface. On the other hand, MR displays such as the HoloLens 2 from Microsoft have a fairly limited field of view (52 degrees) which increases the need for head movements and were said to be impractical for bright light settings [66].

In this research, MR, i.e. a HoloLens 2 device, will serve as the technological platform of the 2D and 3D visualization as it allows the perception of both 2D and 3D dimensions through head movements and the layout of information in 3D space.

Mixed Reality Applications

MR applications have been designed and developed for many different sectors ranging from healthcare and education to industrial settings and environmental monitoring. For example, in healthcare, a HoloLens as see-through AR could be used by engineers and doctors to study anatomical structures as 3D images [12]. In the educational sector, MR was put to use in order to investigate the learning motivation of student teachers using such technology [69]. MR has also been used for remote collaboration. In one study, a geospatial MR collaboration system was developed which displayed immersive visualizations and allowed multi-modal interactions that helped creating a shared SA in remote collaborative settings [70].

Another common application of MR is industrial maintenance work. For instance, in one study, a HoloLens was employed to superimpose the 3D model of a digital twin on the physical aircraft component to help the inspector report the damage directly on the 3D model to be used for later repair. Through allowing remote experts to be involved in the documentation process, the inspection was less limited by the number and availability of the experts and transportation costs could be saved [71]. In another study, a Hololens was put to work in order to scan QR codes of infrastructure assets to increase awareness of infrastructure in the inspection work. As the technology should not hinder the inspector to carry out their main task, it was developed to be pervasive rather than intrusive. By augmenting human perception through the use of technology and not relying solely on the human capabilities to sense, MR could be a powerful visualization technique for prompt decision-making [20] which goes in line with the idea of the Industry 5.0 [18].

While some AR applications exist that have an environmental focus, little research has been done in the area of MR. In an attempt to survey and categorize VR / AR / MR applications based on their environmental contribution, Rambach et al. [72] divided applications into those with a direct and indirect environmental impact. A direct impact meant that those applications aimed at environmental monitoring, creating ecological awareness, resource efficiency or promoting environmental education, having a direct environmental purpose. An indirect impact suggested that they were about remote collaboration, assistance, tourism and used immersive technology for simulation or training, removing the need to travel far distances.

As this research was geared towards promoting environmental awareness, applications which fell under this category were the most interesting. One application which was developed to promote ecological awareness was the mobile AR application *Aire* which visualized air pollution and contaminants present in the environment through colored particles [8]. Another example was a mobile AR application which allowed the scanning of markers on home appliances to visualize their energy consumption as numerical text [73]. Since MR simulations allow the user to easily and safely access hard-to-reach environments for learning, a HoloLens application was developed to learn more about maritime habitats [22]. While the air pollution application was not evaluated at all and the energy consumption app was merely tested for its functionality, the third application was evaluated only with knowledge questions and the User Experience Questionnaire (UEQ) to compare the user's experience with traditional text books.

To conclude, while MR technology has already been introduced into industrial work settings, no previous research could be found which made an attempt to use MR in a gas leak surveying context. Furthermore, while some environmental applications in MR exist, it is still more of a niche research area and a paper claiming to investigate environmental or ecological awareness was limited to the use of knowledge questions [22]. Although cognitive awareness in the form of knowledge should be part of the evaluation, no study has evaluated the entire environmental awareness concept (cognitive, affective, conative) in MR yet as explained in section 2.1.2. Thus, this research seeks to bring the topic of environmental awareness into current MR research.

2.1.5 Information Visualization

The term *information visualization* was defined by Mackinlay & Card [74] as "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition". Such a graphical transcription of numerical data allows users to grasp a concept which can not be fully understood without seeing it [9].

To understand the impact of data visualizations, it is important to consider how they are processed by the human brain. Human visual analysis involves two forms of processing: One is preattentive processing which understands simple, spatial features in parallel and the other is the later processing phase which requires focused attention to combine these separate features into more coherent objects [75]. Preattentive processing includes features such as color, closure and orientation which can be perceived in about 200 ms and allow humans to assess the approximate amount and location of objects [76] which results in a lower cognitive load [77]. For example, semantic depth of field is a preattentive feature which blurs out content that currently does not require attention [78]. Concerning the dimensionality of data, a study by Lugmayr et al. [79] concluded that visualizations in 3D, VR and immersive technology supported cognitive functions such as pre-attentive processing, suggesting that 3D visualizations might lead to a cognitive benefit despite the additional dimension which will be explored in this research through the measurement of cognitive load and cognitive environmental awareness. In the following sub-sections, previous research regarding the dimensionality (2D vs. 3D) and form of the presented visualizations will be discussed.

Dimensionality of Visualizations - 2D and 3D perspectives

Comparing 2D and 3D visualizations, previous research pointed out that 3D visualizations allow humans to use their spatial memory and therefore outperform 2D representations when it comes to spatial data. This benefit was for example found by visualizing biomedical data such as microscopic images and anatomical data sets for clinicians [12]. Spatial data was also explained to be relevant for environmental monitoring which is the continuous measurement and observation of environmental parameters to find a phenomenon that requires switching between visualizations (plots, numbers) and the real world (maps, photographs) to understand the origin of a phenomenon [80]. Furthermore, AR / MR was said to be well suited for enhancing learning of three-dimensional spatial data as visual, motoric and proprioceptive feedback can be given during the interaction with the virtual objects in the real world [81]. In fact, one study suggested that 3D representations increased learning and were less cognitively demanding when used with see-through googles and spatial AR such as immersive museum environments compared to 2D visualizations [11].

However, while 3D might be useful for spatial data, this might not be the case for single-point data measurements for which 3D would be rather ”overrated” [11], suggesting that the benefit of MR depended on the type of data. Apart from the type of data, the level of expertise might also be essential. Through the conduct and analysis of expert interviews, one paper concluded that experts might prefer using 2D, while non-experts understood 3D visualizations better. Especially for work in unfamiliar environments, 3D visualizations were discussed to be more useful than 2D visualizations for environmental monitoring tasks [80].

Other downsides of 3D visualizations were mentioned as depth perception and occlusion problems [76], [82]. It means that the missing of natural depth cues such as occlusion or shadows can hinder users to understand the spatial relationship between virtual content and real objects which is essential to achieve visual coherence [83].

Lastly, one paper proposed the use of immersive media such as AR / VR / MR to convert abstract numbers and statistics into something more palpable. For example, a product's ingredients in a supermarket could be conveyed in a verbal, visual, metaphorical, realistic, spatial or symbolic manner and it could be enhanced with additional information such as its transportation way or the energy and labor required [84].

In conclusion, as previous research emphasized the benefit of 3D visualizations for spatial data, this research will try to extend current 2D / 3D research by exploring visualizations of localized, invisible data.

Information Visualization - Forms

In their paper, Rist & Masoodian [13] presented a categorization of interactive energy data visualizations and discussed them based on their phenotype, temporal dimension, visual representation, dynamic in time and purpose. They defined six categories: i) charts & graphs, ii) energy gauges, iii) eco-visualizations, iv) visualizations for analytics, v) gamified and serious game visualizations and vi) ambient & physical visualizations. From these categories, i) charts & graphs and vi) ambient & physical visualizations were later used for the 2D and 3D visualization designs described in section 3.3.2. The category charts & graphs will be referred to as statistical from now on based on Chalal et al.'s categorization of visualizations [6]. On top of the statistical and physical categories, icons and numerical text were also used for the visualizations concepts. A discussion of the different visualization forms used in the study can be found below.

Statistical Statistical representations are manyfold, including e.g. line charts, pie charts, (stacked) bar charts or radar charts. In general, charts and graphs show a minimum *base load* and a maximum *peak load* over a certain time span which is useful when comparing the data to another data set or baseline and allows the detection of cyclic patterns [13]. In the case of bar charts, they can be further defined as categorical visualizations as they allow the comparison of categories [85]. Some examples were discussed as the monthly energy consumption [53], water usage by area [86] or, in this case, a bar chart could display the different gas levels per gas type. According to a systematic review by Chalal et al. [6], statistical representations are the most common energy consumption visualizations (42%) which is the reason why they were chosen for the 2D visualization concept.

Iconographic According to Peirce's [87] theory of semiotic, a sign has a triadic structure consisting in three components: The signifying element which signifies the object, the object itself which should be represented and the interpretant which establishes the reasoning based on the object and signifying element [87]. An icon can then be just seen as a sign which conveys some fact about a software feature. For example, in the context of a printer, an object is the print document, a signifying element is the print symbol which results in the interpretant of this sign to be the action of printing the document. The relations between these three components can go wrong in a way that the signifying element has no connection with the object or that the interpretation of either signifying element or object goes wrong [21]. Furthermore Peirce defined different sign types from which one is essential for the 2D visualization design in section 3.3.2. The iconic icon signifies that the signifying element resembles the object [87]. This type of icon design has the advantage that it adheres to a human's real perception of the world which enables fast and cross-cultural understanding as the physical representation should be understood in the same way across cultures [21]. A guideline postulated by Barr et al. [21] stated that icons that represent qualities or system objects, i.e. do not represent system actions or non-metaphoric content, should be iconic signs. Thus, iconic icons were used in this research to signify the represented data.

Physical Physical representations are those which are incorporated into physical environments [13] as is the case with AR / MR which references the real world. In comparison to statistical energy consumption visualizations, this type of visualization is less common, covering only 8.1% of the reviewed papers in a systematic analysis [6]. As explained in section 3.2.1, two types of data are of interest for this research: air particles and wind direction. Both represent physical phenomena in the real world and therefore belong to the category of scientific visualizations which are mainly based on accurately displayed, physical data [88].

Air particles and wind have both been visualized by previous research. With regard to the air particle data, one example was the AR app *Aire* whose purpose was to "augment reality that is invisible to the human eye" by displaying the size, color, material, position, quantity and movement of microscopic particles in the air. The app represented gaseous pollutants as chemical compositions, while rendering any solid pollutant as an actual solid particle. Moreover, it spawned the quantity of particles based on the amount of pollution close to the location of the user [8]. Another research project which should make people aware of the pollution surrounding them and make them realize the importance of pro-environmental decisions was the AR app *AiR*. By taking the GPS coordinates of the user and the time, changes in the pollution levels could be perceived. To categorize the pollutants which were represented as 3D particles based on their chemical structure, a dashboard showed all the pollutants, their concentrations and individual color scales based on their severity. By randomly distributing the pollutants in the air space based on the pollution data, users could experience the visualization from their own perspective through their smartphone and move around to grow a sense of involvement [7]. A last example, which should be mentioned at this point, was an air-purification system which was developed for the HoloLens 1 and displayed the air purification over time as streamline and particle visualizations. Here, particles were represented as small spheres for which a red particle color meant that the space was full of pollutants, while a purification and clearance of pollutants was indicated by a blue color [89].

The second type of data concerns the wind direction or flow. In their paper, Christmann et al. [9] revealed that arrows were a common and understandable way to communicate the direction of flow. In 3D, wind can be represented by slices, iso-surfaces, streamlines and arrows with different colors and sizes [9]. Vectors can be used to visualize the direction of an airflow and the velocity or force through the norm of a vector [90]. Through their experiment, Christmann et al. [9] concluded that a scene was difficult to see through the air flow visualizations, suggesting that the user should filter the area where they want to see the air flow. So instead, the user would only have a partial representation on a surface or a volume which would only show the localized data where a relevant decision had to be made.

To sum up, all presented projects had the goal of making the intangible feel tangible which was also the goal of this research seeking to visualize gases and wind and using these projects as inspiration for the final 3D visualization design in section 3.3.2.

Textual For the sake of completeness, text in the form of labels and numerical text has to be included in the listing. Text is the simplest form of data representation which often accompanies other forms of visualizations. Two principles from the cognitive theory of multimedia learning that are relevant for this visualization form are the multimedia and the spatial contiguity principles. The multimedia principle postulates that when two representations like text and pictures or icons are combined, it has a positive effect on learning due to the fact that different cognitive channels are used. The spatial contiguity principle represents the idea that the text should be close to the picture or icon it describes so that the user does not need to search for information [91]. Both principles regarding text were taken into consideration while designing the visualizations.

While many visualizations utilize graphical representations to communicate the information [6], some visualizations are based on text alone. One example of a pure numerical visualization was a mobile IoT based AR app which presented the power quality and energy consumption per appliance in a room. The app also deployed color coding in white, blue, green and red to convey how close the number was to a certain energy value [92].

Lastly, it has to be noted that since numerical text and labels will be used for both 2D and 3D visualizations to accompany the graphical representations, the textual representation will not be used for the comparison between visualization concepts.

Information Visualization - Evaluation

Information visualizations have been evaluated by previous work using many different methods and having diverse foci. For instance, in a systematic review, Lam et al. [93] presented seven common evaluation scenarios for visualizations which consisted in evaluating environments and work practices, visual data analysis and reasoning, communication through visualization, collaborative data analysis, user performance, user experience and automated evaluation. Mapping this research to their categories, visual data analysis which is about supporting users in developing actionable and relevant knowledge in their domain as well as user performance and user experience were of interest. To test the first category, Lam et al. proposed to use case studies in which domain experts would interact with the visualization to answer questions as well as controlled experiments to limit the research area. To measure performance, they suggested to use controlled experiments and field logs to measure accuracy through the automatic collection of error rates. For gaining insights into the user experience, structured interviews were recommended as a way to gain direct qualitative feedback and questionnaires in a five- or seven-point Likert scale format.

Data visualizations have been either tested on their own or in comparison with others. One example of an independent evaluation was a MR collaborative geospatial visualization in which the participants first performed a training session to get familiar with the data exploration tasks, then performed a series of tasks and filled in a post-session questionnaire [70]. While such evaluations allow to validate a certain idea, they do not allow to gain further truth about underlying phenomena that can be found through comparison. For instance, in one study, two visualization techniques that represented environmental changes, i.e. radar graphs and colored boxes, were compared to foster awareness in sustainable development by evaluating the accuracy of right answers to nine cognitive test measures. The goal was to understand how individual, cognitive differences between users might render one visualization technique more effective than the other [94]. In another paper, immersive analytic applications were evaluated in terms of their effectiveness in helping financial data decision-making. Using a within-subject study comparison of 2D and 3D visualizations in a web interface and MR device where the participants performed a series of tasks, the researchers measured the task completion time, perceived precision and perceived usefulness to understand the suitability of both concepts [82]. Lastly, it should be noted that while task completion time might be necessary to assess in time-critical tasks, it does not predict the effect of properties on visualization comprehension and thus is not relevant for visualization accuracy [95].

To conclude, the evaluation methods of this research were mostly based on Lam et al.'s suggestions [93] along with the recommended methods for environmental awareness and cognitive load discussed in the corresponding sections 2.1.2 and 2.1.3.

Chapter 3

Methodology

This research followed a UCD process which signified that the final MR prototype was iteratively developed by revisiting the first UCD phases and testing out the look and feel through own experimentation and pilot studies. The UCD process encompasses four distinct phases which defined the trajectory of this research: 1) Understand context of use, 2) specify user requirements, 3) design solutions and 4) evaluate against requirements (see figure 3.1) [96]. In the first phase of the UCD process, the researcher collects qualitative data, e.g. through interviews or observations, in order to understand the context of the user. Then, the second phase entails analyzing this qualitative data to define the needs of the users and the requirements of the system. The third phase is the ideation and prototyping phase in which new solutions are designed and developed. Finally, in the last phase, the designed solution is tested out against the user's context and previously defined requirements. Apart from the UCD process, other design methodologies exist. While UCD focuses on understanding the human and their needs, activity-focused design rather centers around the actions humans need or want to take so that they are able to reach a certain goal. Common approaches of this design methodology are task analysis, activity theory and activity-based design [97]. A methodology that moves beyond simple tasks and actions is known as goal-directed design which was coined by Alan Cooper and is rooted in behavioral design. Here, the focus lies on the user's goals, i.e. the motivations for using a product, while the activities and tasks are understood as the concrete steps to help the user achieve their goals [98]. A last methodology which should be pointed out is design thinking which aims at coming up with innovative and create solutions for complex problems. As UCD, it is an iterative process and has five distinct phases known as empathize, define, ideate, prototype and test [99]. It is especially used to solve so-called *wicked-problems* which are complex, sociocultural problems that are difficult to solve because of their interconnected nature [100].

For this research, UCD was chosen over other human-centered methodologies for the following reasons. First, while understanding about the activities and goals of gas leak surveyors was essential to come up with the content for the visualizations, the focus in this research was to understand how form and dimensionality of data visualizations affected a human's awareness and cognition. Therefore, it was not about diving deep into the process of different surveying activities and coming up with a solution to help surveyors achieve their goals in their daily life as it would be the case for either activity-based or goal-directed design. Instead, this research was limited to viewing a visualization with little interaction and interest in the user's overall journey with the system. Finally, design thinking was not chosen since the surveyor's problems seemed to be rather known, less complex and intertwined with sociocultural components. Overall, this research was more focused on understanding about a visualization's qualities rather than solving a problem. Together, these arguments made UCD the most suited design methodology to satisfy the purpose of this research.

To begin, four remote interviews were conducted with ABB employees in the US and Canada who were affiliated with ABB's surveying products and/or had previous experience as surveyors. From this understanding about relevant environmental aspects, the measured data and the surveyor's workflow, the user requirements for the visualizations were formulated. Thereafter, the design space was explored by drawing sketches on paper and then iteratively developing the prototype in Figma and Unity based on own experiences and feedback from pilot participants. In the end, a statistical, iconographic 2D and a physical 3D visualization were tested through a survey with experts and a usability study with non-experts to understand their level of environmental awareness and cognitive load for each visualization concept. To measure the environmental awareness, two different methods were employed. For the cognitive dimension, the SAGAT method was used which is a method that can assess the SA and is commonly used for navigation tasks [37]. For the affective and conative dimensions, a questionnaire from Fraj & Martinez [23] was adapted to fit the context of gas leak surveying. Finally, the cognitive load was measured using the NASA-TLX which is a subjective method with six sub-dimensions in mental and physical demand, pace, success, effort and stress [62].

In general, all activities were carried out solely by the author except for the initial expert interviews for which the whole UX research team of five UX researchers was present and where the questions of this research were part of a larger framework. A complete overview of the methodology mapped to each of the four phases of the UCD can be viewed in figure 3.2.

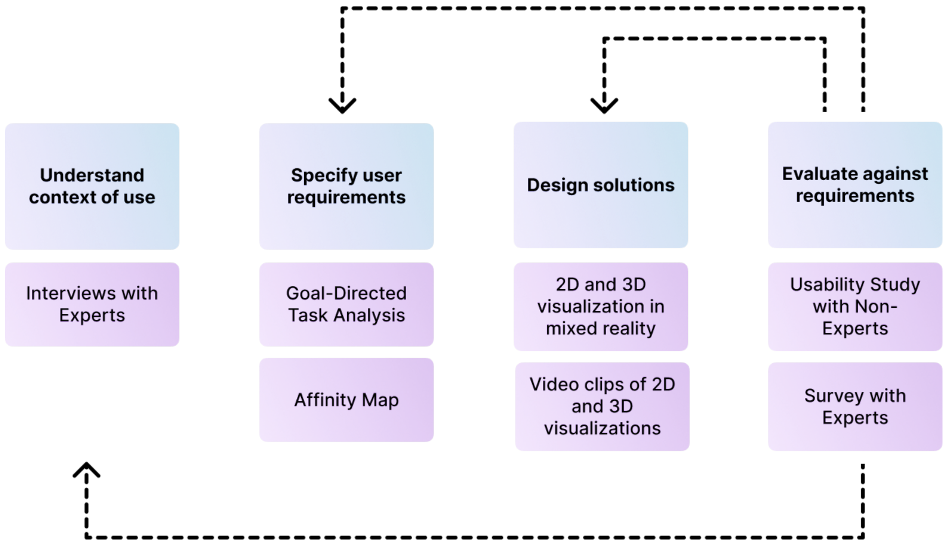


Figure 3.1: The user-centered design process consists in four phases: i) Understand context of use, ii) specify user requirements, iii) design solutions and iv) evaluate against requirements (based on [96]). It displays the activities of this research for each phase, starting with interviews with experts to understand the data types through an affinity map and the informational requirements through a goal-directed analysis. Then, a statistical, iconographic 2D and physical 3D visualization were built in MR which was evaluated through a usability study with non-experts. Moreover, video clips of the MR prototype were created and evaluated through a survey with experts.

3.1 Phase 1 - Understanding context of use

3.1.1 Interviews With Experts

To become familiar with the domain of surveying and the sustainability aspect of surveying, a series of four interviews was remotely performed with subject-matter experts who had experience with the ABB surveying suite (mobile app, car, drone) and the surveying workflow through own surveying experiences or observations. Interviews were chosen as an exploratory method because they allowed to gain rich and detailed data about a topic [101] which was important in this case since the domain of interest required highly technical knowledge in leak detection, the involved gas data, the technologies used as well as navigational and environmental hurdles. The interviews were semi-structured to give researchers the freedom to deviate from the interview script and ask follow-up questions in case a topic needed further clarifications [101]. Moreover, the interview questions were open-ended to gain rich, qualitative data about each participant's unique perspectives. As all the participants were located at a different location than the researchers, the interviews were per-

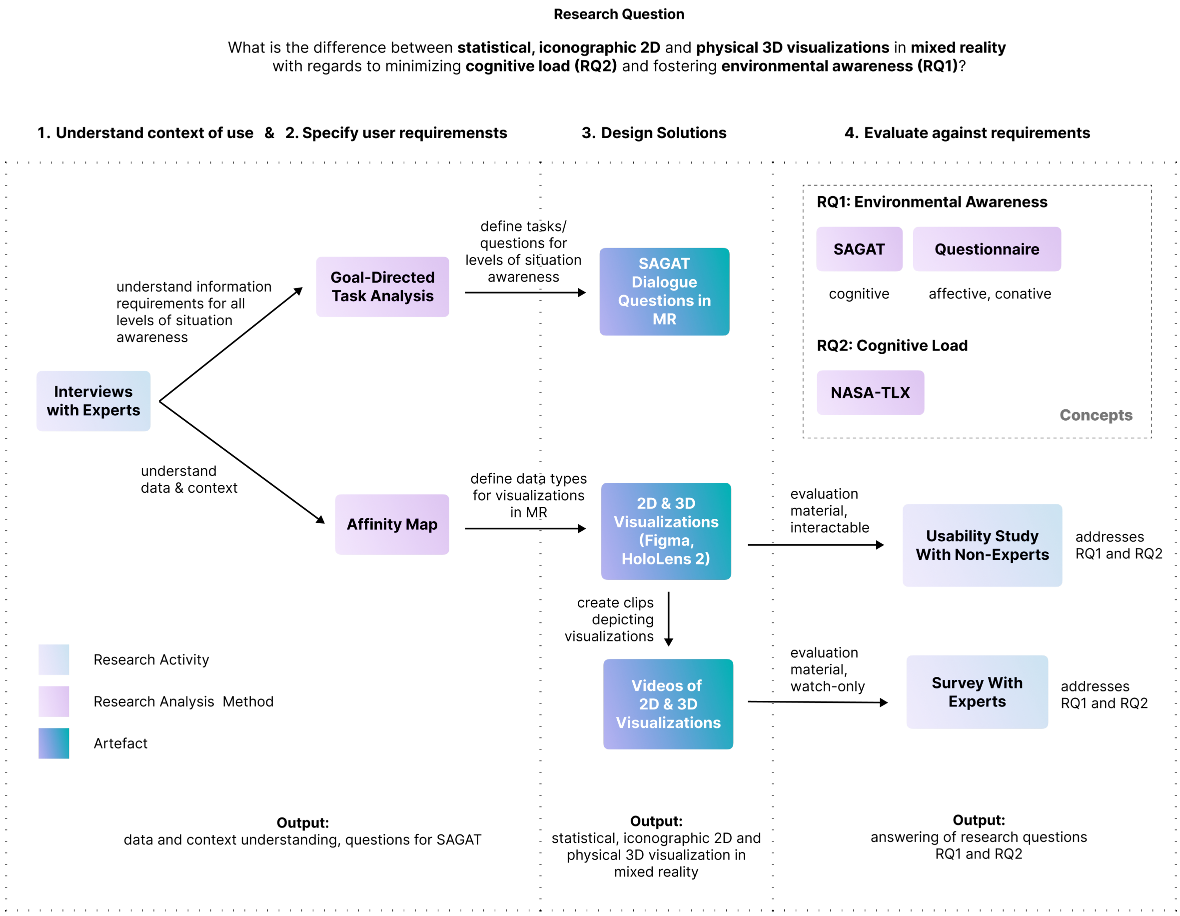


Figure 3.2: Methodology overview which shows the activities for each phase of the UCD. In the first two phases, interviews with experts were analyzed through a goal-directed task analysis and an affinity map. In the third phase, sketches were made and translated into a Figma prototype which was then coded in Unity and filmed to create scenario videos. Lastly, in phase four, a usability study and survey were conducted with non-experts and experts using SAGAT, NASA-TLX and other questionnaires.

formed remotely. Concerning the content of the interviews, the participants were asked to provide information about their role and experience with surveying, current surveying practices including the tasks and devices involved, surveying challenges as well as sustainability related questions about their attitudes, their work's impact and the environmental aspects related to gas leaks. The questions were adapted for subsequent interviews to fill in knowledge gaps based on previous interview data in order to gain a holistic view of the area of interest. An outline of the interview questions relevant for this research can be found in the appendix A.1.

Participants

The interview participants were ABB employees situated in the US and Canada who were over 18 years old and had experience with the ABB surveying suite and leak detection in general, either through their own surveying experience or through observations of working surveyors. As interviews can be categorized as a qualitative research tool, the sample size was chosen based on the principle of saturation which defines the point at which further data does not add any new theoretical insights to the existing data [102]. The saturation of data could be reached with a low number of participants (n = 4) since leak detection involved a clearly defined work procedure with little deviation in the data measured or reported.

Procedure

Each interview was performed with the whole UX team present and lasted 60-120 minutes. The session started by a short briefing in which the participant received information about the procedure, the data treatment and their rights as study participants. Each interview session was audio-recorded for later analysis. During the interview, three UX researchers, who were assigned certain parts of the interview flow, asked questions, while the other two could interrupt to pose follow-up questions. For each following interview, the questions were adjusted to reduce redundant information and allow more in-depth questions to fill any existing knowledge gaps. After the interview session, the interviewee was informed about how the data would be used in the future and asked for consent to follow-up questions by mail.

3.2 Phase 2: Specify user requirements

The resulting data was analyzed through an affinity diagram which meant that data was written on Post-its, in this case virtual Post-its in the collaborative web-tool Miro, and then clustered into groups based on their similarity [103]. Affinity diagramming was chosen over more meticulous research methods like a thematic analysis, which involves coding data to find patterns and overall themes [104], because the goal of this activity was to organize the contextual data to form a knowledge foundation rather than discovering some underlying user needs. The outcomes of this activity were different types of environmental data measured and reported through the surveying app, data which should be measured and reported but currently is not captured yet, current leak detection visualizations and challenges surveyors' face with these visualizations, e.g. based on their workflow and other external influences. A complete version of the affinity map including the discovered groups can be viewed in appendix A.2.

Aside from an affinity diagram, a goal-directed task analysis (GDTA) was performed to investigate the informational requirements for all levels of SA, identifying the surveyors' goals and dividing them into sub-goals. While the affinity diagram provided the knowledge to be tested in an evaluation, the GDTA produced the tasks or questions that the participants would need to solve. A GDTA has the advantage that it is technology and process independent and produces accurate informational requirements and key goals for individuals [105], as it was for instance the case for the context of maritime navigation [106].

3.2.1 Affinity Diagram

The affinity diagram allowed the clustering of information gained through the expert interviews in multiple areas of interest such as the data types, current and envisioned visualization ideas, the sustainability aspect of surveying such as the environmental hazards of leaks and the environmental impact of surveying. The correctness of the information shown in the affinity diagram was verified through an online literature review and follow-up questions to the experts. In the following sections, only the information used for the evaluation will be explained. To review the complete affinity diagram, it can be found in appendix A.2.

MicroGuard - The mobile app

MicroGuard consists in a surveying application within a mobile phone which is attached on a physical wand together with an analyzer in the backpack to measure different gasses and understand their meaning. A gas sample can be taken from the tip of the wand and move through the wand into a tube connected to the analyzer. The analyzer then communicates the analysis results to the mobile app (P1, P2, P3, P4) (see figure 3.3).

Leak Definition

A leak was described as the highest methane concentration point in a certain area (P3). Moreover, it was said to be more of a qualitative rather than a quantitative concept, merely indicating the existence of a leak rather than defining the exact location of a leak (P1). Another aspect which was mentioned during the interviews was the difference between a leak and an emission. While every leak represents emissions, not every emission indicates a leak which is the reason why a surveyor needs to go to the emission location and verify if a leak is present (P2). Concerning the grading of leaks, a leak is commonly categorized into three grades which are

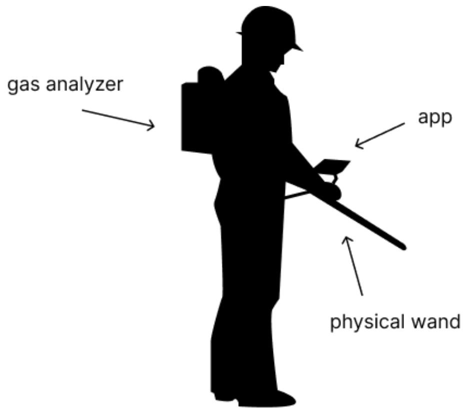


Figure 3.3: A surveyor holding a surveying wand which can suck in a gas sample at its tip. The sample then moves from the wand into a tube which is directly connected to a gas analyzer in the backpack. The result of the analysis is displayed on a phone attached to a physical wand (based on [107]).

based, amongst others, on the methane concentration level and the distance to critical infrastructure [108], therefore revealing a connection between the aspects of sustainability and safety (P3). A grade 1 leak is hazardous and requires immediate repair, a grade 2 leak is non-hazardous and requires a scheduled repair, while a grade 3 leak is non-hazardous and only requires monitoring [108].

Data Types

Methane Concentration (in ppm): Methane was said to be a component of natural gas and a potent greenhouse gas which is more potent than CO_2 and can be measured in parts per million (ppm) (P1). The concentration was defined as a point sample which means that the concentration is measured at a certain point in space and can be therefore described as localized (P4). While methane was described as relatively non-toxic, it was said to possess a large heat index which made it hazardous, potentially leading to explosions and posing a safety risk (P1).

Ethane / Methane Ratio (in %): Ethane was mentioned as another component of natural gas which served as a fingerprint to decide if there is a leak or not by differentiating between biogenic and thermogenic sources (P1). Biogenic gas was described as a result from bacterial activity in landfills, waste and wetlands, while thermogenic gas was a result from petroleum transformation (P2). A misconception of the biogenic and thermogenic categories relying solely on the ethane concentration rather than an ethane / methane ratio was later resolved in the first survey experiment with a surveyor (see subsection 3.4.3) who examined the designed material and questions to see if there were any wrong misconceptions, correcting that the threshold of thermogenic gas was a 3 % ratio between ethane and methane rather than a 3 % ethane concentration.

Methane Leak Rate (in kg/s): The leak rate was defined as the quantification of a leak in kg/s or l/min (P2). More concretely, it was portrayed as a flow rate through a medium such as a pipe or in the case of a mass flux rate, the expansion of a gas plume per unit time (P2). While the leak rate was not yet measured through the mobile surveying app, all participants explained that the leak rate was an essential measurement to become more sustainable. Apart from the environmental impact, a measurement of the leak rate also affected the business as a whole by becoming a tool for customers to communicate how many emissions were emitted through a leak and by allowing them to take credits for the repair work (P3). As the leak rate was a new data element not yet integrated into the current surveying app, the severity of the leak rate in the visualizations was based on Log & Pedersen's logarithmic table for safety and non-safety related fugitive emissions [109].

Wind Direction: According to P3, the wind direction revealed how a surveyor should navigate a terrain. As no anemometer was used yet which could measure the wind direction and speed, other indirect ways of understanding the wind direction were used. For example, P2 mentioned pieces of leaves which were thrown up to see how they were carried by the wind, while P3 mentioned a red string attached to the surveying wand which could move with the wind.

Wind Speed: Wind speed was described as a parameter which could give indication on how far away a leak was. A faster wind speed meant that the force to carry the gasses away was stronger and a slower wind speed meant that less gas was carried away (P2). As the wind speed was not yet integrated into the surveying app, the Beaufort wind scale [110] was used to come up with three wind speed situations for the visualizations.

Sustainability Impact

According to P2, leak detection was intrinsically a sustainable action as no leaks meant no emissions (P1) and the more leaks could be fixed, the better for the environment (P3). The environmental hazards from leaking gas were emissions (P1, P2, P3, P4), contamination of soil and (underground) water (P2, P4) and dead environments due to explosions (P3). None of these hazards was currently measured or reported since emissions do not have an accurate quantification yet (in the form of the leak rate) and dead spots on grass were only used as an indicator for a leak

(P1). Moreover, while explosions could result in harm for both humans and the environment, they were only seen as a safety risk (P3). To become more sustainable, all the participants explained that a better quantification of a leak would be needed which would also render the product more attractive for customers.

3.2.2 Goal-directed task analysis

The goal-directed task analysis displayed one main goal which was to assess the environmental impact of a leak. To achieve this main goal, three sub-goals were formulated with their corresponding critical decision and informational requirements to become aware of the situation. These sub-goals were defined as to first identify a leak, then to understand the short-term, local impact and finally to understand the long-term, global impact. The informational requirements were split into the three SA levels of perception as level 1, understanding as level 2 and projection as level 3. The complete analysis is depicted in figure 3.4.

Subgoal 1: Identify a leak

The first sub-goal was declared as identifying the existence of a leak because the environmental impact of a leak could only be assessed if a leak was actually present. Therefore, the critical decision was to understand if a leak was found and should be reported. The lowest informational requirements on level 1 were defined as the methane concentration and ethane ratio (corrected after the first survey in section 3.4.3), wind direction and speed as well as the location of the pipeline and wand. Based on this data, level 2 information requirements were formulated and included the questions of whether the wand was above a pipeline, whether the ethane ratio was biogenic or thermogenic, whether the methane concentration was above the atmospheric background level and whether there was a chance that the leak would be further away based on the current wind direction and speed. Finally, the third level of SA referred to the questions of whether a leak was found and whether a surveyor needed to stay or continue moving because the highest concentration might be further away.

Subgoal 2: Understand the short-term, local impact

The second sub-goal of understanding the short-term, local impact involved the critical decision of how the leak should be prioritized and how it would affect its imminent surroundings. The level 1 information were defined as the methane concentration, the wand location and the location of critical infrastructure. On level 2, the questions were formulated as categorizing the leak based on its grade and deciding whether

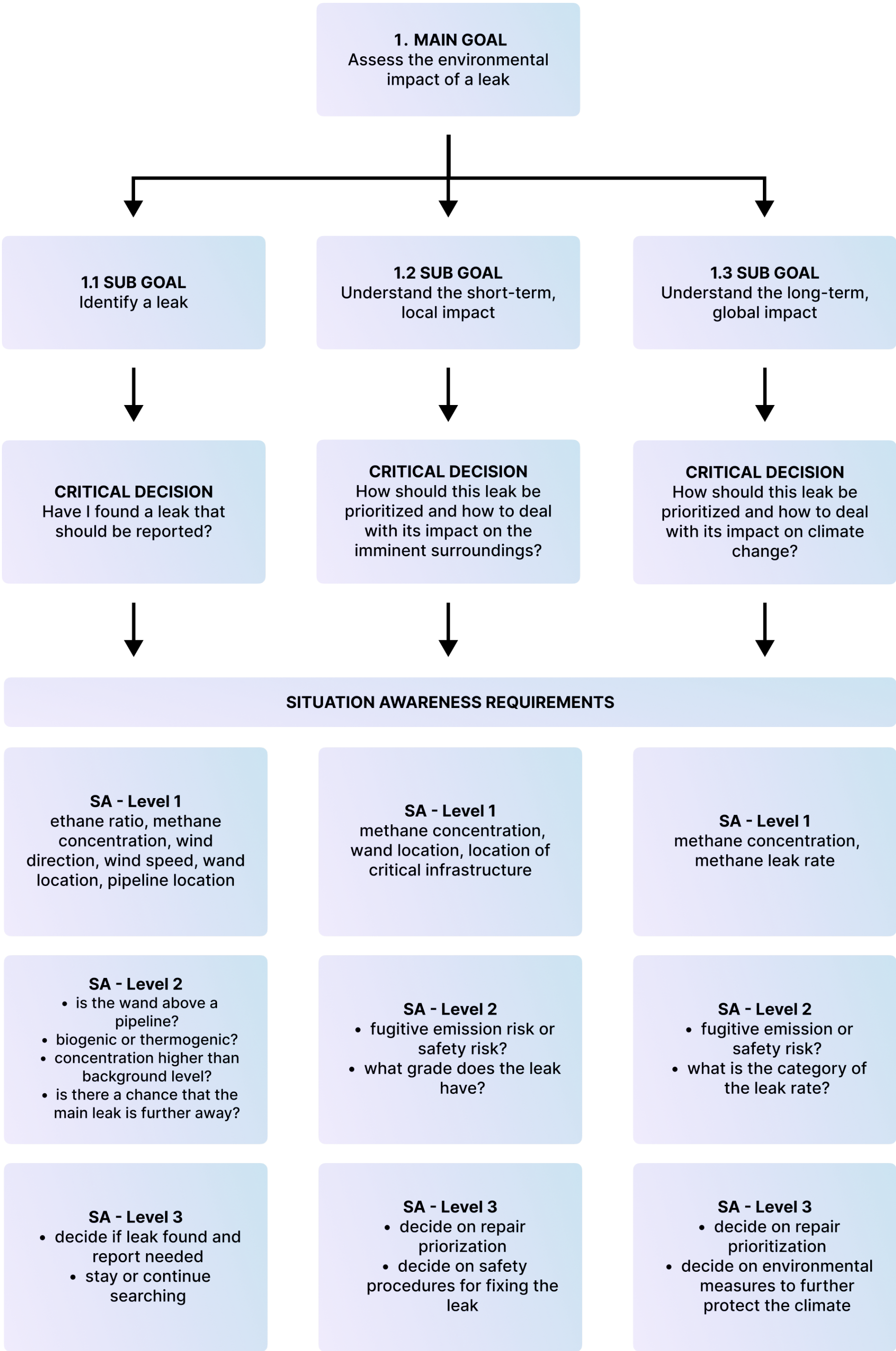


Figure 3.4: Goal-directed task analysis which defines as its main goal the assessment of the environmental impact of a leak. This main goal is further divided into three sub-goals and critical decisions, namely to i) identify a leak, ii) understand the short-term, local impact and iii) understand the long-term, global impact. For each sub-goal, a list of informational requirements can be found which corresponds to the three levels of SA (1. perception, 2. understanding, 3. projection).

the leak only resulted in the risk of fugitive emissions or whether it even passed as a safety risk. The third level involved the question of how the leak should be prioritized for repair and what kind of safety procedures would be needed to be followed by the repair team.

Subgoal 3: Understand the long-term, global impact

The last sub-goal implied the decision of how the leak should be prioritized and how decision-makers would need to deal with it based on its long-term, global impact on climate change through harmful methane emissions. For this decision, the methane concentration and leak rate were relevant on level 1. On level 2, it would be again important to understand whether the leak only resulted in fugitive emissions or whether it actually posed a safety risk. Lastly, level 3 would involve the question of how the leak should be prioritized for repair and what measures should be taken to further protect against a harmful environmental impact.

3.3 Phase 3: Design Solutions

The design phase was characterized by two activities: 1) An initial sketching phase and 2) an iterative and alternating process of designing a high-fidelity prototype in Figma and implementing a MR prototype in Unity for the HoloLens 2 device.

3.3.1 Sketches

To come up with first design concepts, sketching was used as an early, exploratory ideation method which involved drawing out visualization concepts to help suggest, refine and question thoughts in a palpable way. The advantage of sketching was that it represented a quick, low-cost method which allowed the creation of many ideas to explore multiple design directions simultaneously [111]. For instance, sketching was used in previous AR / VR research as a workshop technique to play with the first and third person perspectives of users to both understand the field of view and the big picture of the experience [112].

The sketches for this research focused on visualizations that would be shown at the tip of the wand and depicted different forms of data representations such as particle boxes, line charts, radial stacked bar chart, maps, scales and compasses as well as multiple color codings based on the different data levels (see figure 3.5). A decision was made to focus on the wand visualization and not include other external information such as pipeline data to minimize the complexity of the visualization and focus on data from the physical environment. The Figma prototype explained in the following section was based on these initial ideas.

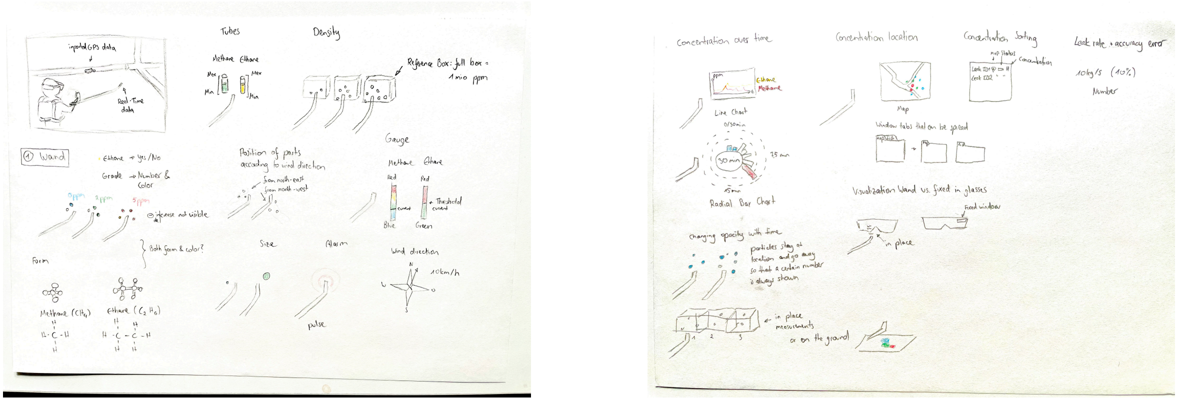


Figure 3.5: Sketches which depict different visualization ideas at the tip of the wand from free particles and constrained particle boxes to statistical line charts, stacked radial charts, compasses and color codings.

3.3.2 MRTK3 Figma Prototype

To conceptualize the experimental setup and data visualizations to be used with the HoloLens 2, the MRTK3 Figma Toolkit was employed which came with a ready-to-use set of UI components for MR development. Two visualization concepts were developed: A statistical, iconographic 2D and a physical 3D visualization. Both concepts were meant to be static in order to ease the development process which signified that a limited amount of data states were designed. The final look and feel was a result of six iterations which were sparked by own experimentation and the conduct of two pilot studies in which feedback about the visualizations was collected (see section 3.4.2). Both visualizations used five data elements: Methane concentration and leak rate, ethane ratio, wind direction and speed. An overview of the visualization decisions per data type can be viewed in table 3.1. Finally, the Figma design of both visualizations can be found in the appendix B.1.

2D Representation

The 2D visualization concept consisted in the statistical representation of two bar charts, icons and text. A statistical visualization was chosen as it was the most common way for visualizing energy eco-feedback systems according to a systematic review [6]. For the methane concentration and ethane ratio, a bar chart was used in order to compare the current value at the top of the bar with different threshold levels on the y-axis. For the wind direction, a compass icon was designed which could rotate based on the wind direction data. Both wind speed and methane leak rate had a static icon which did not change during data transitions. Finally, all data elements were accompanied by a label of what the data represented as well as the current value.

3D Representation

While the 2D visualization was designed to be flat, the 3D visualization rather mimicked the physical properties of the data. The methane concentration and ethane ratio were physically represented by a partial particle visualization which meant that particles were enclosed in two cubical boxes. This representation was inspired by an AR visualization depicting air pollutants as particles flowing in the air [7] and a realistic particle visualization where a partial view was recommended to only focus on a selected area in space [9]. The wind direction and speed were portrayed as a direct, flow visualization of an arrow pointing in a direction [9], [113], differing from previous research by displaying one instead of multiple arrows. The direction was indicated by where the arrow pointed at, while the wind speed or force was represented by the arrow's magnitude, mimicking a mathematical vector. Lastly, the methane leak rate also embodied a physical representation in the form of particles that came out of the methane concentration box to represent the fleeing methane emissions shaping a gas plume.

Data Levels

Both visualizations used colors to differentiate between different threshold levels or grades. For the methane concentration and leak rate, blue meant a low level, green a medium level and red a high one. The concentration levels were adapted from the mobile surveying app which used the same color schema. The ethane concentration was illustrated by a yellow color for a biogenic level and red color for a thermogenic level. In the 2D visualization, the ethane ratio and methane concentration were shown through the height of the bars, while in the 3D visualization, they were defined by the number of particles in the cubes. As described earlier, the wind speed was defined by the arrow length and the methane leak rate was noticeable through the speed and expansion volume of the particles. Each data element had three defined levels except for the wind direction which could point in eight directions. An overview of the different data levels can be seen in appendix B.2.

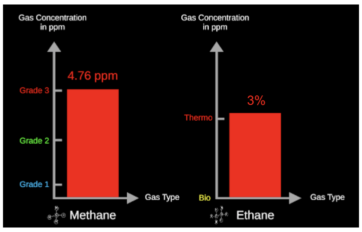
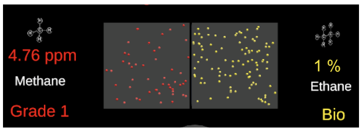
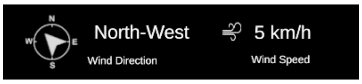
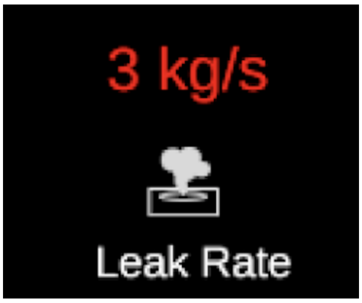
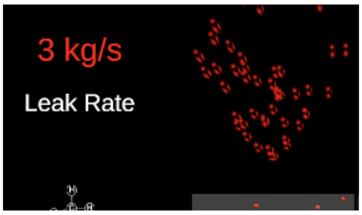
Data Type	2D Representation	3D Representation
methane concentration (ppm) and ethane ratio (%)	 statistical representation in the form of two bar charts	 physical representation in the form of a partial particle visualization in two volumetric boxes
wind direction and speed (km/h)	 an iconic compass icon which rotates based on the wind direction and an iconic wind icon to indicate the wind speed	 direct, flow visualization which depicts the air movement as an arrow to indicate the wind direction and the magnitude of the arrow as the wind force
methane leak rate (kg/s)	 iconic icon to indicate the leak rate	 physical representation in the form of a gas plume of leaking and expanding particles

Table 3.1: Data types and their corresponding representations in 2D and 3D.

Dialogues

A set of dialogues was designed to navigate the users through the MR experience, providing them with a tutorial for becoming familiar with the surveying context and showing them the SAGAT questions that popped up after each visible leak scenario (see figure 3.6).

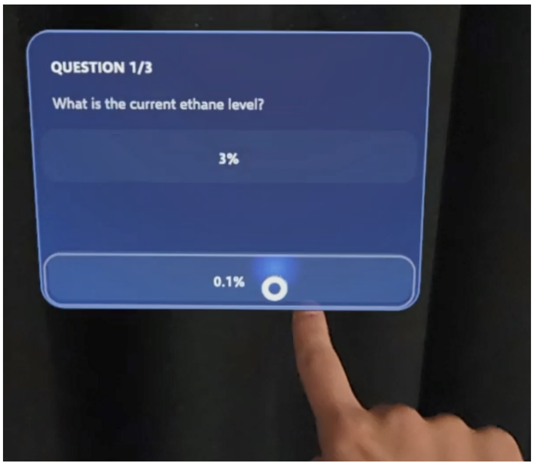


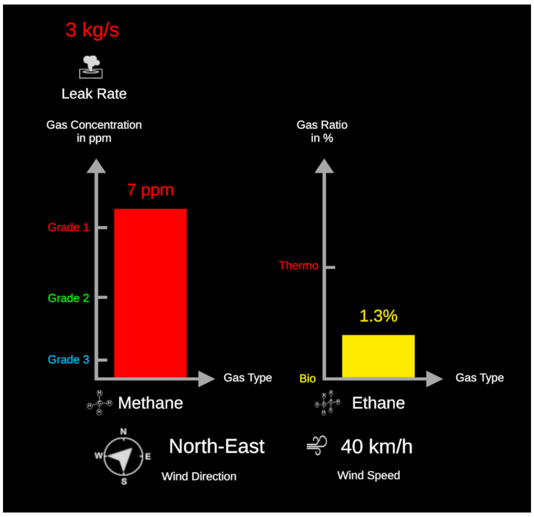
Figure 3.6: SAGAT dialogue window which displays a SAGAT question corresponding to the first level of SA and two possible answers. The space between the answers is large to avoid inaccurate responses.

3.3.3 HoloLens 2 Prototype

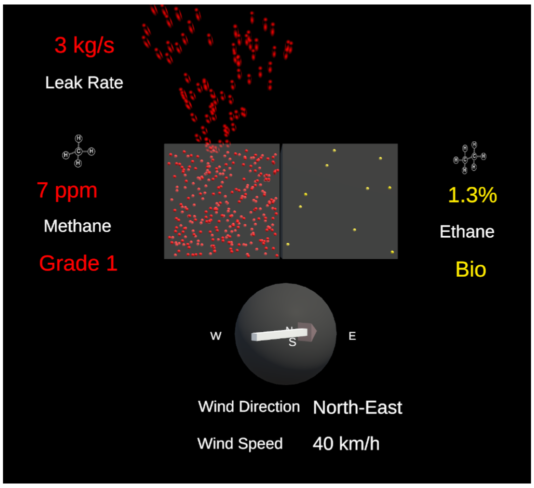
A MR prototype was built in Unity for the HoloLens 2 device which offers a diagonal 52 degrees field of view (FOV) [114]). The MR project used Unity version 2021.3.19f1, OpenXR 1.8, Visual Studio 2022 and the MRTK3 Microsoft toolkit for MR development. It can be viewed in figure 3.7.

The prototype was implemented based on the visualizations and dialogues designed in Figma (see section 3.3.2) and exhibited the whole experimental setup. Three JSON files were created to connect the SAGAT questions, which are listed in the appendix C.2, with the visualization data levels and participants since each question set belonged to a different leak scenario and each participant would have a different sequence of scenarios shown.

Each pressed answer was automatically logged into a text file which saved the participant id, visualization type, question id and the correctness of the chosen answer. The visualizations were shown at a fixed position in space and were purely static except for the 3D methane leak rate which had particles dynamically moving out of the box. Lastly, the only way of interacting with the dialogues was through near hand interaction as hand ray interaction resulted in unintentional triggers.



(a): Statistical, iconographic 2D visualization: It displays the leak rate at the top through colored text and an icon. Beneath it, two colored bar charts are shown which represent the methane concentration and ethane / methane ratio, accompanied by the grading of the gas on the y-axis. At the bottom, the wind direction and speed are illustrated by white text and a rotating compass and static wind stream icon.



(b): Physical 3D visualization: It depicts the leak rate through text and moving particles coming out of a box. Both methane concentration and ethane / methane ratio are displayed through the number and color of particles in two transparent boxes. The wind direction and speed are communicated through the direction and length of an arrow in a sphere as well as white text.

Figure 3.7: 2D (left) and 3D (right) visualization concepts developed in Unity for the HoloLens 2 device using the same layout.

HoloLens 2 Prototype Videos

Videos were made from the HoloLens visualizations showing the visualizations at the top of a stick which symbolized the real wand. The experience started from a central view, moving to the left and then the right. The videos were created through recordings from the Windows Device Portal which allowed the recording of the HoloLens screen and had a resolution of 3840 x 2160 pixels. Each video was edited through the video-editing software DaVinciResolve to shorten it to 20 seconds and make it usable for the surveys in section 3.4.3.

3.4 Phase 4: Evaluate against requirements

The forth phase encompassed three key activities. First, two pilot studies were conducted together with members of the UX project team to improve the current study design and make last changes on the visualizations. Then, an evaluation took place with two different research methods and two different population samples. On the one hand, a survey with embedded videos was sent out to participants having previous experience with surveying or knowledge about gas leak detection products. On the other hand, a usability study was conducted with non-experts who directly interacted with the MR prototype. As the subject-matter experts were not located at the same or close location to the MR device and could therefore not try the real experience themselves, videos embedded into a questionnaire were chosen as an alternative way which allowed them to perceive the visualizations in a limited way. To validate the developed surveying questions and correct possible mistakes and misconceptions, the first evaluation was done as a survey that was sent to an ABB employee with actual surveying experience who could provide feedback on the visualizations and questions.

In the following sections, the evaluation material, pilot studies and evaluation procedures will be presented.

3.4.1 Evaluation Material

The material used for both the survey and the usability study consisted in an introductory sheet of the surveying rules (see appendix C.1a), the SAGAT questions (see appendix C.2), a Microsoft Forms questionnaire which contained demographic questions about the gender, age, surveying and MR experience (see appendix C.3.1), the NASA-TLX (see appendix C.3.2) as well as the adjusted affective and conative environmental questions (see appendix C.1 and C.2). Finally, the evaluation also included a series of post-interview questions for both survey and usability study (see appendix C.4).

SAGAT questions

The SAGAT questions found in appendix C.2 consisted in six question sets which included one question per SA level with two possible answers. Each question set had one corresponding question set in which the opposite answers were correct, denoted with the characters a and b. In general, although the SAGAT questions reflected true intuitions of the surveying context and were checked by a surveyor before conducting the experiments (see section 3.4.3), they still represented simplifications of the real surveying work which required to consider many more parameters to assess the gravity of a leak. However, this limitation did not reduce the power of this research as it was aimed at exploring the domain of information visualizations and the surveying context just served as an application scenario. For the evaluation, the SAGAT questions were randomized for each participant during the survey and usability study evaluations. This meant that while the order of questions remained from sub-goal 1 to 3 in each evaluation, the order of a- and b-questions and the order of appearance for each visualization concept was randomized, leaving 2 x 2 potential combinations (2Da x 3Db, 2Db x 3Da, 3Da x 2Db, 3Db x 2Da). Since a-questions often displayed more serious leak scenarios, the letter a only denoted the first a-question so that the remaining questions alternated between a and b (e.g. a, b, a, ...) for one evaluation sequence.

Microsoft Forms questionnaire

Both the NASA-TLX and environmental awareness questions had a seven-point Likert scale format, ranging from ”Very Low” to ”Very High” for the NASA-TLX and from ”Strongly Disagree” to ”Strongly Agree” for the environmental awareness questions. While the original NASA-TLX embodied 21 points, a seven or ten-point Likert scale rating was also possible [25], [115] and preferred over the original size to render the questionnaire less lengthy and time-consuming.

Post-Interview questions

The post-interview questions consisted in questions about the understandability, mental demand, usefulness, awareness and potential improvements of each visualization concept as well as which visualization was more helpful in displaying a certain data type (see appendix C.4).

Survey

In the case of the survey-based evaluation, a questionnaire with embedded videos (see section 3.3.3) was attached by mail which showcased different surveying scenarios for both visualization concepts and asked the participant to answer the SAGAT questions corresponding to the displayed scenario. Within this questionnaire, two links were embedded to the other questionnaire which tested the cognitive load, affective and conative environmental awareness. Furthermore, to help the participant understand the shown scenarios in the videos, a PDF of the surveying rules was also attached to the mail.

Usability Study

The usability study was conducted with the real HoloLens prototype (see subsection 3.3.3). Moreover, as it was conducted with non-experts, a paper-based introduction into the surveying context was provided (see appendix C.1b) and the participants had access to a paper sheet of the surveying rules during the experiment. As for the survey, the study required participants to fill in twice the questionnaire with the NASA-TLX, affective and conative environmental awareness.

3.4.2 Pilot Study

In order to test the effectiveness of the usability study design and to correct any flaws in the visualizations, two pilot studies took place in an office space with UX research project members. Their participation resulted in some changes in the study material which will be explained below.

Concerning the visualizations, P1 was confused by the fact that both methane and ethane, which used very different scales, were initially presented in one bar chart, tempting the viewer to compare both levels. Although the bar chart was subsequently split for the pilot with P2, P2 still tried to relate both bar charts or gases with each other. Thus, to better separate the methane and ethane levels, thresholds were added to the y-axis to indicate that their heights and units had different meanings.

Another visualization element which was improved for the real usability study was the location of the visualization. As both 2D and 3D visualizations initially followed the user's head movement, P2 criticized this functionality as limiting for the visibility of the 3D visualization as it was always shown from the same angle and did not give the allure of a 3D effect any more. Rather than seeing it as attached to the screen, P2 proposed to either make it interactable or located at a fixed position in the real world. Moreover, according to P2, participants should have a wand prop to more clearly understand the underlying idea of a visualization shown at the tip of a wand even if it was not functional. As a consequence, the final design was spawned at a fixed position in space and a stick prop was given to the participants during the usability study. Moreover, to better visualize the direction of the 3D wind arrow, the color of the arrow tip was changed based on P2's feedback.

The study design itself was improved through introducing a progress bar and timer to indicate when the SAGAT question would be shown as P1 perceived a lack of feedback when the visualization was visible. Furthermore, as P1 had troubles choosing the right answer in the dialogue panel, the space between both answers was increased to improve input accuracy. P1 also mentioned that an introduction into the surveying rules should be done on paper instead of in MR as it was done initially, since reading was easier on paper than through the MR device. This was confirmed by P2 who, instead of accessing the surveying rules which could be toggled in MR, only used the paper sheet of the surveying rules to understand the data. Based on this observation and feedback, the previously developed grading tables in MR were completely omitted.

Finally, two adjustments were made to the Microsoft Forms questionnaire. First, the Likert-scale questions were changed into rows instead of squeezing them into columns as P1 criticized the format and readability of the questions. Second, the level of surveying experience was transferred from the post-interview to the questionnaire since P2 noted that it would allow an easier filtering of the data based on the level of experience.

3.4.3 Survey With Experts

The survey was sent out to all four previously interviewed experts. It was first sent to the expert with real surveying experience to get feedback on the self-defined surveying rules and SAGAT questions to check if the current intuitions and understanding reflected the real work experience and context. After this verification step, an improved version was sent out to the remaining three experts. In all cases, an email with the main questionnaire with the embedded visualization videos and a PDF of the surveying rules was sent out. To begin, the participant needed to go through the surveying rules for a couple of minutes until the information was clear, since new information like the methane leak rate was added which was not yet measured through the current gas leak detection solution. Then, the participant needed to fill in the main questionnaire which meant that they had to watch a randomized series of six leak scenarios for each visualization concept. After each video of a leak scenario which was 20 seconds long and needed to be watched twice, they had to answer three questions about it corresponding to the three levels of SA. Once the end of the series of six scenarios was reached, the participant had to fill in the supplementary questionnaire with the NASA-TLX and environmental awareness (affective, conative) questions so that it had to be filled in twice, one time per visualization concept. After that, a remote, semi-structured, audio-recorded interview was scheduled which took 10-20 min and allowed the gathering of more qualitative data about the participant's opinions of the visualizations.

Surveyor

A first survey was sent out to the expert who had free text fields for each question in the main questionnaire to be able to comment on flaws or mistakes in the defined questions. The videos could be re-watched by the expert surveyor as often as wanted because the focus lied on the improvement of the questions and the experience of the visualizations.

Other Experts

Unlike the surveyor's questionnaire, the survey for the other experts contained no free text fields to comment on the correctness of the questions and the participant's ability to re-watch the video was limited to two times (40 seconds) in order to reflect the visibility time in the usability study.

3.4.4 Usability Study With Non-Experts

A usability study in the form of a within-subject study was chosen as it required fewer participants [116] and allowed a comparison and discussion of both visualization concepts.

Participants

The participants of the usability study were ABB employees of the Corporate Research Center in Västerås, Sweden. As surveyors were difficult to reach for this physical evaluation, nearby non-experts were chosen instead and the evaluation included an introduction into the surveyor's work to get them familiar with the context. The participants were reached through a recruiting screener in which the only requirement for participation was defined to be that they were aware of the risks associated with wearing a HoloLens such as cyber-sickness and other forms of physical or mental discomfort [117].

Procedure

The usability study was conducted by the author in a meeting room in the ABB Corporate Research Center in Västerås, Sweden.

At the beginning, the participant received a short, verbal briefing about the study procedure as well as an information letter, a consent form and the surveying information in the form of four paper sheets where one sheet included the surveying rules that was used during the experiment. Since the participants had no or only a limited amount of knowledge about the surveying context, an introductory lecture about the surveyor's work and important terminology was given to ensure that they could understand the scenarios that were shown in the MR experience. After the paper-based introduction, the participant put on the HoloLens 2 in which they saw an introduction that described them as a surveyor who would go on two surveys. The MR introduction entailed a training session with two leak scenarios, each displaying a textual visualization and requiring them to answer a sample question in order to help them become familiar with the context and HoloLens device. A familiarization phase with the HoloLens was for example done in Bagher et al. [25].

The experience with the HoloLens 2 was video-recorded through the web-based Windows Device Portal which collected both screen-recording data of what the participant could see through the HoloLens 2 as well as audio-data of comments made during the experiment. Each survey was accompanied by one visualization concept and comprised six leak scenarios, each lasting 40 seconds, in which a data visualization was shown at a fixed position in the experiment room at shoulder height and the participant had to answer three questions about it, relating to the three levels of SA. The order of the first shown visualization concept as well as the corresponding SAGAT questions were randomized for each participant (see section 3.4.1). To help the participants imagine the visualization being shown at the tip of a surveying wand, they could hold a physical stick prop (see appendix C.1). After encountering the six leak scenarios, the participant had to put aside the MR device and fill in an online questionnaire about their demographics, cognitive load as well as affective and conative environmental awareness. After this step, they would go on their second survey with the second visualization concept which had the same structure, including the completion of the same online questionnaire.

At the end of both surveys, a semi-structured post-interview took place which lasted 10-20 min and was audio-recorded through the dictation mode of the web version of Microsoft Word.

Chapter 4

Evaluation

4.1 Results

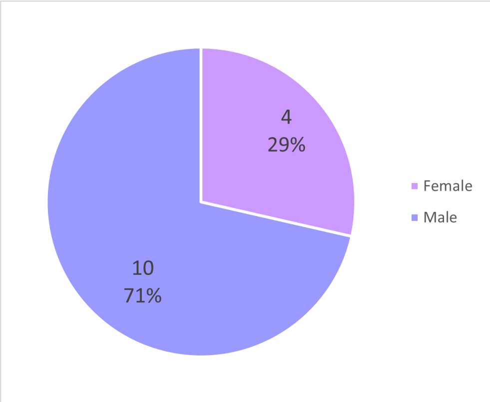
In this chapter, the results of the evaluation will be presented, starting with the demographic distribution of the participants, the statistical analysis of the answers and questionnaire data and ending with the themes discovered through the post-interviews.

4.1.1 Participants

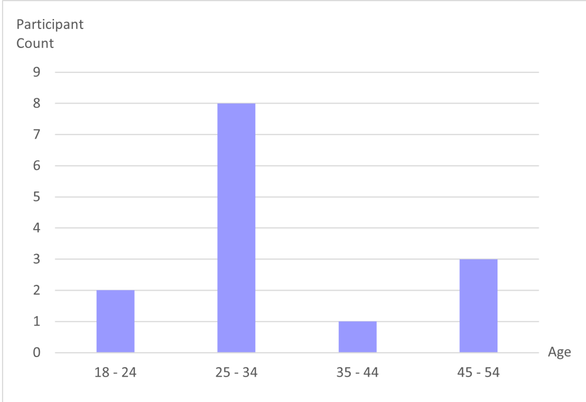
The evaluation was performed with 14 participants, four experts who saw the survey with embedded videos and ten non-experts who interacted with the prototype during a usability study. From these 14 participants, ten identified as male, four as female and their ages ranged from 18-24 to 45-54 with most being between the ages 25-34 (see figure 4.1). Concerning their level of experience with MR, half of the participants had no previous experience with it, 43 % had tried a MR device once or a couple of times before and only one participant was experienced in the use of MR technology (see figure 4.2). What concerned their level of surveying experience, 64 % of the participants had no experience with gas leakage surveying, 29 % were familiar with the context and surveying products while one participant had actual experience in surveying. Among the ones who were familiar with the context, one identified as the project lead for a mobile surveying application while the other participant listed experience in supervising leak surveyors, completing leak surveys with surveyors and having training as an operator (see figure 4.3).

4.1.2 Statistical Analysis

For the statistical analysis, the leak scenario answers from the SAGAT method and the Likert scale results from the questionnaires were analyzed. To be able to ana-



(a): Gender



(b): Age

Figure 4.1: Gender (left) and age (right). Most participants were male and between 25-34 years old.

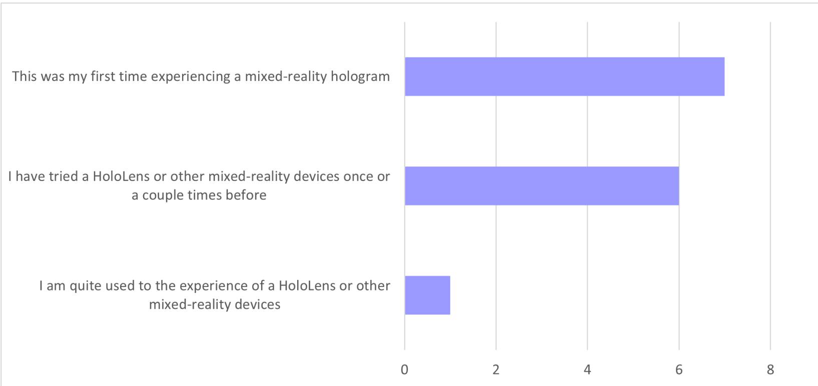


Figure 4.2: MR experience level, showing that almost an equal amount of participants had either no experience with MR or tried a MR device once or a couple of times before.

lyze the responses to the SAGAT questions, invalid answers had to be removed first. For instance, the grading levels on the y-axis in the 2D visualization turned out to be reversed during the first experiments which is the reason why the SAGAT answers concerning the methane grade were excluded from the final calculations (see appendix C.4 sub-goal 2, question: 2). The screen-recordings helped understanding when the participant accidentally pressed the wrong answer which happened twice and was then manually corrected. After the data cleaning phase, a frequency analysis was performed by counting the number of errors for each visualization concept and then calculating the corresponding error rate or percentage of incorrect answers for each condition (2D, 3D) [37]. The error rates were then also calculated in light of other conditions such as expert vs. non-expert, which visualization was started with and the different levels of SA (see table 4.2).

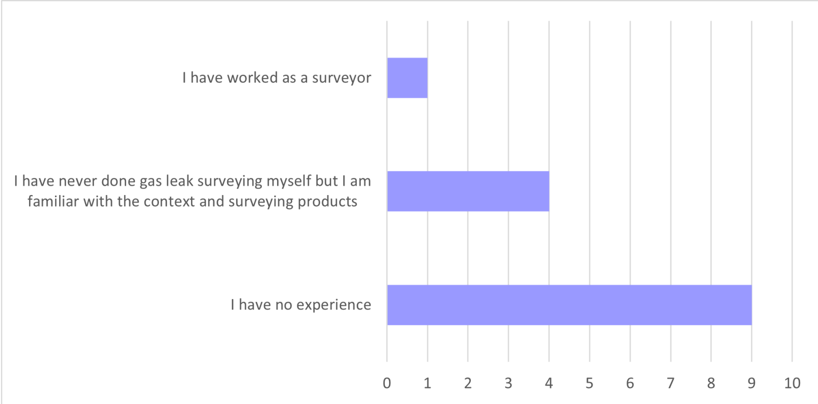


Figure 4.3: Surveying experience level indicating that most participants had no experience in surveying, while only one person had actual surveying experience.

For the Likert scale results, the data was first coded into numeric values and then reverse coded for five questions which had the opposite meaning. As Likert scale data is ordinal data for which the distance between two points is unknown, only the median could be calculated for each statement and concept [118]. Moreover, while the mean was not applicable, the standard deviation was still calculated to provide a sense of the variability or spread of the data [119]. Then, a two-tailed non-parametric Wilcoxon Signed-Rank test was performed which can compare the difference of median ranks between two paired groups [120] and can be seen as a non-parametric version of a paired t-test [121]. To understand if there is a significant difference between two groups, the Wilcoxon-Signed Rank test first calculates the difference between two groups for each measurement. Then, a rank is calculated for each difference, taking into account negative differences with a negative rank (-1) and tied ranks for the same differences with a rank's average. Once each difference has an assigned rank, both positive and negative ranks are summed up. Lastly, the Wilcoxon (W) test-statistic which is represented by the smallest sum is compared to a critical value (t-critical) defined by the test. In case the value for W is smaller than the critical value, the difference is statistically significant (t_{sig} = t-statistic $<$ t-critical) [121].

A non-parametric test was chosen because they are suited for small sample sizes [122] and because the sample size of 14 was lower than the sample size of 30 for which a normal distribution can be assumed according to the central limit theorem [123]. Moreover, as Likert scales produce only ordinal data and not interval data for which a normal distribution could be calculated, no normal distribution could be assumed, excluding parametric tests such as the t-test [118]. In addition, to not only assess the statistical significance, i.e. if the results were created by chance, but also understand the magnitude of the relationship between a visualization concept (2D, 3D) and examined concept (cognitive load, environmental awareness), the effect size was also provided which has meaning independent of the used sample size [124]. As no normal distribution was assumed, the formula $r = z/\sqrt(n)$ provided by Rosenthal [125] was used to calculate the effect size, using Cohen's [126] categorization of small (r = 0.1 - 0.3), medium (r = 0.3 - 0.5) and large (r $>$ 0.5) effects. In this calculation, the 3D visualization was treated as the pre-condition, while the 2D visualization was treated as the post-condition.

Lastly, as this evaluation was performed with two different sample groups (expert vs. non-expert) or formats (survey vs. usability study), the medians and standard deviations were also calculated for each group / format. Differences that exceeded at least one score were included in the discussion below. Since the overall sample size was low and the sample size difference between both groups was high, no attempt to calculate the significance was made and the results should be interpreted rather as indications worth investing than clear significant results.

Environmental Awareness

Cognitive Environmental Awareness The analysis of the SAGAT answers revealed a mean error rate of 10.5 % with 2D having a higher error rate (11.76 %) than 3D (9.24 %). While this difference was not significant (t_{sig} = 24 $<$ 17, α = 0.05), a small effect (r = 0.24) could be found (see table 4.1). Looking more closely at the target group, it can be stated that the error rate was on average higher for the expert group (12.5 %) than the non-expert group (10.26 %) and that they made more mistakes with the 2D type (17.65 %) than the 3D one (7.35 %). On the other hand, the non-expert group struggled a bit more with the 3D visualization (11.11 %) than the 2D one (9.41 %). Since the order of the visualization concepts might have affected the error rate, it was also interesting to see how the error rates differed based on which visualization was shown first. The results indicated that when the participant started with the 3D concept, they had a higher mean error rate (13.45 %) than when starting with the 2D one (7.98 %). Interestingly, the visualization concept the participants started with had in both cases a lower error rate than the second one. When the participant started with the 2D visualization, their error rate was 7.56 % for 2D and 8.4 % for 3D. Similarly, when the participant started with the 3D visualization, their error rate was 10.92 % for 3D and 15.97 % for 2D. The results can be seen in table 4.2.

To gain insights into which types of situational knowledge were difficult to retain, the error rates per question type were also calculated. In figure 4.4, it can be seen that the wind direction and speed questions resulted in the highest error rates (42.86 % and 38.1 %), then came the methane prioritization question (21.43 %), the ethane

α	0.05
n	11
t-critical	17
t-statistic	24
significance	no
r (small effect)	0.24

Table 4.1: Wilcoxon-Signed-Rank test and effect size: SAGAT error rate was not significant, but a small effect could be found.

level question (9.52 %), the methane safety question (7.14 %) and lastly the leak rate question with no errors at all. The number in parentheses indicates the order of appearance, suggesting that the number of errors was higher in earlier than later questions. Taking into consideration the error rates per SA level, it could be seen that SA level 1 questions resulted in the lowest average error rate (4.17 %), while SA level 2 had a 13.1 % and SA level 3 a 12.5 % average error rate. Finally, one thing to note was that the SA level 3 error rate was much higher for 2D (15.48 %) than for 3D (9.52 %) (see table 4.2).

	2D	3D	Mean
Total Error Rate (%)	**11.76**	9.24	10.5
Expert Error Rate (%)	**17.65**	7.35	**12.5**
Non-Expert Error Rate (%)	9.41	**11.11**	10.26
Start With 2D (%)	7.56	**8.4**	7.98
Start With 3D (%)	**15.97**	10.92	**13.45**
SA Level 1	**4.76**	3.57	4.17
SA Level 2	13.1	13.1	**13.1**
SA Level 3	**15.48**	9.52	12.5

Table 4.2: SAGAT error rates in % for all samples, per role (expert, non-expert), per first visualization type (2D, 3D) and SA level.

Affective Environmental Awareness For the affective environmental awareness, the median of the 2D visualization type was a 6 on a 7-point Likert scale (σ = 1.27) and a 5.5 for the 3D one (σ = 1.27) which both indicate a farely high emotional awareness of the environmental impact of a leak (see figure 4.5). However, the result from the Wilcoxon Signed-Rank test suggested that these outcomes were not significant (t_{sig} = 213.5 $<$ 137, α = 0.05) (see table 4.3a). Calculating the effect size, there was also no effect (r = -0.07).

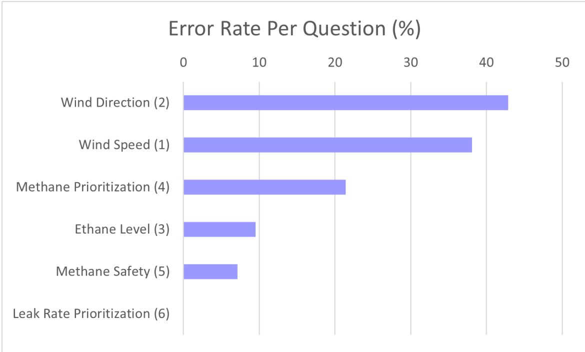


Figure 4.4: SAGAT error rate per question type in %, sorted in descending order with the order of appearance in parentheses. The wind questions had the highest error rates. Moreover, earlier questions had a higher error rate than later questions.

Although there was no large difference in terms of affective environmental awareness for both visualization concepts, there could still be a difference between experts and non-experts or survey vs. usability study. Comparing the medians between both experts and non-experts, it can be seen that experts rated their affective environmental awareness at 6 (σ = 1,62 (2D), σ = 1,62 (3D)) for both visualization types, while non-experts rated theirs at 5 (σ = 1,1 (2D), σ = 1,1 (3D)) (see figure 4.7). There were also differences between the individual statements that were used to measure the affective component. Non-experts were more frightened of how much methane is emitted to the atmosphere for 2D (median = 6, σ = 0.83) and 3D (median = 5.5, σ = 0.87) than experts (median = 4.5 (2D, 3D), σ = 1.79) (see figure 4.8). On the other hand, experts became more incensed when thinking about the harm being done to the climate by leaking gas for both visualization types. Here, the median results for experts were 6.5 for 2D (σ = 2.06) and 6 for 3D (σ = 1.5), while the ones for non-experts were 5 for both 2D and 3D (σ = 0.8) (see figure 4.9). Lastly, experts also disagreed more with rarely worrying about the effect of methane emissions on themselves and their families for the 2D visualization concept (median = 2, σ = 0.43) in comparison to non-experts (median = 3.5, σ = 1.56) (see figure 4.10).

Conative Environmental Awareness The 2D (σ = 0.98) and 3D (σ = 1.25) visualization concepts both resulted in a 6 out of 7 for the conative environmental awareness, indicating a high willigness to act and reduce the environmental impact of leaks (see figure 4.6). As for the affective environmental awareness, the results from the Wilcoxon Signed-Rank test suggested no significant difference between 2D and 3D (t_{sig} = 263 $<$ 235, α = 0.05) (see table 4.3b). Looking however at the effect size, a small effect (r = -0.25) could be found.

A comparison between the medians of experts vs. non-experts revealed that

α	0.05
n	30
t-critical	137
t-statistic	213.5
significance	no
r (no effect)	-0.07

(a): Affective

α	0.05
n	38
t-critical	235
t-statistic	263
significance	no
r (small effect)	-0.25

(b): Conative

Table 4.3: Wilcoxon-Signed-Rank test and effect size of the affective (left) and conative (right) environmental awareness. Both turned out to have no significant difference in 2D and 3D. Althought the affective dimension had no effect, a small effect could be found for the conative dimension.

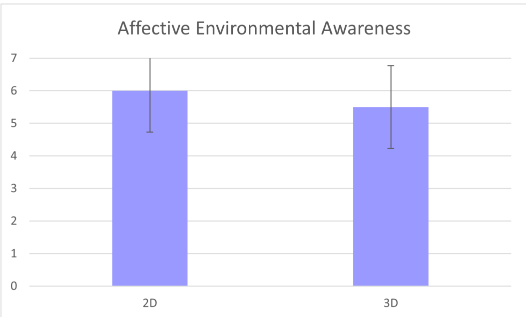


Figure 4.5: Affective environmental awareness was slighter higher for the 2D visualization.

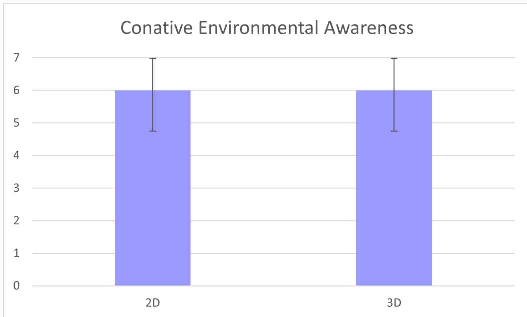


Figure 4.6: Conative environmental awareness was the same for both 2D and 3D visualizations.

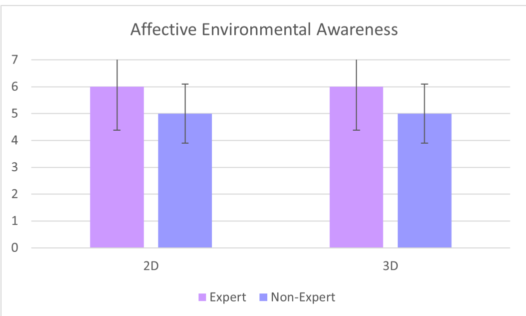


Figure 4.7: Experts had a higher affective environmental awareness for both 2D and 3D visualizations.

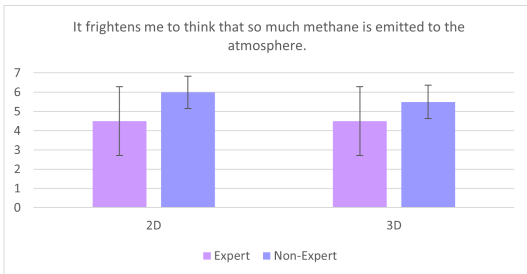


Figure 4.8: Non-experts were more frightened about methane emissions for both 2D and 3D visualizations.

experts had a higher conative environmental awareness of 6 (σ = 1.09) than non-experts who score a 5 (σ = 0.92) for the 2D visualization type (see figure 4.11). Moreover, experts showed a higher willingness to diligently do their job to reduce methane emissions for both visualization types. For the 2D type, experts had a 6.5

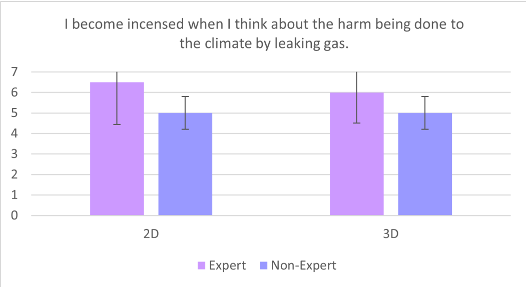


Figure 4.9: Experts became more incensed about the harm of leaking gas for both 2D and 3D visualizations

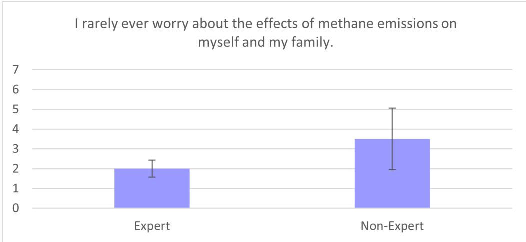


Figure 4.10: Experts worried more about the effect of methane emissions on themselves and family.

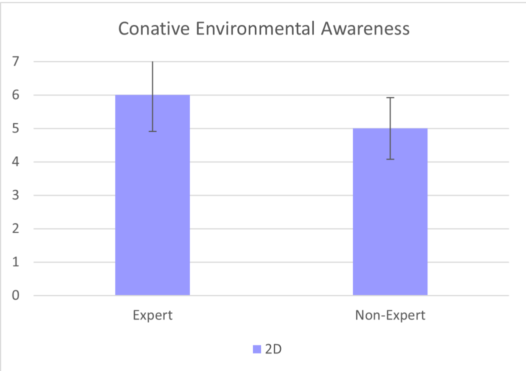


Figure 4.11: Experts had a higher conative environmental awareness for the 2D visualization.

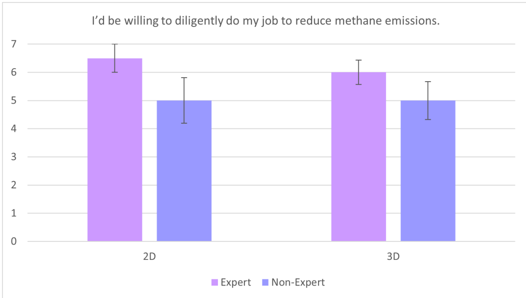


Figure 4.12: Experts were more willing to dilligently do their work for both 2D and 3D visualizations.

median (σ = 0.5) in comparison to non-experts who had a median of 5 (σ = 0.8). For the 3D type, experts scored a median of 6 (σ = 0.43) and non-experts a median of 5 (σ = 0.67) (see figure 4.12).

Cognitive Load

The cognitive load was at 3 on a 7-point Likert scale for both 2D (σ = 1.4) and 3D (σ = 1.54) visualizations, indicating a low to medium mental load (see figure 4.13). The analysis of the Wilcoxon Signed-rank test suggested that the cognitive load results were not significant (t_{sig} = 469 $<$ 343, α = 0.05) and had no effect (r = -0.08) (see table 4.4).

Since cognitive load is a concept which can be divided into six sub-dimensions, there was still a chance that the sub-dimensions could have some significant effects which will be presented in the following. The mental demand component was not significant (t_{sig} = 10 $<$ 0, α = 0.05) and had no effect (r = -0.04). The physical demand component turned out to be not significant as well (t_{sig} = 6 $<$ 0, α = 0.05), however a

α	0.05
n	45
t-critical	343
t-statistic	469
significance	no
r (no effect)	-0.08

Table 4.4: Wilcoxon-Signed-Rank test and effect size for the NASA-TLX results which had no significant effect.

small effect could be found (r = -0.18) (see table 4.14). The pace was significant (t_{sig} = 2 $<$ 17, α = 0.05) and had a large effect size (r = -0.77) for 2D (median = 3) and 3D (median 3.5). The success component also turned out to be significant for both visualization types (t_{sig} = 12 $<$ 17, α = 0.05) and showed a small effect (r = -0.13) for 2D and 3D (median = 3) (see table 4.16). No significant results (t_{sig} = 39 $<$ 13, α = 0.05) and no effect (r = 0) could be found for the effort component. Lastly, while the stress dimension had no significant difference (t_{sig} = 26 $<$ 17, α = 0.05), there was a medium effect (r = 0.4) (see table 4.16).

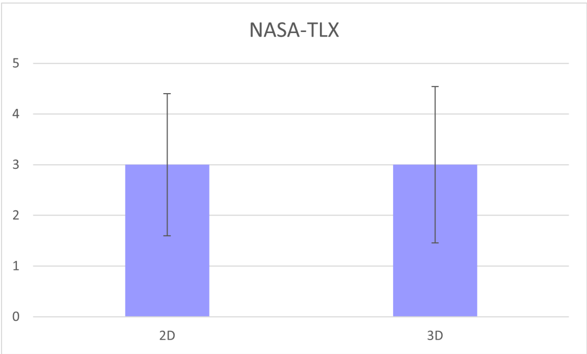


Figure 4.13: Cognitive load measured through NASA-TLX was the same for both 2D and 3D visualizations.

Although there was no difference in medians between 2D and 3D, a comparison of medians between experts and non-experts revealed that for 3D, experts had a lower cognitive load (median = 3, σ = 1.11) in comparison to non-experts (median = 4, σ = 1.64) (see figure 4.17). There were also differences of at least one score for three of the six cognitive load sub-dimensions. The mental demand of experts was lower for both 2D and 3D (median = 3, σ = 1.09 (2D), σ = 0.433 (3D)), while the mental demand for non-experts was much higher for both 2D and 3D (median = 5, σ = 1.35 (2D), σ = 1.57 (3D)) (see figure 4.18). Experts also felt more successful in accomplishing what they were asked to do for the 3D visualization (median = 6, σ = 0.433) in comparison to non-experts (median = 4.5, σ = 1.28) (see figure 4.19).

α	0.05
n	6
t-critical	0
t-statistic	10
significance	no
r (no effect)	-0.04

(a): Mental demand

α	0.10
n	5
t-critical	0
t-statistic	6
significance	no
r (small effect)	-0.18

(b): Physical demand

Figure 4.14: Wilcoxon-Signed-Rank test and effect size for the mental and physical demand dimensions of the NASA-TLX which both turned out to be not significant. Though the mental demand had no effect, the physical demand displayed a small effect.

α	0.05
n	7
t-critical	17
t-statistic	2
significance	yes
r (large effect)	-0.77

(a): Pace

α	0.05
n	7
t-critical	17
t-statistic	12
significance	yes
r (small effect)	-0.13

(b): Success

Figure 4.15: Wilcoxon-Signed-Rank test and effect size for the pace and success dimensions of the NASA-TLX. Both dimensions had significant differences between 2D and 3D and displayed a large effect for pace and a small effect for success.

α	0.05
n	12
t-critical	13
t-statistic	39
significance	no
r (no effect)	0

(a): Effort

α	0.05
n	8
t-critical	17
t-statistic	26
significance	no
r (medium effect)	0.4

(b): Stress

Figure 4.16: Wilcoxon-Signed-Rank test and effect size for the effort and success dimensions of the NASA-TLX which both were not significant. While the effort component had no effect, the stress component displayed a medium effect.

The perceived effort of the experts was also lower (median = 3, σ = 1.09) for both visualization types, while the effort of non-experts was lower for 2D (median = 4, σ

= 1.27) than for 3D (median = 5, σ = 0.36) (see figure 4.20).

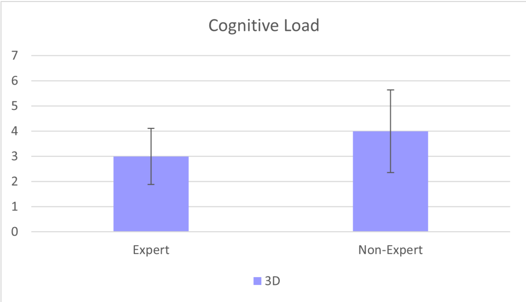


Figure 4.17: Experts had a lower cognitive load for the 2D visualization concept.

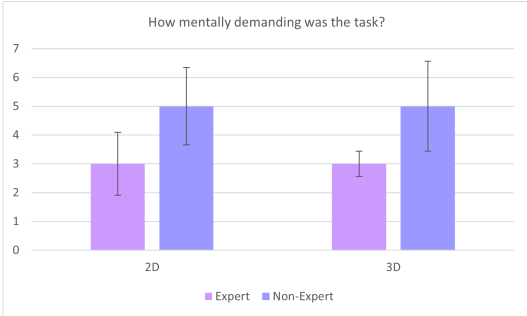


Figure 4.18: Experts had a lower mental demand for both 2D and 3D visualizations.

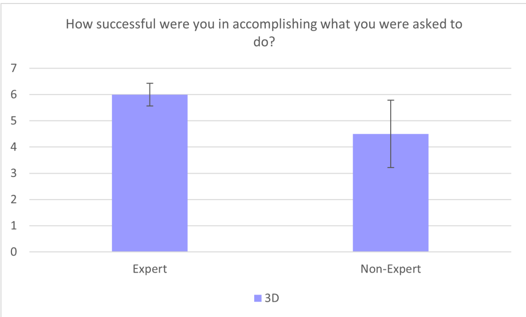


Figure 4.19: Experts felt more successful in accomplishing their tasks for the 2D visualization concept.

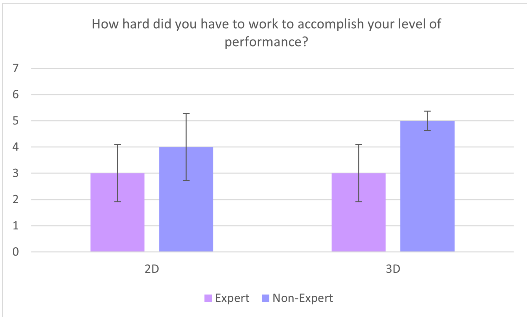


Figure 4.20: Experts experienced a lower effort for both 2D and 3D visualizations.

Since the order of what visualization concept was shown first could have affected the perceived cognitive load due to learning effects or progressing tiredness, the results which started with a certain visualization concept were also compared to all the results of a visualization concept. Comparing the results which started with 2D with all 2D visualization results showed no difference in scores (both median = 3). However, a comparison of the mental demand sub-dimension results revealed that when participants started with the 2D visualization, their cognitive load was lower (median = 3, σ = 1.36) than when being the first or second visualization (median = 4, σ = 1.36). The opposite relationship could be found for the 3D visualization. When the participants started with the 3D type, their cognitive load was higher (median = 5, σ = 1.48) than for all 3D visualization data (median = 4, σ = 1.56) (see table 4.21).

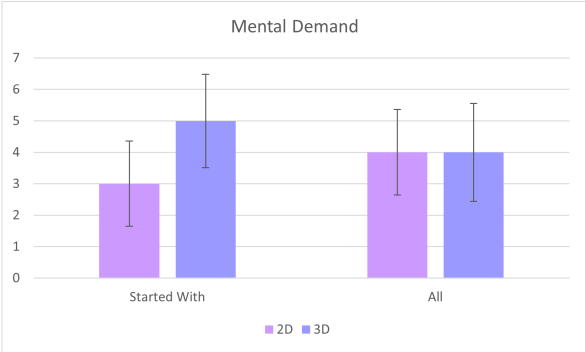


Figure 4.21: Mental demand was lower when the experiment started with the 2D visualization and higher when it started with the 3D visualization in comparison to the mental demand results of the same visualization concept.

4.1.3 Qualitative Data

The qualitative data was analyzed through a thematic analysis which is a method that generates codes from data to form overall patterns or themes [104]. In this analysis, both deductive top-down and inductive bottom-up approaches [104] were combined, starting with the investigated concepts of environmental awareness and cognitive load and then finding new themes that could be related to them. The analysis was performed in the qualitative research tool MAXQDA. To better understand the answers based on the surveying experience level, different letters were used for the participants with S = Surveyor, B = Business person (expert), P = Participant (non-expert). In addition, the screen recordings of the HoloLens interactions were analyzed by noting down any procedural errors or deviations from the general procedure such as an invisible visualization at the beginning of the experiment and any comments made that could enrich the post-interview data.

In the following sections, themes that emerged for both concepts will be discussed. As cognitive environmental awareness and cognitive load were related concepts that influenced each other, i.e. a high cognitive load could decrease the SA understanding [127], the themes regarding both concepts were put together.

Cognitive Environmental Awareness and Load

Perspective Visibility To be able to understand the presented leak situation, it was necessary to first perceive or see the provided visualization. Visibility at a certain distance was mentioned by P3 as a criterion to choose between the 2D and 3D visualization, as the thick 2D bars might be easier to notice than small 3D particles when the visualization would be shown on a wand:

"If it's further away [on a stick], it's harder to notice the details like the leaking particles [in 3D] even with color coding, so thick bars that are like fully red or fully green would be easier to notice." (P3)

Furthermore, while the 2D visualization was fully visible from one angle, participants (P1, B1, B3) complained that they struggled to fully see the 3D wind arrow from their perspective, suggesting to shift the arrow by 20-30 degrees (P1) or changing the arrow size instead of the arrow length since a low wind speed made it hard to see the direction (B2). Other participants solved this problem by actually switching perspectives through movements around the visualization (P6, P8). For instance, P8 noticed that while moving around the visualization, it "still pointed north" (P6) and therefore acted "like a compass" (P8).

Spatial Mapping Besides the perspective visibility, the 3D visualization of the wind also seemed to encourage participants to relate the arrow to their bodies. While the 2D compass icon was only associated with the wind direction, the 3D wind arrow also suggested to be a navigational recommendation of where they should go (P3, B3) which B3 declared as "misleading".

In general, some participants noted that the wind direction and speed were the hardest parts to understand (P1, P3, P6) and therefore the most cognitively demanding. For instance, P3 claimed to lack contextual knowledge to understand the "meaning" of what P3 saw. This signified that although P3 could perceive the information (SA level 1), P3 was unable to draw conclusions about the location of the leak (SA level 2) and appropriate next actions (SA level 3). As a result, multiple participants (P6, B2, B3) recommended to present navigational hints of where to go:

"Maybe another solution would be better for me, like you should go this, maybe only an arrow showing which direction I should go. So not that I need to read the compass and then make a decision. Should I go towards the wind. Or against the wind or orthogonal. So maybe just a quick feedback you should go this way from my point of view." (P6)

While B3 thought of replacing the wind direction information with the navigation information, B2 recommended instead to combine multiple perspectives, e.g. three arrows in 3D: The physical wind direction, the human head direction and the navigational hint direction. Moreover, B3 proposed a sort of virtual scouting functionality like in Google Street where a human could virtually walk a street beforehand to understand where to go.

Another way of aiding the understanding of the wind direction in 3D were proposed by P7 and B3 who suggested to use the plume and its blowing direction as an indirect marker to understand if one was walking towards the leak or not. This would mean that multiple elements in the visualization could encode redundant information to ensure that difficult information could be interpreted correctly through different visual cues. Moreover, the 3D visualization could rely on the physical understanding and relation of elements in the real world.

While most participants treated the 2D visualization like a typical user interface, P1 also tried to map it spatially to the environment as it was done with the 3D visualization and then was confused by the fact that the 2D compass icon was laid out vertically, pointing to the sky, instead of horizontally like a real compass. Whereas the 2D information of the wind first needed to be translated into the 3D space, P3 and B3 praised the 3D wind arrow as it allowed to "just see which direction the wind is going" (P3):

> "Wind direction was much easier in 3D, because, like AR is, in 3D, it's easier to kind of map that information directly. So having a 2D like overview, because what it's basically, it's talking about directions that you're kind of standing in, right in 3D. But then you have to mentally map." (P3)

One danger of the spatial mapping existed which concerned the layout of the visualization. For instance, B3 was confused by the 3D layout, believing that the layout mirrored some physical reality:

> "So spatially, I thought it was a spatial distribution where you have something on the right and something on the left. So since it's distributed in this space, I think you expect that too [methane on the left and ethane on the right]." (B3)

Attention Another aspect which influenced the SA of the participants in both 2D and 3D scenarios was the question of what drew their attention and what, as a consequence, lost their attention. For instance, in the 2D visualization, P3's attention was directly caught by the bar chart and then went straight downwards, missing the leak rate which was positioned above it:

> "Because, like the bar charts, if there were big, like big red ones, they immediately caught my attention. And then, I naturally went down from there. But then, I had to go up for the leak speed about particles." (P3)

To overcome this problem, some participants (P3, P5, P6) suggested a different layout for both 2D and 3D visualizations which should adhere to the natural, human reading direction from top left to bottom right and avoid putting the most attention grabbing element in the center.

Another aspect which was mentally demanding for some participants was the number of elements to keep track of (B3, P9, P10) which was more difficult to handle in the 2D visualization as there were ”many icons and many colors”(2D), while P3 found the 3D visualization easier to grasp as it seemed more ”compact”. To mitigate the problem of too many items on a small space, P10 recommended to increase the spacing of elements as AR was not limited by the confines of a traditional screen but instead could utilize the whole physical space:

> ”Lot of information to handle at the same time. So if it could be split up in some way. [...] Its augmented reality, and I can move around my head like this. There's so much space. The space can be used. So no need to make it all compact.” (P10)

Besides the effect of the information layout, the color coding also influenced the attention direction of the viewer and created more cognitive load by overstraining the visual eye. For instance, while the red color immediately triggered the sense that it was ”dangerous” (P4), P10 failed to notice and remember the ”white text” for the wind direction in the 2D and 3D visualizations as the ”eyes were only looking at the ”red”, making it ”hard [...] to get away from that”. On the other hand, when the concentrations were low, when there were ”only like a few dots” and the colors were blue and green, P10 found it easier to ”process” and was able to see the other information as well. Finally, especially for the 2D visualization, the bars appeared to be more ”striking” (P9) because of their size or width (P9, B3) and solid color (S).

In addition to a spatial focus, P4 also mentioned the possibility of a temporal focus, comparing looking at the visualizations to watching a movie that allows to discover new elements every time, it would be rewatched:

> ”Because when you see something in the first time, you are so focused on seeing what, what it's showing. [...] It's like seeing a movie sometimes. OK, when you look at it the second time or the third time, it's even get better and better and think it's even more funny or not funnier.” (P4)

Relatability One theme which also emerged from the data was the concept of relatability. For instance, B1 described the 3D visualization with the ground and the emitting particles as something that B1 was not ”traditionally used to” and that it was ”not a graph anymore” but just an ”illustration” lacking the realism of the physical world:

> ”It [3D] didn’t bring any value for me because, if you take all, if you just take the illustration out and just see the numbers, it’s telling me the same thing cause the illusion, like the illustration, doesn’t mean there’s only four dots there.” (B1)

As a consequence, P6 struggled to understand ”what the boxes were” at the beginning of the experiment but then realized that it represented the leakage when the concentration was high. While the 2D visualization seemed more like a traditional user interface, like ”charts or Excel” that P5 was used to, the increased realism of the 3D visualization could, for some participants, in fact help in understanding the data better by being more ”intuitive” (P3) and close to the real world (P3). In terms of MR, P5 felt like the 2D visualization would be ”an interface that would work on the screen as well as in AR”, while the 3D visualization would be suitable only for AR / MR because it would be ”harder to view on the screen”.

In the 2D visualization, especially the 2D wind icon was explained as being more familiar, reminding of a traditional compass (P8, B1) that allowed to ”just see the icon” and then knowing what it was about (B3). P6, however, also warned that the relatability with an icon depended on the knowledge of the object the icon symbolized, stating that compasses were not used any more for a couple of years and a simple arrow might be more understandable to younger generations.

There was also another form of relatability which regarded more the context, i.e. the surveying context. Some participants simply claimed that they were not ”familiar” with the context (P3, P4). For instance, P3 noted a lack of experience in interpreting the methane leak rate in terms of SA level 2 and 3:

> ”So, if it’s slow there is no repair needed, even though the visualization showed very clearly that it was slow, but it’s more like, what does it mean?” (P3)

On the other hand, being familiar with the context allowed to easier ”click” (B1) with both visualizations:

> ”It took me a second to understand what it was trying to tell me, but because I’m familiar with emissions and how pipelines work and what’s important and what mobile guard [mobile surveying app] reports out, it was, I was able to click.” (B1)

Being more familiar with the context then also indicated a stronger focus (B1) and need for numbers e.g. on the y-axis of the 2D bar chart (B2), while P7 claimed that the 3D visualization was easier for P7 as a non-expert.

Categorization An often stated difference between the 2D and 3D visualizations was the perceivability of thresholds or data levels for the methane concentration and ethane ratio. In the 2D visualization, some participants praised the bar chart's grading labels on the y-axis that allowed them to easily understand the thresholds based on the bar's height (P4, P6, P8, P10, B1, B2), showing "how close [you are] to the kind of breaking point" (B1). However in the 3D visualization, it was criticized that the 3D boxes only showed "lots or less or even less" (P3), having only the color coding and accompanying text to perceive the grade.

"The first one I saw, the bars, and immediately could decide if it was grade three or two or one, thanks to the bars, while in this one, I couldn't really take a decision on 'now it's 100 % of the particles inside the box and now it's only 10 %'. I couldn't take that decision in the first place." (P6)

Not only were these categories harder to perceive in the 3D visualization, but also the proportion between methane and ethane was harder to understand in it. While P3 found the 3D boxes more intuitive and less demanding because P3 was able to connect a low ratio with fewer dots and a higher ratio with more dots, the 3D boxes lacked accuracy by telling what exactly the ratio was (P5):

"Like the bar charts are easy 'cause you can like see like oh, this one is double the amount of that one next to it, but it's hard to see when it's like the concentration of these atoms in the box, it's hard to see it like that one is double that one. You get an idea, but you don't get like the exact, but you you get the numbers, so you can watch them." (P5)

To get the same sense of a threshold in 3D as in 2D, P6 therefore suggested to "fill [the boxes] up" like liquid to communicate the fullness of the boxes.

While the thresholds could be perceived more easily through the 2D charts, the understanding of the categories was however more mentally demanding according to some participants as it required "reading" the charts (P7). Moreover, according to B3, it was "less interesting to have to read the axis" in VR / AR in order to understand if the bars were high or low:

"It [3D] was not mentally demanding compared to the other one. Because most of the things were already displayed there, I mean the grade or biogenic or thermogenic and the rate. So there, it will already classify and, the other one, you should have a look at the chart and see if the level is higher or not". (P7)

Environmental Awareness - Affective

Realism One theme that emerged was that the affective awareness was increased by the realism of the 3D visualization (e.g. S, P3, P6, B2, B3). While the 2D visualization was associated with terms such as ”information” (P7, B2), ”numbers” (P3), ”simplistic” (P2) and ”stats” (P6), the 3D visualization was rather described as a ”picture” (P6) of the real world:

”Because you saw all the particles going away, and I think that made me think more about it: That they actually leak and that it affects the environment. While the first one was more static and more like, let’s say information or stats, numbers showing something, and while the other one showed a picture.” (P6)

More concretely, the 3D visualization created more awareness because the leak rate particles had ”some movement to it” (S) and were ”spilling out from the top [...] (looking) very bad” (P5), indicating that ”there’s something in the air” (B3). Furthermore, a large number of dots in the 3D boxes gave a sense of danger (P10). On the other hand, the 2D visualization, as a more statistic representation, failed to trigger more affective feelings for the environmental impact of a leak (P5, B3), lacking the connection to a ”real outcome” (B3):

”It does not give the sense of that something is going in the air and that it has an environmental impact because it’s just like numbers and bar height and it’s just not tie(d) [..] to a real outcome.” (B3)

However, B1 also noted a lack of realism in the 3D visualization describing it as ”only an illustration” and as such not bringing ”any value” because the essential information could be understood through the numbers alone. Similarly, P8 did not feel as if the ”particles differed in speed depending on the grade”, creating no impact for P8.

Colors Besides the realism of a visualization, the color coding also influenced the affective environmental awareness of the participants. Even with a lack of contextual knowledge, the red color successfully conveyed that the situation was ”dangerous” (P4) and could immediately communicate the situation (P6):

”Because when I got in [...] and everything was red, then I understood that immediately and then started to look at the numbers.” (P6)

While the 3D visualization created awareness through e.g., movements (S), the 2D visualization created awareness because the bars were more ”compact so that made it a lot more red” (P8).

In the case of the ethane ratio, the color coding was even declared as ”misleading” for the differentiation of biogenic and thermogenic gas, indicating that colors could also create a wrong sense of awareness. As B3 explained, the ethane ratio was only a ”quality indicator” (B3) and the red color in the thermogenic condition presumed that it was abnormal or alarming, although it was simply to indicate when a surveyor needed to pay attention:

> ”Put things just like ethane ratio 1 % or biogenic source to somewhere in the corner. And the color could change like to blue biogenic and then it switched to a brown or something like that. That is not, like triggering any sense of alarm and when it's a high concentration as you see now, it becomes red.” (B3)

Severity One theme that also emerged was that awareness was created when there were bad leakage situations shown in the 2D or 3D visualization. For example, during the experiment, P9 shouted out that ”here we have things happening” when the methane and ethane levels were high in the 2D visualization. Similarly, P3 exclaimed that ”Uff, this is not good”, looking at high stats in the 2D visualization. While low levels did not trigger any verbal or emotional reaction, high levels better communicated the environmental impact of a leak:

> ”The bars in the first one was more understandable and because [...] in the first questions, I didn't really understand what the boxes were and the particles. But then, when it came out very bad case where a lot of leakage were and the methane concentration was high and so on, and I started to realize, oh, that's the leakage.” (P6)

Environmental Awareness - Conative

Sensing Urgency One theme that emerged was that the willingness and urgency to act depended on the understanding of the situation and an emotional reaction which was intensified by the realism of the 3D visualization:

> ”I could see the rate and I could feel that maybe someone needs to go there to stop it in a way, but in the second chart, no, it was just information.” (P7)

Seeing the particles escaping when it was grade one triggered the thought that ”someone needs to stop” them from escaping, serving as a ”nudge” that ”we need to take care of it" (P3).

"Kind of tricks me to thinking about this is actually, this one particle is escaping now. The next one is going to escape. So like someone needs to stop." (P3)

While the visualization of the invisible methane particles could help create affective awareness in the 3D visualization, in the 2D visualization, it was more the red color that created a sense of urgency which made P10 "really want to act if" seeing all that red.

Personal Awareness and Responsibility There were different levels of personal awareness and responsability. Some participants simply felt not responsible (P2, P4). For instance, P1 explained that "it's a job" and that "the surveyor would invest as much time as is needed for the job at the minimum".

B1 confirmed a willingness to act, but not because of the environmental impact but because of the "safety of the community to make sure that nothing is going into someone's house and it being the possibility of being blown up". B1 explained it by being more "passionate" about the safety of the families, mentioning that for the grading only the infrastructure and risk for the community were important, not the leak rate:

"Because I know pipeline operations, we base our reaction by the grade and the risk of the community. Immediate repair would be grade one. Everything on grade two would just be fixed within the year and then for grade three, it just has to be rechecked. I would still have the same philosophy, so even if it's emitting very high at grade two, maybe I would put it 100 % at grade two as long as it's away from our structure, it would still fall as grade two." (B1)

Furthermore, B2 noted that there was a tension between personal and collective responsibility as "a lot of surveyors will [...] feel like they're at the whims of their company's policies" and country's regulations which decide which leaks need to be reported. Moreover, as infrastructures and regulations varied between countries like the US and India, the detection thresholds could vary a lot (B3).

Still, to increase the personal willigness to act, P10 suggested to more clearly show the direct, local environmental impact such as "trees dying" or "animals running away":

"I did feel some responsibility in regards to how the like methane and ethane and all of those hazards, how they would affect my surrounding. It was a little bit difficult to grasp because I couldn't see any like effect of the like on that red visualization. If I were too see, you know, like some example pictures of, you know, like grass dying or maybe like, you know, beautiful trees and plants, maybe that would you know make me more aware of the sustainability aspects.” (P10)

4.1.4 Summary of Results

To help prepare the following discussion, a summary of the most important results will be provided in this section.

Environmental awareness and cognitive load On a higher level, figure 4.22 displays the quantitative and qualitative results for each (sub-)dimension of environmental awareness and cognitive load, ignoring deeper comparisons between expertise levels and order of appearance. For RQ1 which was about the concept of environmental awareness, it could be concluded that the physical 3D visualization resulted in a higher environmental awareness. Although no significant difference could be found for all three dimensions, this conclusion was drawn by achieving a small effect for both cognition (11.76 % (2D), 9.24 % (3D), r = 0.24) and conation (median = 6 (2D, 3D), r = -0.25). While the affective dimension yielded no significant effect (median = 6 (2D), median = 5 (3D)), the qualitative data revealed that the physical 3D visualization was seen as more realistic and illustrative, triggering more emotions than the informational 2D concept. Apart from a more illustrative approach to indicate different levels of severity, they also indicated that the 3D visualization might have increased SA by relying on the mental model of how the physical world, and within it the relationships between physical data elements, works.

The results for RQ2, which was about the concept of cognitive load, suggested that three sub-dimensions (physical demand (r = -0.18), pace (r = -0.77), success (r = -0.12)) had a small to large effect for the 3D concept, while only the sub-dimension stress (r = 0.39) seemed to impact the 2D concept more. As more sub-dimensions were affected by the 3D visualization, it could be concluded that although a general comparison yielded no significant effect (median = 3 (2D, 3D)), a comparison of sub-dimensions showed that the statistical, iconographic 2D visualization resulted in a lower cognitive load. Factors that might have positively affected the cognitive load in the 2D visualization could have been a more exact representation of methane and ethane thresholds, the ability to be viewed from only one perspective in 3D space and a familiarity with the compass icon as an unambiguous indication of the wind direction. Especially the last aspect was not given for the 3D concept as the participants tried to relate the 3D wind arrow to their own body in space.

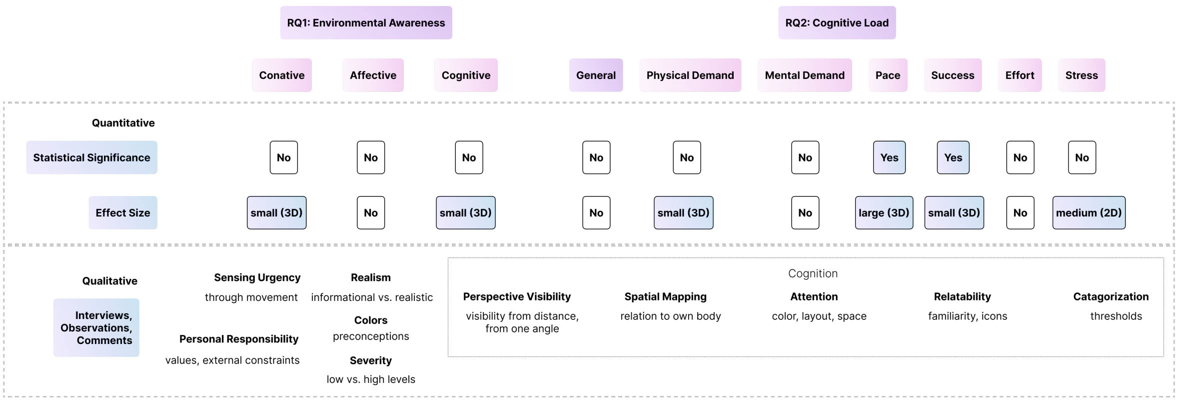


Figure 4.22: Overview of the quantitative and qualitative results. In regards to the quantitative results, the cognitive load sub-dimensions pace and success yielded statistical significant results. A small to large effect could be found for the conative and cognitive dimensions of environmental awareness as well as the physical demand, pace, success and stress sub-dimensions of cognitive load. The qualitative results have the themes ”Sensing Urgency” and ”Personal Responsibility” for the conative dimension of environmental awareness. The affective dimension has the themes ”Realism”, ”Colors”and ”Severity”. Both cognitive environmental awareness and cognitive load were united under the term cognition and entailed the themes ”Perspective Visibility”, ”Spatial Mapping”, ”Attention”, ”Relatability” and ”Categorization”.

Experts vs. non-experts Even if no statistical significance could be calculated for a comparison between experts and non-experts, the results revealed that experts did more errors when answering the SAGAT questions for the statistical, iconographic 2D visualization (17.65%), while non-experts had more difficulties with the physical 3D visualization (11.11 %). In terms of affect and conation, non-experts had a lower level of emotions for both 2D and 3D concepts (median = 5) than experts (median = 6), while they only had a lower level of willingness to act for the 2D one (median = 5), median = 6 (experts)). Non-experts also had a higher cognitive load for the 2D visualization (median = 4, median = 3 (experts)).

Order of appearance Both the order of the concepts and SAGAT questions seemed to affect the SA of the participants. Starting with the 3D visualization resulted in higher error rates than when starting with the 2D one (7.98 % (2D), 13.45 % (3D)). This statistical result was reflected in the theme of relatability which meant that the 2D visualization was better understood since 2D interfaces are more familiar to users. Moreover, the error rate decreased with the flow of questions, being highest for the wind-related questions which could be explained by the attention problem that color and layout made the wind information less perceivable than the methane and ethane concentration.

4.2 Discussion

This section served the goal of relating both quantitative and qualitative results to each other and interpreting the data with respect to the examined concepts of environmental awareness and cognitive load as well as to relate the findings to related work. In general, there were no significant differences in the environmental awareness and cognitive load of the 2D and 3D visualizations in all dimensions (cognitive, affective, conative) which might be the result of a small sample size (n = 14) and the use of a non-parametric test which is generally less powerful than a parametric test [122]. Another explanation could be that both visualizations had their strengths and weaknesses which made B2 for example suggest to combine the 3D wind arrow and 2D bar chart. As five data types were combined into one visualization, they might have counterbalanced each other. A more detailed discussion of each subresearch question as well as a more general comparison to previous related work will be shown in the following sections.

4.2.1 RQ1: Environmental Awareness

As a reminder, the first research question had the following form:

RQ1: *What is the environmental awareness of statistical, iconographic 2D and physical 3D visualizations in mixed reality?*

To summarize the results, the cognitive environmental awareness was high which was illustrated by low error rates of 11.76 % for the 2D and 9.24 % for the 3D visualization. On a seven-point Likert scale, the affective environmental awareness turned out to be high for the 2D (median = 6) and 3D (median = 5.5) visualization. Likewise, a median of 6 for both 2D and 3D visualizations suggested a high conative environmental awareness. No significant difference could be found between the 2D and 3D visualization in terms of all environmental awareness dimensions (cognitive, affective, conative). While no effect could be found for the affective dimension, a small effect in favor of the 3D visualization was noticeable for both cognitive (r = 0.24) and conative (r = -0.25) dimensions. While no significant difference was found, the effect sizes together with the qualitative data analysis suggested that the environmental awareness was higher for the 3D than for the 2D visualization. A more detailed analysis of each dimension can be found below.

Cognitive Environmental Awareness

The overall cognitive environmental awareness could be stated to be high for both 2D and 3D visualizations with a mean success rate of 89.5% and error rates of 11.76 % for 2D and 9.24 % for 3D. While these differences in favor of the 3D visualization were not significant, a small effect (r = 0.24) could be found, indicating that the 2D visualization had a higher error rate than the 3D visualization.

Experts vs. non-experts A differentiation between expert/survey and non-expert/ usability study gave a more thorough picture. The experts struggled more with the 2D visualization (17.65%) than with the 3D one (7.35%). This could be explained by the fact that the 2D visualization relied more on numbers and text (P3, P6) which might have been harder to perceive through the video that generally had a poorer quality than the real experience. The inability to view the visualization in the physical world might have also decreased the immersion and attention which could have explain why the overall mean error rate was higher for experts (12.5%) than for non-experts (10.26%). On the other hand, non-experts who performed the usability study had more errors with the 3D visualization (11.11%) than with the 2D one (9.41%). Potential reasons could be that some participants initially had trouble understanding the meaning of the 3D boxes (P6), did not move side to side to understand the perspective of the 3D wind arrow and were confused if the 3D arrow signified the wind direction or the direction of where to go (P1, P3, P6).

Understanding the 3D visualization in 3D space through interactions with a physical intermediary While the participants were encouraged to walk around the 3D visualization which was positioned at shoulder height at a certain location in 3D space, only few looked at it from all perspectives. An ability to directly rotate and interact with the visualization might have eased the understanding of the wind direction in the case of the 3D wind arrow. However, this method was not chosen as the visibility of the visualization was time-constrained and rotation by hand was cumbersome in MR. Instead, attaching the visualization to a real wand might help solve the problem of the perspective visibility of the 3D wind arrow since this would allow the viewer to rotate the visualization through the rotation of the wand. Moreover, the wand would then naturally be tilted down to the ground and the visualization would be visible from the top rather than from the front as it was the case in the experiment. Furthermore, using a physical wand which does not only serve as a non-functioning prop but rather as a holder of the visualization would allow to understand more about a suitable size or distance between the visualization and the viewer to make sure that the visualization is fully visible (P3) from the intended medium.

Order of visualization concepts and SAGAT questions In general, starting with the 3D visualization (13.45 %) led to higher error rates than when the participants started with the 2D one (7.98 %) which could be caused by the fact that most participants (n = 9) were unfamiliar with the surveying context and did not have the same associations with the 3D visualization as experts did. Interestingly, the visualization the participants started with had less errors for 2D (1st: 7.56 %, 2nd: 8.4 %) and 3D (1st: 10.92 %, 2nd: 15.97 %) which suggested that although the participants perceived some learning effects, the attention seemed lower towards the end of the experiment.

Regarding the error rates per question, the error rates approximately reflected the order of appearance with earlier questions having higher error rates than later questions, indicating some learning effects that were also mentioned by multiple participants (P1, P2, P3, P9). Another reason could be that the first questions which were related to the wind direction and speed were simply more mentally demanding as explained during the post-interviews (P1, P3, P6) and more difficult to answer which was reflected in a high error rate of 42.86 % for the wind direction and an error rate of 38.1 % for the wind speed.

Another reason why the wind questions led to more errors than the other questions could have been caused by the lack of attention on the wind information due to more attention grabbing elements. For instance, some participants complained about the saturation of the thick, red bars which drew all their attention (P7, P10), distracting them from the white (wind) text (P10). However, although color could affect the attention, shifting it towards or away from certain elements, it also helped getting an immediate understanding of the severeness of the situation, supporting especially non-experts in grasping the context who proclaimed having no idea what they were doing (P4).

Situation awareness levels While the participants generally perceived the information which was indicated by a mean error rate of 4.17 % for the first level of SA, level 2 and 3 yielded much higher error rates (2: 13.1 %, 3: 12.5 %) which P3 explained in the interview as the inability to understand the data's meaning.

Strategies to increase a cognitive awareness To help interpret the wind information, indirect hints were suggested to be integrated into the visualization such as the movement of the plume according to the wind direction (P7, B3). This suggestion indicated that the realistic representation might more encourage a connection of the visible elements to adhere to preconceptions about the real world.

Affective Environmental Awareness

While the affective environmental awareness was high for both 2D (median = 6) and 3D (median = 5.5), no significant effect could be found. From the interviews, it could be inferred however that the 3D visualization with the moving leak rate particles could create more awareness by materializing them into the air and visualizing how they were released into the environment (S, P3, P6, B2, B3). On the contrary, participants described a lower affective environmental awareness for the 2D visualization, describing it as information (P7, B2) or stats (P6) and associating it with numbers (P3). While the realism and movement of the particles created more affection for the 3D visualization, the thickness and redness of the red bars could communicate the impact for the 2D visualization (P7, P10). The color coding also had this effect in the 3D visualization, helping especially non-experts with no surveying knowledge understand the situation (P4). The awareness was then especially high when the situation looked severe (P3) and things were happening (P9), while a low level of methane and ethane did not prompt any emotional reaction (P6).

Experts vs. non-experts A comparison between experts and non-experts revealed that experts had a higher affective environmental awareness (median = 6) than non-experts (median = 5). While the significance of this difference is unknown, this result indicated that experts were more emotionally aware due to their information advantage. This knowledge gap could also explain why non-experts were generally more frightened by the amount of methane emissions (2D: median = 6, 3D: median = 5.5) than experts (median = 4.5). While this statement was more about a vague feeling of fear, experts felt more worried (median = 6) than non-experts (median = 4.5) about the effect of these emissions on themselves and their families for the 2D visualization. Moreover, experts were also angrier about the harm done to the climate by leaking gas (2D: median = 6.5, 3D: median = 6) than non-experts (median = 5). These differences could potentially be a result of being more aware of the amount of these emissions and their consequences.

Strategies to increase affective awareness Several strategies to increase the affective awareness were provided by the participants. For instance, the visualization should be more real and less illustrative (B1) by changing the speed of the particles (P8) and showing the direct impact on the environment e.g. through dying trees or grass (P10), demonstrating not only that particles are released but also how they affect the local environment.

Conative Environmental Awareness

The conative environmental awareness was high (median = 6) for both visualization types. While this difference turned out to be not significant, a small effect (r = -0.25) still indicated that the willigness to act and reduce the methane emissions was higher when viewing the 3D visualization than when viewing the 2D visualization. In the interviews it was mentioned that especially the realism of the 3D visualization, seeing the moving particles escape (P3), could create a sense of urgency that someone needed to go there and stop the particles from leaving (P7).

Experts vs. non-experts In general, experts showed a higher conative environmental awareness (median = 6) in comparison to non-experts (median = 5), indicating that they were more involved with the surveying impact as it was their job. Looking more closely at the individual statements, experts also revealed a higher willingness to diligently do their job (2D: median = 6.5, 3D: median = 6) than non-experts (median = 5) with a very low standard deviation (2D: σ = 0.5; 3D: σ = 0.5).

Strategies to increase a conative awareness This devotion could result from a contribution to environmental safety or human safety as B1 was for example more passionate about the safety of families, emphasizing that the leak categorization did not yet include the harm done to the environment. Furthermore, B2 mentioned a tension between personal and collective responsibility as surveyors had to rely on a country's and company's regulation of what leak should be reported which restricted their freedom to act.

Both answers indicate that triggering different user motivations and improving external conditions could help in increasing a willingness to make a sustainability impact.

4.2.2 RQ2: Cognitive Load

The second research question, which will be answered in this section, had the following content:

RQ2: *What is the cognitive load of statistical, iconographic 2D and physical 3D visualizations in mixed reality?*

The cognitive load was rather low (median = 3) for both 2D and 3D visualizations, demonstrating no significant effect. Nonetheless, the analysis of the six subdimensions led to some statistically relevant results. While the physical demand was not significant, a small effect (r = -0.18) could still be found, indicating that the 3D visualization had a higher physical demand than the 2D one. While a possible reason could be that the viewer was forced to move around to fully see the 3D visualization, it could also be a result of the participants holding the wand and/or paper sheet which was optional and was generally not done for both subsequent surveys. The difference in pace turned out to be significant and to have a large effect (r = -0.77) with the 3D visualization survey having a .5 higher pace than the 2D one. The reason why the 3D visualization was perceived to have a higher pace could be because of the perspective visibility problem that viewers had difficulty interpreting the 3D wind arrow (P3, P6, B3) and that the meaning of the 3D boxes was not clear at the beginning (P6). Another reason could be that the 2D visualization was just more informative (B2) and through the bar chart's y-axis more straightforward (P8) and easy to understand (e.g. P4, B1). A comparison of the reverse-coded success dimension also yielded a significant, small effect (r = -0.12), indicating that participants felt slightly more successful with the 2D visualization which might be because they felt more familiar with the 2D interface (B1) and struggled to understand what the 3D visualization meant at first (P6). Moreover, while the difference in stress was not significant between 2D and 3D, a medium effect (r = 0.39) was still found, implying that the stress was higher when looking at the 2D visualization. One reason might be the attention grabbing thick, red bars (P7, P10) which distracted from noticing the white text (P10).

Experts vs. non-experts Looking at the difference between experts and non-experts, experts had a lower cognitive load (median = 3) compared to non-experts (median = 4). Furthermore, there were also a few differences in the six sub-dimensions when comparing experts and non-experts. The mental load was lower for experts (median = 3) than for non-experts (median = 5), indicating that contextual knowledge decreased the mental demand in understanding and interpreting the data that non-experts struggled with (e.g. P3, P4, P6). Another reason could also be related to the procedure of comfortably viewing the visualization in a survey from home versus conducting a usability study in a lab and viewing the visualization in 3D space. Concerning the effort dimension, experts also had a lower effort for both 2D and 3D visualizations (median = 3) than non-experts (2D: median = 4, 3D: median = 5). While the procedural differences could have led to this result, the advantage of having contextual knowledge might have also played a role.

Order of visualization concepts As the order of the visualizations could have impacted the cognitive load of the participants, a comparison between those visualizations which came first and the overall median score revealed some differences.

On the one hand, when starting with the 2D visualization, the cognitive load was lower (median = 3) than the overall median (median = 4). On the other hand, when starting with the 3D visualization, the cognitive load was higher (median = 5) than the overall median (median = 4), indicating that the cognitive load was higher when the participants started with the 3D visualization than when they started with the 2D one. One explanation can again be the difficulty to initially understand the 3D boxes (P6) as well as the difficulty to interpret the 3D wind arrow (e.g. P1, P3, B3).

Strategies to lessen the cognitive load A proposition to positively influence the cognitive load and SA was to improve the information layout for both 2D and 3D visualizations (P3, P5, P6) which should consider the most attention grabbing elements and the normal gaze or reading direction from top to bottom (P3). As e.g. participant P3 noted, the gaze always went from the bar chart down which made P3 not notice the leak rate which was above the bar chart. While this was less of a problem for the 3D visualization where the moving leak rate particles caught attention, the 3D visualization had other attention problems, namely the lack of attention on the 3D ethane ratio box (P8, B3).

Finally, a remark by P4 who compared the viewing of the visualizations to a movie in which new information would be noticed with each rewatch, there could be different layers of attention, hiding and showing crucial information at different temporal events.

4.2.3 Overall Discussion

This research contributed to previous AR / MR research by comparing 2D and 3D visualizations in MR with regard to environmental awareness and cognitive load, bringing the three-dimensional concept of environmental awareness (cognitive, affective, conative) from social sciences research [32] into current AR / MR research which often employs knowledge questions to evaluate the environmental awareness gained through interacting with immersive environmental applications [22].

One aspect which emerged from the data was the tension between a numerical exactness which was more represented by the 2D visualization and a more illustrative realism characterizing the 3D one. Although there was no significant difference between 2D and 3D visualizations for both examined concepts, an analysis of the effect sizes and qualitative data as well as a more in-depth analysis of the six cognitive load sub-dimensions yielded some results which postulate that the physical 3D visualization was better at increasing environmental awareness, while the statistical, iconographic 2D visualization was better suited to lower cognitive load.

Information hiding in the temporal and spatial dimensions As some participants criticized that too many elements were shown at the same time, especially in the 2D visualization, the temporal dimension of information could be taken into consideration, dividing data into those which need constant monitoring like the wind direction and the methane concentration and those which are only relevant at a specific moment in time, e.g. when being close to a leak and needing to categorize if and how big a leak was. These temporal events could be based on the corresponding decisions that need to be taken and would allow to limit the number of visible elements at a time. Such *information hiding* technique was for instance applied to a 3D information visualization of a map displaying a large time-dependent health data set, using an event-based approach to e.g. only toggle data exceeding a certain threshold in a certain time period [128]. Not only the temporal but also the spatial dimension could be relevant to decrease the cognitive load by guiding the gaze direction more intentionally from the most important information to the least important one, taking into consideration the natural gaze flow of a reader. Furthermore, information could be hidden based on the distance to the viewer. Such a spatial approach was for example taken by Chen et al. [129] who developed an interactive screen with three communication zones (interaction, notification, ambient) whose level of detail and interactivity diminished with increased distance to the viewer. By considering both the temporal and spatial dimensions, the visualization could develop a kind of *data story* which unravels over time and draws the attention on important highlights throughout a viewer's journey.

Colors as attention guides The color coding turned out to directly draw the attention which goes in line with previous research which defined colors as a preattentive feature [76]. By acting as an attentional starting point, the color density and size impacted the gaze movement and finally which elements would be seen and which not. To not cause a wrong sense of awareness, signalling colors should not be selected for data which is not inherently bad or simply less interesting such as the ethane ratio (B3) since color are loaded with preconceptions [130].

Perspective visibility through interaction or multiple views Another aspect that became apparent was how important perspective visibility was to correctly interpret the 3D visualization. While a 2D visualization could be fully perceivable from one direction, a 3D visualization either would require more interaction to actively change the perspective and rotate the visualization or other views or perspectives to passively consume the data from another angle. This could be achieved by rotating the visualization on a physical wand or adding a bird's eye or side-view to the application, as it was for example done in a mobile AR application for environmental monitoring that enabled a shared, remote view through pan-tilt and infrastructure cameras [80].

Encoding of data in the physical environment To reduce the confusion of the single 3D wind arrow being either the wind direction or the navigation of where to go, instead of visualizing everything together within a compact visualization, the information could also be spread out and encoded in the physical environment. By representing the wind direction and speed not through one arrow but rather a vector-like field or streamlines as it is common for representations of the wind [9], the viewer could better understand that this information originates from the physical world and not from a recommendation of the system. While spacing out the elements might help digest the information and ease focus, the limited field of view of the HoloLens 2 device also needs to be taken into consideration.

Using metaphors and forming causal relationships between data elements The analysis of the qualitative interview data indicate that although there was no effect and significant difference between both visualization concepts and the affective environmental awareness dimension, the 3D visualization with its more illustrative, physical way of displaying the intangible and invisible wind and gas data could elicit a more emotional response with regards to the environmental impact of a leak. Thus, these results extend current 2D and 3D research which described AR / MR to be well suited to present spatial data in 3D for a better understandability and performance [11], [12] by indicating a stronger emotional response due to the physicality of the data. Using more illustrative and metaphorical approaches such as eco-visualizations [131] have been a common way to promote ecological literacy, representing e.g. energy consumption behavior through a healthy and dying tree [132].

To increase the usefulness of the 3D visualization in gaining a SA and to render it less illustrative, the 3D visualization should be more authentic by using real data and mirroring the preconceptions of the viewer. This could include the provision of indirect hints, e.g. for the navigation, as the realism of the 3D visualization encouraged a form of causal connection that the wind could affect the direction of the 3D plume. These indirect hints could be based on an understanding of the real, physical world and its inherent physical effects, rendering the invisible visible as it was done e.g. for a city's air pollution [8]. While the 3D visualization was mostly associated with the real world, the use of iconic icons in the 2D visualization could also invoke certain associations to the real world as it was explained by Atkin [87] by using a compass icon to communicate the wind direction. Here, to be relatable, the knowledge of the viewer should be taken into consideration to ensure that icons are timeless and understood throughout technological trends.

Cognitive load and situation awareness

The results suggest that while the physical 3D visualization led to a lower error rate and therefore higher SA, the statistical, iconographic 2D visualization seemed better suited to lower cognitive load. While the relationship between both concepts was not investigated as part of this research, previous work often puts both concepts in an asymmetrical relationship in which the increase of one concept signifies the decrease of the other [33]. For instance, one study employed the NASA-TLX and SART to measure cognitive load and SA for human users of autonomous vehicles who viewed AR visualizations of semantic segmentation information. Their conclusion was that the AR succeeded in reducing cognitive load and increasing SA [133]. In another paper about space explorations, it was explained that mission control room workers could experience high cognitive load which could lead to a loss of situation awareness and potentially hazardous situations [134].

Cognition and affect

Although cognition and affect were viewed as distinct dimensions within the concept of environmental awareness, it has to be noted that both cognition and affect do not function independently from each other. For instance Bower et al. [135] reject the notion of humans as rational creatures and *information-storage* devices who can look objectively at things without any bias. They pointed out that emotions influence memory and judgment as humans attend to and select stimuli that match their current feelings. As they experience more feelings and interest towards a matter, the material becomes more salient and a deeper processing can take place. They explained this phenomenon through the fact that humans tend to remember events which evoked a strong emotional response (mood-congruity effect). Secondly, they pointed out that emotions could help to remember memories that were associated with that emotion or mood (mood state-dependant retrieval). Lastly, they mentioned that a momentary mood could influence the cognitive understanding or judgment of a human's impression of others, their thoughts about the future and their competencies. In a similar notion, Duncan & Barrett [136] described affect and cognition as phenomenologically but not ontologically distinct. They argued that affect was necessary for conscious experience, language fluency and memory since affect filters what sensory information is encoded more fully and therefore memorized in the human brain. As they put it there is no *non-affective thought* even if humans experience emotions separately from thought. Thus, it is important to not view the cognitive and affective dimensions of environmental awareness as entirely separate categories but rather as overlapping, affecting each other.

Experts vs. non-experts The evaluation of the SA through SAGAT resulted in a slightly better performance in the 3D visualization and a small effect in favor of it, indicating that the cognitive environmental awareness was higher for the 3D visualization. As non-experts represented the majority of the target sample and as the interviews revealed that experts relied more on numbers (B1) and wanted to have more of them (B2), there is support for findings from previous research [80] which suggested that 3D visualizations were more understandable for non-experts when doing environmental monitoring work in unfamiliar environments.

Trust and agency with automated systems Even if some interviewees suggested to replace the wind direction arrow with a navigational arrow, the question arises how much responsibility should be given to the system and how much decision power should remain with the surveyor. As surveying could require the use of non-technical leak detection methods which rely on the natural senses (e.g. smelling, hearing) [3] and as every surveying situation might depend on multiple factors [108], a fully automated navigational system might eventually result in a lack of trust in the system [137]. Instead, providing the wind data in all three SA levels, i.e. the wind direction, the meaning for one's own location and where to go next, could be a more suitable approach.

Empowering the individual and reinforcing corporate responsibility Although there was no significant difference for the conative environmental awareness between 2D and 3D visualizations, a small effect could be found, indicating that there was a higher willingness to act and contribute to the topic of sustainability for the 3D visualization. This was also expressed by the qualitative data in which participants described a sense of urgency through escaping particles into the air.

However, despite a willingness to act, the individual contribution might still be constrained by a collective responsibility enforced by a country's or company's regulations which specifies at which threshold an emission should be considered a leak and which leaks should be reported, therefore overruling an individual's personal responsibility. As current SHCI practices focus largely on individual awareness and responsibility and less on industrial responsibility [17], making attempts to increase corporate social responsibility or empowering the individual in a corporate environment could be future endeavours worth taking. Especially in the spirit of the industry 5.0, a job should not only remain a job but also create meaning for the employee [18], balancing out the need to be efficient with the need to understand the impact of one's work and developing a desire to act.

Bridging the gap between cause and effect This persuasion might be easier by visualizing the direct, local impact close to the viewer instead of showing the more abstract, global impact whose effect is much more delayed and less palpable. This problem of having a gap between cause and effect was also brought forth by Moser [4] who noted that there was in fact a delayed gratification for climate change actions and who described a series of issues which challenged a successful communication and willigness to act. These were, amongst others, the invisibility of causes which was addressed in this research, the distance of impacts, a lack of immediacy and direct experience of the impacts and a lack of gratification for taking mitigative actions.

4.2.4 Limitations

As in every research, this research project involved some limitations which are important to mention.

An essential limitation was that the 2D and 3D visualizations did not only differ in their dimensionality but also their form since the 2D visualization had a statistical, iconographic representation and the 3D visualization had a physical one. This means that the results cannot be applied to the dimensionality alone but must be tied to the used forms.

Furthermore, while the questions and corresponding visualizations were checked by a surveyor before conducting the other experiments, the leak scenarios merely represented a simplification of the surveying scenario and did not encompass all the information relevant for the reporting of a leak.

Another important limitation of this research was the inaccessibility of experienced surveyors with whom the usability study should have been conducted and who would have had the necessary contextual experience to assess the usefulness of both concepts for their work. In order to still get their feedback, questionnaires with embedded videos were distributed which had a lower resolution and video quality than the real HoloLens experience due to the screen-recording capabilities of the device, displaying blurry text and numbers with sharp contours. On the other hand, as the participants who joined the usability study had no experience in surveying, their knowledge only consisted in the distributed surveying rules and instructions as well as the introduction by the researcher which made it difficult for them to answer the questions from the perspective of a surveyor and might have negatively impacted the validity of this research.

Moreover, although differences between experts (n = 4) and non-experts (n = 10) were drawn, both groups were quite unequal in their numbers and the described differences between experts and non-experts could have been both caused by their level of contextual surveying knowledge and the applied study method of a survey or usability study.

Regarding the overall sample size, this research was conducted with a low number of participants (n = 14) which could have been a reason why there was no significant difference between both visualization concepts in all dimensions of environmental awareness and cognitive load. A conducted power analysis with a significance level of 0.05 % and an effect of 0.8 would have suggested a required sample size of 20 participants [138] which was not achieved due to the unavailability of surveyors and the secrecy of the research which made it unavailable to external participants. A higher sample size would have been appropriate, especially since some differences led to some small effects that could have turned out to be significant with a higher number of participants.

There were also essential limitations regarding the SAGAT method. Since each leak scenario only differed in the visualization concept and the data used and no other interaction was required except to look at them, it raises the question if the SAGAT method was used to its full potential. This holds especially true as the nature of these scenarios were rather repetitive, short and simplistic, while SAGAT traditionally presents more varied scenarios in a long procedure such as air flight control where questions would be generally asked in two to three minutes time intervals in a 20 min long sequence [37]. Moreover, the question arises whether SAGAT could be used over a survey as it is traditionally done in simulations where the operator can directly view and interact with the scenario [33]. While the SAGAT method implied random freezes [37], in this research, freezes were deterministic after 40 seconds, posing the question if implemented freezing negatively affected the validity of the SAGAT results. Using SAGAT through the questionnaire also had the drawback that videos could be rewatched more often than the 40 seconds and paused so that the survey participants had an advantage in answering the SAGAT questions.

While the overall scores for the affective and conative environmental awareness were high, it is not clear whether they stemmed from the 2D and 3D visualization or were simply a result of a general high awareness as no pre-test condition was evaluated to verify their previous awareness level. Such pre- and post test method was for instance performed to compare the knowledge of maritime habitats through traditional textbooks with the immersion in a MR experience [22]. However, this approach would have required a between-subject design with a higher number of participants to be able to compare the before and after condition instead of the between condition as it was done in this research.

Another aspect which was not addressed in this research was the accessibility of the visualizations. As one participant was older and had difficulty viewing the numbers in the visualizations, more research would have been needed to come up with an appropriate font-size, layout spacing and distance to the viewer.

Besides these issues, there were also a couple of other procedural mistakes which might have negatively affected the reliability of the results. For instance, in P1's usability study, the visualizations were too far away from the participant so that the readability of the text was lower than it should be. In a couple of cases, the first visualization needed to be repositioned at the beginning of the experiment which led to a few lost seconds in the visibility of the first leak scenario. However, since the SA level 1 error rate was very low in comparison to the error rates of the other two SA levels, it can be assumed that the data was still perceived and the operational mistakes did not influence the results too much.

Finally, the survey with a surveyor led to a correction of the ethane understanding and a change of the methane and ethane numbers. Nonetheless, the results of this survey were included in the general analysis because the exclusion would have meant that no surveyor would have been represented in the results.

Chapter 5

Conclusions and Recommendations

5.1 Conclusions

The results of this research suggested that there was no significant difference in the creation of environmental awareness in all its three dimensions - cognitive (2D: error-rate = 11.76%, 3D: error-rate = 9.24 %), affective (2D: median = 6, 3D: median = 5.5), conative (median = 6) - between the statistical, iconographic 2D and physical 3D representation. There was also no significant difference in terms of cognitive load (median = 3). The lack of a significant difference could have been caused by having multiple data elements (wind, ethane, methane) combined into one visualization, counterbalancing each other as both visualizations had different problems of understandability and attention. Moreover, a small sample size of 14 participants and the use of a less powerful non-parametric test due to the rejection of the normal distribution assumption might have resulted in this lack of significance.

Despite the absence of statistically significant results, the calculation of the effect size yielded a small effect for the cognitive environmental awareness (r = 0.24), suggesting a lower error rate for the 3D visualization than for the 2D one that might have been caused by a lack of understanding on SA level 3 and a low video quality of the embedded survey videos. Another small effect (r = -0.18) could also be found for the conative environmental awareness, indicating that the willigness to act and reduce methane emissions was higher when viewing the 3D visualization than when viewing the 2D visualization. From the qualitative analysis, it could be concluded that this willingness was invoked by the realism of the 3D visualization creating a sense of urgency to act through escaping particles into the air. Similarly, although no significant effect could be found for the affective environmental awareness, the qualitative analysis revealed that the realism of the leaking particles helped sensing the impact and feeling the severeness of a leak.

A comparison of medians between experts and non-experts without statistical significance indicate that the conative environmental awareness of experts was higher for both visualization types. The affective awareness of the 2D visualization was also higher for experts than non-experts. These results suggest that a relatability with the context had a positive impact on both awareness dimensions.

A closer look at the six sub-dimensions of cognitive load yielded some significant effects even though the overall concept did not. The pace had a large significant effect (r = -0.77), rendering the 2D visualization slower paced than the 3D one. In addition, the reverse-coded success dimension also had a small significant effect (r = -0.13) in favor of the 2D visualization. Possible reasons could be that the 2D visualization felt more informative, especially the presentation of the methane gradings on the bar chart were positively highlighted during the interviews. There was also a slightly lower physical demand for the 2D visualization with a non-significant but small effect (r = -0.13), suggesting that more movement was needed to understand all the perspectives of the 3D visualization. Finally, a non-significant but medium effect (r = 0.4) could be found for the stress dimension which suggested that the 2D visualization was more stressful which could result from the more attention grabbing elements in the 2D representation for serious leak conditions.

To conclude, the research results suggest that a physical 3D visualization could lead to a higher environmental awareness in all its three dimensions, while a statistical, iconographic 2D visualization might result in a decreased cognitive load.

5.2 Recommendations

Some recommendations can be provided which concern the improvement of the research done and possible future work directions.

A first set of recommendations concerns the procedural and study design decisions which could be altered in a future iteration of the research. To begin with, the visualizations should also be evaluated with experts in person instead of using surveys as the difference between evaluation designs might have affected the results and the research's validity. Furthermore, it can not be concluded how much the affective and conative environmental awareness were a result of the visualizations and how much this awareness existed independently from viewing them. This problem was caused by using general statements which did not directly address the visualizations themselves and by not testing the awareness before the intervention or experiment. Therefore, a pre-test and post-test could be used in a between-subject design to understand if and how much both visualization types influence the affective and conative environmental awareness dimensions. Concerning the SAGAT method, less time could be given to better trigger the preattentive processing rather than a memorization process as SAGAT was criticized of being overly reliant on working memory [37]. Moreover, as the order of questions for each data type was fixed, starting with the wind and ending with the leak rate question, a random order might reduce any learning effects and provide more evidence that the wind questions were not solely more difficult because they appeared at the very beginning.

Another recommendation to improve the presented research is related to the selection of the sample size. Increasing the sample size could increase the strength of this research by allowing the use of a parametric test with a higher likelihood to yield significant results [122]. As this research was restricted by the regulation of secrecy, only ABB employees could be chosen as participants who are not representative of a general population. Furthermore, as most ABB employees were not familiar with the context of gas leak surveying and had to be given some introductory information, the lack of familiarity with the context could have impacted the environmental awareness and cognitive load results. Therefore, more experts should be invited to participate in this research who can better evaluate the chosen visualizations in terms of their relevance for their work and their understanding of their work's impacts on the environment.

In the following, some recommendations shall be given that could become questions worth investigating through future work.

One potential area of future work could be to assess the visualizations in a non-static, outside environment with no predictable lightening conditions where dynamic cues might compete with the visualizations for attention. The goal could be to understand how an outside usage would affect the suitability of 2D and 3D visualizations under different weather and lighting conditions. This would fill the gap of existing research which mostly focuses on indoor usage or small work spaces, leaving out environments with a large amount of dynamic variables [83] such as obstacles in a rough terrain or approaching entities like cars that could occur in a gas surveying context.

A recommendation which more directly builds up on the presented research concerns the relation to a physical object. As some MR research simply spawns virtual objects into 3D space [66], [82], the visualization could be improved and prototyped in a way that the visualization is attached at the tip of a real surveying wand through image or object tracking. Thus, instead of using a non-functional stick prop during the usability study, a functional wand could be used to move and interact with the visualization. This would enable the investigation of how the 2D and 3D visualizations would be perceived and understood at a certain distance and how moving the visualization through a prop would affect the awareness level, including e.g. the perspective visibility of the visualization.

Instead of using an object to change the perspective of the visualization, future work could also experiment with a combination of multiple perspectives such as a birds-eye and side view since the 3D wind arrow caused some visibility issues. As some participants requested navigational hints and brought up maps like Google Street View, the visualization could also incorporate a map showing a bird's eye view that could either be in a 2D or 3D perspective using the additional z-axis. Focusing more on navigational than qualifying features would allow to significantly increase the SA and decrease the cognitive load as the wind questions appeared to be the most difficult ones where most errors happened on the SA levels 2 and 3. Using multiple perspectives helped for example increase spatial awareness of scientists in environmental monitoring tasks [139].

Moreover, while this research evaluated static visualizations that were shown in one location in space, future research could improve the visualizations in this research to make them dynamic and change depending on real gas leak data such as the current methane concentration and other dynamic data like the wind direction and speed. This would allow to investigate the temporal dimension of the data, exploring transitions between states which could complement and confirm the findings from this research regarding 2D and 3D visualizations. Potential questions could concern the understandability and attention shift of when a threshold was passed for both visualization concepts and how cognitively demanding the visualizations are in low and high level conditions. Especially low level conditions, when no leak is found yet, could be interesting to investigate for a longer period of time as this is probably the most common scenario in a survey setting which permits the surveyor to save some cognitive resources.

Another aspect worth investigating by future research could be the attention shift through the spatial and temporal dimensions of the visualization. More concretely, this would involve the question of how the data should be spaced out and how it could evolve over time to increase the environmental awareness and decrease the cognitive load. This would involve the exploration of which information would need constant monitoring, which information would only be shown at a certain moment in time and what level of detail and interactivity should be chosen based on the distance to the viewer. Especially for MR, there is a possibility to lay out and encode information in 3D space such as the air movement [9] or air pollution [7], allowing a clearer understanding of where the data belongs to.

Furthermore, there was an indication that the data shown in the physical 3D visualization was not apprehended independently from each other but that there was a mental preconception that, for instance, the wind direction should have an effect on the direction of the plume. Therefore, another research direction could be to look more into how indirect cues could be provided in all data elements to mirror the mental model of the viewer and increase understanding.

To explore more the topic of environmental awareness, future research could also address alternative challenges of climate change communication which were mentioned by Moser [4]. Apart from the invisibility of the causes, other aspects like the lack of immediacy and direct experience of impacts as well as the lack of gratification could be aspects worth exploring through information visualizations in MR. By better visualizing the direct local impacts and bringing the delayed, distant and global impacts closer to the viewer, a positive effect on the level of environmental awareness might be achieved.

Apart from these aspects, previous research suggested that user characteristics like cognitive abilities, personality traits and domain expertise might have an effect on a visualization's effectiveness [140]. While this research compared the domain expertise level of the participants, especially the cognitive or spatial abilities might be an important factor to consider and evaluate to understand more about the suitability of 2D and 3D based on the user's profile.

Lastly, while the visualizations were solely viewed by one participant at a time, in the real world, collaboration between different entities might be plausible, e.g. with the person in the control center who collects all the reported data and decides on the prioritization of leaks. Interactions between multiple entities in MR could for instance include the joint annotation of data or real-time data influx from other persons and devices to establish a shared SA. Collaboration in a MR setting has already been explored by previous research in the form of flood risk visualizations [66], showing that collaborative MR interfaces could be useful for understanding geospatial data.

5.3 Reflections

This research positively contributed to society by tackling the 12th SDG of responsible consumption and production [30] in the gas and oil industry who represents a major culprit of harmful methane emissions in the atmosphere [1]. By visualizing environmental data like the leak rate which constitutes a quantification of methane emissions that can help surveyors correctly assess the situation, it facilitated the reduction and communication of methane emissions, thus contributing to environmental safety. Ethically, this research investigated how dimensionality and form could decrease the cognitive demands of surveyors in MR which indirectly affect their bodily safety. Through the exploration of how environmental awareness can be increased, it put emphasis on the meaningfulness of one's work, sharing the ideals of the Industry 5.0 to use technology in a way to enhance the abilities of the human and uphold their human dignity. Lastly, this research also impacted the economic dimension by bringing forth the environmental data which can render leak detection solutions more attractive to customers and by exploring factors that can help in the surveyors' work, thus decreasing the economic damage made by leaking gas.

Bibliography

[1] "Driving Down Methane Leaks from the Oil and Gas Industry – Analysis," 2021. [Online]. Available: https://www.iea.org/reports/driving-down-methane-leaks-from-the-oil-and-gas-industry

[2] H. Lu, T. Iseley, S. Behbahani, and L. Fu, "Leakage detection techniques for oil and gas pipelines: State-of-the-art," *Tunnelling and Underground Space Technology*, vol. 98, p. 103249, Apr. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0886779818311131

[3] P.-S. Murvay and I. Silea, "A survey on gas leak detection and localization techniques," *Journal of Loss Prevention in the Process Industries*, vol. 25, no. 6, pp. 966–973, Nov. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950423012000836

[4] S. C. Moser, "Communicating climate change: history, challenges, process and future directions," *WIREs Climate Change*, vol. 1, no. 1, pp. 31–53, 2010. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.11

[5] J. Rodgers and L. Bartram, "Exploring Ambient and Artistic Visualization for Residential Energy Use Feedback," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2489–2497, Dec. 2011, conference Name: IEEE Transactions on Visualization and Computer Graphics.

[6] M. L. Chalal, B. Medjdoub, N. Bezai, R. Bull, and M. Zune, "Visualisation in energy eco-feedback systems: A systematic review of good practice," *Renewable and Sustainable Energy Reviews*, vol. 162, p. 112447, Jul. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032122003537

[7] N. S. Mathews, S. Chimalakonda, and S. Jain, "AiR: An Augmented Reality Application for Visualizing Air Pollution," in *2021 IEEE Visualization Conference (VIS)*, Oct. 2021, pp. 146–150.

[8] N. G. Torres and P. E. Campbell, "Aire: visualize air quality," in *ACM SIGGRAPH 2019 Appy Hour*, ser. SIGGRAPH '19. New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 1–2. [Online]. Available: https://dl.acm.org/doi/10.1145/3305365.3329869

[9] O. Christmann, S. Fleury, J. Migaud, V. Raimbault, B. Poussard, T. Guitter, G. Gorisse, and S. Richir, "Visualizing the invisible: User-centered design of a system for the visualization of flows and concentrations of particles in the air," *Information Visualization*, vol. 21, no. 3, pp. 311–320, Jul. 2022, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1177/14738716221085965

[10] K. Ito, Y. Ban, and S. Warisawa, "AlteredWind: Manipulating Perceived Direction of the Wind by Cross-Modal presentation of Visual, Audio and Wind Stimuli," in *SIGGRAPH Asia 2019 Emerging Technologies*, ser. SA '19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 3–4. [Online]. Available: https://dl.acm.org/doi/10.1145/3355049.3360525

[11] J. Buchner, K. Buntins, and M. Kerres, "The impact of augmented reality on cognitive load and performance: A systematic review," *Journal of Computer Assisted Learning*, vol. 38, no. 1, pp. 285–303, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12617

[12] M. Venkatesan, H. Mohan, J. R. Ryan, C. M. Schürch, G. P. Nolan, D. H. Frakes, and A. F. Coskun, "Virtual and augmented reality for biomedical applications," *Cell Reports Medicine*, vol. 2, no. 7, p. 100348, Jul. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S266637912100197X

[13] T. Rist and M. Masoodian, "Promoting Sustainable Energy Consumption Behavior through Interactive Data Visualizations," *Multimodal Technologies and Interaction*, vol. 3, no. 3, p. 56, Sep. 2019, number: 3 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2414-4088/3/3/56

[14] A. Lavalle, M. A. Teruel, A. Maté, and J. Trujillo, "Improving Sustainability of Smart Cities through Visualization Techniques for Big Data from IoT Devices," *Sustainability*, vol. 12, no. 14, p. 5595, Jan. 2020, number: 14 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2071-1050/12/14/5595

[15] A. Al-Kababji, A. Alsalemi, Y. Himeur, R. Fernandez, F. Bensaali, A. Amira, and N. Fetais, "Interactive visual study for residential energy

consumption data," *Journal of Cleaner Production*, vol. 366, p. 132841, Sep. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652622024350

[16] A. Kollmuss and J. Agyeman, "Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior?" *Environmental Education Research*, vol. 8, no. 3, pp. 239–260, Aug. 2002. [Online]. Available: https://doi.org/10.1080/13504620220145401

[17] S. Prost, E. Mattheiss, and M. Tscheligi, "From Awareness to Empowerment: Using Design Fiction to Explore Paths towards a Sustainable Energy Future," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15. New York, NY, USA: Association for Computing Machinery, Feb. 2015, pp. 1649–1658. [Online]. Available: https://doi.org/10.1145/2675133.2675281

[18] D.-G. for Research and Innovation (European Commission), M. Breque, L. De Nul, and A. Petridis, *Industry 5.0: towards a sustainable, human centric and resilient European industry*. LU: Publications Office of the European Union, 2021. [Online]. Available: https://data.europa.eu/doi/10.2777/308407

[19] D. Romero, J. Stahre, T. Wuest, O. Noran, P. Bernus, F.-B. Fasth, Åsa, and D. Gorecky, "Towards an Operator 4.0 Typology: A Human-Centric Perspective on the Fourth Industrial Revolution Technologies," Oct. 2016.

[20] D. Maharjan, M. Agüero, D. Mascarenas, R. Fierro, and F. Moreu, "Enabling human–infrastructure interfaces for inspection using augmented reality," *Structural Health Monitoring*, vol. 20, no. 4, pp. 1980–1996, Jul. 2021, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1177/1475921720977017

[21] P. Barr, J. Noble, and R. Biddle, "Icons R icons," in *Proceedings of the Fourth Australasian user interface conference on User interfaces 2003 - Volume 18*, ser. AUIC '03. AUS: Australian Computer Society, Inc., Feb. 2003, pp. 25–32.

[22] A. Taulien, A. Paulsen, T. Streland, B. Jessen, S. Wittke, and M. Teistler, "A Mixed Reality Environmental Simulation to Support Learning about Maritime Habitats: An Approach to Convey Educational Knowledge With a Novel User Experience," in *Proceedings of Mensch und Computer 2019*, ser. MuC'19. New York, NY, USA: Association for Computing Machinery, Sep. 2019, pp. 921–925. [Online]. Available: https://doi.org/10.1145/3340764.3345382

[23] E. Fraj and E. Martinez, "Ecological consumer behaviour: an empirical analysis," *International Journal of Consumer Studies*, vol. 31, no. 1, pp. 26–33, 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1470-6431.2006.00565.x

[24] Z. Yiyuan, Y. Tangwen, D. Dayong, and F. Shan, "Using NASA-TLX to evaluate the flight deck design in Design Phase of Aircraft," *Procedia Engineering*, vol. 17, pp. 77–83, Jan. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877705811026890

[25] N. Bagher Zadeh Ansari, Léger, and M. Kersten-Oertel, "VentroAR: an augmented reality platform for ventriculostomy using the Microsoft HoloLens," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 0, no. 0, pp. 1–9, Dec. 2022. [Online]. Available: https://doi.org/10.1080/21681163.2022.2156394

[26] S. Beitzel, J. Dykstra, P. Toliver, and J. Youzwak, "Network Anomaly Analysis using the Microsoft HoloLens," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 62, no. 1, pp. 2094–2098, Sep. 2018. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1541931218621472

[27] E. Blevis, "Sustainable interaction design: invention & disposal, renewal & reuse," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '07. New York, NY, USA: Association for Computing Machinery, Apr. 2007, pp. 503–512. [Online]. Available: https://doi.org/10.1145/1240624.1240705

[28] B. Raghavan and D. Pargman, "Means and Ends in Human-Computer Interaction: Sustainability through Disintermediation," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, May 2017, pp. 786–796. [Online]. Available: https://doi.org/10.1145/3025453.3025542

[29] S. Scuri, M. Ferreira, N. Jardim Nunes, V. Nisi, and C. Mulligan, "Hitting the Triple Bottom Line: Widening the HCI Approach to Sustainability," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–19. [Online]. Available: https://doi.org/10.1145/3491102.3517518

[30] "Goal 12 | Department of Economic and Social Affairs." [Online]. Available: https://sdgs.un.org/goals/goal12

[31] E. R. Hilgard, "The trilogy of mind: Cognition, affection, and conation," *Journal of the History of the Behavioral Sciences*, vol. 16, no. 2, pp. 107–117, Apr. 1980. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/1520-6696(198004)16:2⟨107::AID-JHBS2300160202⟩3.0.CO;2-Y

[32] M. Ham, D. Mrčela, and M. Horvat, "INSIGHTS FOR MEASURING ENVIRONMENTAL AWARENESS," *Ekonomski vjesnik : Review of Contemporary Entrepreneurship, Business, and Economic Issues*, vol. 29, no. 1, pp. 159–176, Jun. 2016, publisher: Sveučilište J.J.Strossmayera u Osijeku, Ekonomski fakultet u Osijeku. [Online]. Available: https://hrcak.srce.hr/161021

[33] M. Endsley, "Situation awareness global assessment technique (SAGAT)," in *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, May 1988, pp. 789–795 vol.3.

[34] G. Salvendy and C. St, *Handbook of human factors and ergonomics fourth edition*, Jul. 2022.

[35] H. Brauers, "Natural gas as a barrier to sustainability transitions? A systematic mapping of the risks and challenges," *Energy Research & Social Science*, vol. 89, p. 102538, Jul. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214629622000457

[36] F. Durso, C. Hackworth, T. Truitt, J. Crutchfield, and D. Nikolić, "Situation Awareness As a Predictor of Performance in En Route Air Traffic Controllers," *Air Traffic Control Quarterly*, vol. 6, p. 15, Jan. 1999.

[37] M. R. Endsley, "A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM," *Human Factors*, vol. 63, no. 1, pp. 124–150, Feb. 2021, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/0018720819875376

[38] S. Selcon, R. Taylor, and E. Koritsas, "Workload or Situational Awareness?: TLX vs. SART for Aerospace Systems Design Evaluation," *Proceedings of the Human Factors Society Annual Meeting*, vol. 35, no. 2, pp. 62–66, Sep. 1991, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1518/107118191786755706

[39] M. R. Endsley, S. J. Selcon, T. D. Hardiman, and D. G. Croft, "A Comparative Analysis of Sagat and Sart for Evaluations of Situation Awareness," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*,

vol. 42, no. 1, pp. 82–86, Oct. 1998, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/154193129804200119

[40] K. van de Merwe, H. Dijk, and R. Zon, “Eye Movements as an Indicator of Situation Awareness in a Flight Simulator Experiment,” *The International Journal of Aviation Psychology*, vol. 22, pp. 78–95, Jan. 2012.

[41] J. Woodward and J. Ruiz, “Analytic Review of Using Augmented Reality for Situational Awareness,” *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2022, conference Name: IEEE Transactions on Visualization and Computer Graphics.

[42] A. Rowen, M. Grabowski, J.-P. Rancy, and A. Crane, “Impacts of Wearable Augmented Reality Displays on operator performance, Situation Awareness, and communication in safety-critical systems,” *Applied Ergonomics*, vol. 80, pp. 17–27, Oct. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000368701830766X

[43] M. P. Maloney, M. P. Ward, and G. N. Braucht, “A revised scale for the measurement of ecological attitudes and knowledge,” *American Psychologist*, vol. 30, pp. 787–790, 1975, place: US Publisher: American Psychological Association.

[44] S. J. Breckler and E. C. Wiggins, “Affect versus evaluation in the structure of attitudes,” *Journal of Experimental Social Psychology*, vol. 25, no. 3, pp. 253–271, May 1989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/002210318990022X

[45] M. Anderson, “New Ecological Paradigm (NEP) Scale,” *Berkshire Encyclopedia of Sustainability*, vol. 6, pp. 260–262, Jan. 2012.

[46] T. Milfont and J. Duckitt, “The Environmental Attitudes Inventory: A valid and reliable measure to assess the structure of environmental attitudes,” *Journal of Environmental Psychology*, vol. 30, pp. 80–94, Mar. 2010.

[47] M. Abbas and R. Singh, “A Survey of Environmental Awareness, Attitude and Participation Amongst University Students: A Case Study,” *International Journal of Science and Research (IJSR)*, vol. 3, pp. 1755–1760, May 2014.

[48] S. A. R. Khan, Z. Yu, and M. Umar, “How environmental awareness and corporate social responsibility practices benefit the enterprise? An empirical study in the context of emerging economy,” *Management of Environmental Quality: An International Journal*, vol. 32, no. 5, pp. 863–885,

Jan. 2021, publisher: Emerald Publishing Limited. [Online]. Available: https://doi.org/10.1108/MEQ-08-2020-0178

[49] D. A. R. Robledo and M. S. Prudente, ""A Virtual Fieldtrip": Effects of Google Earth Learning Activities (GELA) on Students' Environmental Awareness and Environmental Attitudes," in *2022 13th International Conference on E-Education, E-Business, E-Management, and E-Learning (IC4E)*, ser. IC4E 2022. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–8. [Online]. Available: https://doi.org/10.1145/3514262.3514293

[50] F. Pittarello and T. Pellegrini, "Designing and experimenting a web3d application for increasing the environmental awareness," in *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*, ser. CHItaly '13. New York, NY, USA: Association for Computing Machinery, Sep. 2013, pp. 1–9. [Online]. Available: https://doi.org/10.1145/2499149.2499167

[51] C. Boletsis, "The Gaia System: A Tabletop Projection Mapping System for Raising Environmental Awareness in Islands and Coastal Areas," in *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '22. New York, NY, USA: Association for Computing Machinery, Jul. 2022, pp. 50–54. [Online]. Available: https://doi.org/10.1145/3529190.3535340

[52] L. Fu, Z. Sun, L. Zha, F. Liu, L. He, X. Sun, and X. Jing, "Environmental awareness and pro-environmental behavior within China's road freight transportation industry: Moderating role of perceived policy effectiveness," *Journal of Cleaner Production*, vol. 252, p. 119796, Apr. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652619346669

[53] N. Castelli, S. Taugerbeck, M. Stein, T. Jakobi, G. Stevens, and V. Wulf, "Eco-InfoVis at Work: Role-based Eco-Visualizations for the Industrial Context," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. GROUP, pp. 02:1–02:27, Jan. 2020. [Online]. Available: https://doi.org/10.1145/3375182

[54] R. C. Atkinson and R. M. Shiffrin, "Human Memory: A Proposed System and its Control Processes," in *Psychology of Learning and Motivation*, K. W. Spence and J. T. Spence, Eds. Academic Press, Jan. 1968, vol. 2, pp. 89–195. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079742108604223

[55] J. Sweller, “Cognitive load during problem solving: Effects on learning,” *Cognitive Science*, vol. 12, no. 2, pp. 257–285, Apr. 1988. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0364021388900237

[56] F. Paas, A. Renkl, and J. Sweller, “Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture,” *Instructional Science*, vol. 32, no. 1, pp. 1–8, Jan. 2004. [Online]. Available: https://doi.org/10.1023/B:TRUC.0000021806.17516.d0

[57] J. Sweller, J. J. G. van Merriënboer, and F. Paas, “Cognitive Architecture and Instructional Design: 20 Years Later,” *Educational Psychology Review*, vol. 31, no. 2, pp. 261–292, Jun. 2019. [Online]. Available: https://doi.org/10.1007/s10648-019-09465-5

[58] J. Sweller, “Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load,” *Educational Psychology Review*, vol. 22, no. 2, pp. 123–138, Jun. 2010. [Online]. Available: https://doi.org/10.1007/s10648-010-9128-5

[59] S. Kalyuga, “Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction,” *Educational Psychology Review*, vol. 19, no. 4, pp. 509–539, Dec. 2007. [Online]. Available: https://doi.org/10.1007/s10648-007-9054-3

[60] W. Huang, P. Eades, and S.-H. Hong, “Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective,” *Information Visualization*, vol. 8, no. 3, pp. 139–152, Sep. 2009, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1057/ivs.2009.10

[61] P. Hoonakker, P. Carayon, A. P. Gurses, R. Brown, A. Khunlertkit, K. McGuire, and J. M. Walker, “Measuring workload of ICU nurses with a questionnaire survey: the NASA Task Load Index (TLX),” *IIE Transactions on Healthcare Systems Engineering*, vol. 1, no. 2, pp. 131–143, Apr. 2011. [Online]. Available: https://doi.org/10.1080/19488300.2011.609524

[62] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” in *Advances in Psychology*, ser. Human Mental Workload, P. A. Hancock and N. Meshkati, Eds. North-Holland, Jan. 1988, vol. 52, pp. 139–183. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166411508623869

[63] G. B. Reid and T. E. Nygren, “The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload,” in *Advances in Psychology*, ser. Human Mental Workload, P. A. Hancock

and N. Meshkati, Eds. North-Holland, Jan. 1988, vol. 52, pp. 185–218. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166411508623870

[64] T. E. Nygren, "Psychometric Properties of Subjective Workload Measurement Techniques: Implications for Their Use in the Assessment of Perceived Mental Workload," *Human Factors*, vol. 33, no. 1, pp. 17–33, Feb. 1991, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/001872089103300102

[65] D. Mutlu-Bayraktar, V. Cosgun, and T. Altan, "Cognitive load in multimedia learning environments: A systematic review," *Computers & Education*, vol. 141, p. 103618, Nov. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S036013151930171X

[66] R. Rydvanskiy and N. Hedley, "Mixed Reality Flood Visualizations: Reflections on Development and Usability of Current Systems," *ISPRS International Journal of Geo-Information*, vol. 10, no. 2, p. 82, Feb. 2021, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2220-9964/10/2/82

[67] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE Trans. Information Systems*, vol. vol. E77-D, no. 12, pp. 1321–1329, Dec. 1994.

[68] W. Hoenig, C. Milanes, L. Scaria, T. Phan, M. Bolas, and N. Ayanian, *Mixed reality for robotics*, Sep. 2015.

[69] C. Wyss, W. Bührer, F. Furrer, A. Degonda, and J. A. Hiss, "Innovative Teacher Education with the Augmented Reality Device Microsoft HoloLens—Results of an Exploratory Study and Pedagogical Considerations," *Multimodal Technologies and Interaction*, vol. 5, no. 8, p. 45, Aug. 2021, number: 8 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2414-4088/5/8/45

[70] T. Mahmood, W. Fulmer, N. Mungoli, J. Huang, and A. Lu, "Improving Information Sharing and Collaborative Analysis for Remote GeoSpatial Visualization Using Mixed Reality," in *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Oct. 2019, pp. 236–247, iSSN: 1554-7868.

[71] S. Utzig, R. Kaps, S. M. Azeem, and A. Gerndt, "Augmented Reality for Remote Collaboration in Aircraft Maintenance Tasks," in *2019 IEEE Aerospace Conference*, Mar. 2019, pp. 1–10, iSSN: 1095-323X.

[72] J. Rambach, G. Lilligreen, A. Schäfer, R. Bankanal, A. Wiebel, and D. Stricker, "A survey on applications of augmented, mixed and virtual reality for nature and environment," Aug. 2020. [Online]. Available: http://arxiv.org/abs/2008.12024

[73] S. Naret, T. Annop, S. Supharoek, and J. Thanadon, "Visualizing electrical energy consumption with virtual augmented technology and the internet of things," *IOP Conference Series: Earth and Environmental Science*, vol. 1094, no. 1, p. 012010, Sep. 2022, publisher: IOP Publishing. [Online]. Available: https://dx.doi.org/10.1088/1755-1315/1094/1/012010

[74] M. Card, *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, Jan. 1999.

[75] A. Treisman, "Preattentive processing in vision," *Computer Vision, Graphics, and Image Processing*, vol. 31, no. 2, pp. 156–177, Aug. 1985. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0734189X85800049

[76] R. Kosara, H. Hauser, and D. L. Gresh, "An Interaction View on Information Visualization," *Eurographics Association*.

[77] J. Moons and C. De Backer, "The design and pilot evaluation of an interactive learning environment for introductory programming influenced by cognitive load theory and constructivism," *Computers & Education*, vol. 60, no. 1, pp. 368–384, Jan. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131512001959

[78] R. Kosara, S. Miksch, H. Hauser, J. Schrammel, V. Seibert-Giller, and M. Tscheligi, "Useful Properties of Semantic Depth of Field for Better F+C Visualization," Aug. 2002.

[79] A. Lugmayr, Y. Lim, J. Hollick, J. Khuu, and F. Chan, "Financial Data Visualization in 3D on Immersive Virtual Reality Displays: A Case-Study for Data Storytelling and Information Visualization of Financial Data of Australia's Energy Sector," May 2019, pp. 118–130.

[80] E. Veas, R. Grasset, I. Ferencik, T. Grünewald, and D. Schmalstieg, "Mobile augmented reality for environmental monitoring," *Personal and Ubiquitous Computing*, vol. 17, no. 7, pp. 1515–1531, Oct. 2013. [Online]. Available: https://doi.org/10.1007/s00779-012-0597-z

[81] B. Shelton and N. Hedley, "Using augmented reality for teaching Earth-Sun relationships to undergraduate geography students," in *The First IEEE International Workshop Agumented Reality Toolkit,*, Sep. 2002, pp. 8 pp.–.

[82] K. Schroeder, B. Ajdadilish, A. P. Henkel, and A. Calero Valdez, "Evaluation of a Financial Portfolio Visualization using Computer Displays and Mixed Reality Devices with Domain Experts," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–9. [Online]. Available: https://doi.org/10.1145/3313831.3376556

[83] S. Zollmann, T. Langlotz, R. Grasset, W. H. Lo, S. Mori, and H. Regenbrecht, "Visualization Techniques in Augmented Reality: A Taxonomy, Methods and Patterns," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 9, pp. 3808–3825, Sep. 2021, conference Name: IEEE Transactions on Visualization and Computer Graphics.

[84] M. McGinity, "Immersive Media for Environmental Awareness," in *2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good)*, Mar. 2018, pp. 1–5.

[85] A. Kirk, *Data Visualisation: A Handbook for Data Driven Design*. SAGE, Jul. 2019.

[86] L. Bartram, J. Rodgers, and K. Muise, "Chasing the Negawatt: Visualization for Sustainable Living," *IEEE computer graphics and applications*, vol. 30, pp. 8–14, May 2010.

[87] A. Atkin, "Peirce's Theory of Signs," Oct. 2006, last Modified: 2022-08-04. [Online]. Available: https://plato.stanford.edu/entries/peirce-semiotics/?ref=superjump#SigEleSig

[88] H. Nagel, "Scientific Visualization versus Information Visualization," Jan. 2006.

[89] M. Kim, S. Yi, D. Jung, S. Park, and D. Seo, "Augmented-Reality Visualization of Aerodynamics Simulation in Sustainable Cloud Computing," *Sustainability*, vol. 10, no. 5, p. 1362, May 2018, number: 5 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2071-1050/10/5/1362

[90] C. Helbig, H.-S. Bauer, K. Rink, V. Wulfmeyer, M. Frank, and O. Kolditz, "Concept and workflow for 3D visualization of atmospheric data in a virtual reality environment for analytical approaches," *Environmental Earth Sciences*, vol. 72, pp. 3767–3780, Nov. 2014.

[91] K. Altmeyer, S. Kapp, M. Thees, S. Malone, J. Kuhn, and R. Brünken, "The use of augmented reality to foster conceptual knowledge acquisition in STEM labo-

ratory courses—Theoretical background and empirical results," *British Journal of Educational Technology*, vol. 51, Jan. 2020.

[92] M. Alonso-Rosa, A. Gil-de Castro, A. Moreno-Munoz, J. Garrido-Zafra, E. Gutierrez-Ballesteros, and E. Cañete-Carmona, "An IoT Based Mobile Augmented Reality Application for Energy Visualization in Buildings Environments," *Applied Sciences*, vol. 10, no. 2, p. 600, Jan. 2020, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-3417/10/2/600

[93] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Seven Guiding Scenarios for Information Visualization Evaluation," report, 2011. [Online]. Available: https://hal.inria.fr/hal-00723057

[94] C. Conati and H. Maclaren, "Exploring the role of individual differences in information visualization," in *Proceedings of the working conference on Advanced visual interfaces*, ser. AVI '08. New York, NY, USA: Association for Computing Machinery, May 2008, pp. 199–206. [Online]. Available: https://doi.org/10.1145/1385569.1385602

[95] M. Velez, D. Silver, and M. Tremaine, "Understanding visualization through spatial ability differences," in *VIS 05. IEEE Visualization, 2005.*, Oct. 2005, pp. 511–518.

[96] "What is User Centered Design?" [Online]. Available: https://www.interaction-design.org/literature/topics/user-centered-design

[97] C. Briggs, "Activity-Focused Design," May 2022. [Online]. Available: https://www.interaction-design.org/literature/article/activity-focused-design

[98] H. Dubberly, "Alan Cooper and the Goal Directed Design Process," *Journal of Design for the Network Economy*, vol. 1, no. 2, 2001.

[99] "What is Design Thinking?" 2016. [Online]. Available: https://www.interaction-design.org/literature/topics/design-thinking

[100] E. Wong, "What Are Wicked Problems and How Might We Solve Them?" Feb. 2022. [Online]. Available: https://www.interaction-design.org/literature/article/wicked-problems-5-steps-to-help-you-tackle-wicked-problems-by-combining-systems-thinking

[101] C. Courage and K. Baxter, Eds., *Understanding Your Users*, ser. Interactive Technologies. San Francisco: Morgan Kaufmann, Jan. 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9781558609358500609

[102] M. Hennink and B. N. Kaiser, "Sample sizes for saturation in qualitative research: A systematic review of empirical tests," *Social Science & Medicine*, vol. 292, p. 114523, Jan. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0277953621008558

[103] R. F. Dam and T. Y. Siang, "Affinity Diagrams: How to Cluster Your Ideas and Reveal Insights," May 2022. [Online]. Available: https://www.interaction-design.org/literature/article/affinity-diagrams-learn-how-to-cluster-and-bundle-ideas-and-facts

[104] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa

[105] C. A. Bolstad, J. M. Riley, D. G. Jones, and M. R. Endsley, "Using Goal Directed Task Analysis with Army Brigade Officer Teams," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, no. 3, pp. 472–476, Sep. 2002, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/154193120204600354

[106] A. Sharma, S. Nazir, and J. Ernstsen, "Situation awareness information requirements for maritime navigation: A goal directed task analysis," *Safety Science*, vol. 120, pp. 745–752, Dec. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753518316412

[107] "Gas Leak Detection | Portable | System | Manufacturer." [Online]. Available: https://new.abb.com/products/measurement-products/analytical/laser-gas-analyzers/advanced-leak-detection/abb-ability-portable-gas-leak-detection-solution

[108] "Section 114.04 - Uniform Gas Leak Classification, 220 Mass. Reg. 114.04 | Casetext Search + Citator." [Online]. Available: https://casetext.com/regulation/code-of-massachusetts-regulations/department-220-cmr-department-of-public-utilities/title-220-cmr-11400-uniform-gas-leaks-classification/section-11404-uniform-gas-leak-classification

[109] T. Log and W. B. Pedersen, "A Common Risk Classification Concept for Safety Related Gas Leaks and Fugitive Emissions?" *Energies*, vol. 12, no. 21, p. 4063, Jan. 2019, number: 21 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1996-1073/12/21/4063

[110] E. a. C. C. Canada, "Beaufort wind scale table," Feb. 2011, last Modified: 2017-03-07. [Online]. Available: https://www.canada.ca/en/environment-climate-change/services/general-marine-weather-information/understanding-forecasts/beaufort-wind-scale-table.html

[111] J. Rojas, "Learn How to Use Sketching as an Ideation Method," Nov. 2020. [Online]. Available: https://www.interaction-design.org/literature/article/etch-a-sketch-how-to-use-sketching-in-user-experience-design

[112] M. Maguire, *An exploration of low-fidelity prototyping methods for augmented and virtual reality*. Loughborough University, Jul. 2020. [Online]. Available: https://repository.lboro.ac.uk/articles/conference_contribution/An_exploration_of_low-fidelity_prototyping_methods_for_augmented_and_virtual_reality/11904909/1

[113] K. A. Schmal, C. Thomas, J. Cushing, and G. Orr, "Visualizing valley wind flow," in *ACM SIGGRAPH 2015 Posters*, ser. SIGGRAPH '15. New York, NY, USA: Association for Computing Machinery, Jul. 2015, p. 1. [Online]. Available: https://dl.acm.org/doi/10.1145/2787626.2792647

[114] "Microsoft HoloLens 2: Full Specification." [Online]. Available: https://vr-compare.com/headset/microsofthololens2

[115] S. Ruan, J. O. Wobbrock, K. Liou, A. Ng, and J. Landay, "Speech Is 3x Faster than Typing for English and Mandarin Text Entry on Mobile Devices," 2016.

[116] W. L. i. R.-B. U. Experience, "Between-Subjects vs. Within-Subjects Study Design." [Online]. Available: https://www.nngroup.com/articles/between-within-subjects/

[117] R. Kirollos and W. Merchant, "Comparing cybersickness in virtual reality and mixed reality head-mounted displays," *Frontiers in Virtual Reality*, vol. 4, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frvir.2023.1130864

[118] G. M. Sullivan and A. R. Artino, "Analyzing and Interpreting Data From Likert-Type Scales," *Journal of Graduate Medical Education*, vol. 5, no. 4, pp. 541–542, Dec. 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3886444/

[119] Barde and P. J. Barde, "What to use to express the variability of data: Standard deviation or standard error of mean? - PMC," *Perspect Clin Res*, vol. 3, no. 3, pp. 113–116, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3487226/

[120] B. Rosner, R. J. Glynn, and M.-L. T. Lee, "The Wilcoxon signed rank test for paired comparisons of clustered data," *Biometrics*, vol. 62, no. 1, pp. 185–192, Mar. 2006.

[121] I. C. Anaene Oyeka and G. U. Ebuh, "Modified Wilcoxon Signed-Rank Test," *Open Journal of Statistics*, vol. 02, no. 02, pp. 172–176, 2012. [Online]. Available: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/ojs.2012.22019

[122] F. Dallmeier, R. C. Szaro, A. Alonso, J. Comiskey, and A. Henderson, "Framework for Assessment and Monitoring of Biodiversity," in *Encyclopedia of Biodiversity (Second Edition)*, S. A. Levin, Ed. Waltham: Academic Press, Jan. 2013, pp. 545–559. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123847195003166

[123] S. G. Kwak and J. H. Kim, "Central limit theorem: the cornerstone of modern statistics," *Korean Journal of Anesthesiology*, vol. 70, no. 2, pp. 144–156, Apr. 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5370305/

[124] G. M. Sullivan and R. Feinn, "Using Effect Size—or Why the P Value Is Not Enough," *Journal of Graduate Medical Education*, vol. 4, no. 3, pp. 279–282, Sep. 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3444174/

[125] R. Rosenthal, *Meta-analytic procedures for social research, Rev. ed*, ser. Meta-analytic procedures for social research, Rev. ed. Thousand Oaks, CA, US: Sage Publications, Inc, 1991, pages: x, 155.

[126] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. New York: Routledge, Jul. 1988.

[127] P. S. Tsang and M. A. Vidulich, "Mental workload and situation awareness," in *Handbook of human factors and ergonomics, 3rd ed.* Hoboken, NJ, US: John Wiley & Sons, Inc., 2006, pp. 243–268.

[128] C. Tominski, P. Schulze-Wollgast, and H. Schumann, "3D Information Visualization for Time Dependent Data on Maps," in *Ninth International Conference on Information Visualisation (IV'05)*. London, England: IEEE, 2005, pp. 175–181. [Online]. Available: http://ieeexplore.ieee.org/document/1509075/

[129] Y. V. Chen, Z. C. Qian, and W. T. Lei, “Designing a Situational Awareness Information Display: Adopting an Affordance-Based Framework to Amplify User Experience in Environmental Interaction Design,” *Informatics*, vol. 3, no. 2, p. 6, Jun. 2016, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2227-9709/3/2/6

[130] D. S. Y. Tham, P. T. Sowden, A. Grandison, A. Franklin, A. K. W. Lee, M. Ng, J. Park, W. Pang, and J. Zhao, “A systematic investigation of conceptual color associations,” *Journal of Experimental Psychology. General*, vol. 149, no. 7, pp. 1311–1332, Jul. 2020.

[131] T. G. Holmes, “Eco-visualization: combining art and technology to reduce energy consumption,” in *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*, ser. C&C ’07. New York, NY, USA: Association for Computing Machinery, Jun. 2007, pp. 153–162. [Online]. Available: https://doi.org/10.1145/1254960.1254982

[132] R. Bühling, “Ästhetisch motivierte Gestaltung als persuasives Element in interaktiven Medien,” Ph.D. dissertation, Jul. 2015.

[133] M. Colley, B. Eder, J. O. Rixen, and E. Rukzio, “Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM, May 2021, pp. 1–11. [Online]. Available: https://dl.acm.org/doi/10.1145/3411764.3445351

[134] S. Roy, T. Smith, B. Coltin, and T. Williams, “I Need Your Help... or Do I?: Maintaining Situation Awareness through Performative Autonomy,” in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. Stockholm Sweden: ACM, Mar. 2023, pp. 122–131. [Online]. Available: https://dl.acm.org/doi/10.1145/3568162.3576954

[135] G. H. Bower, A. Sahgal, and D. A. Routh, “Affect and Cognition [and Discussion],” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 302, no. 1110,, pp. 387– 402, 1983. [Online]. Available: http://www.jstor.org/stable/2396001

[136] S. Duncan and L. F. Barrett, “Affect is a form of cognition: A neurobiological analysis,” *Cognition & emotion*, vol. 21, no. 6, pp. 1184–1211, Sep. 2007. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2396787/

[137] D. Shin, “The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI,” *International Journal of Human-Computer Studies*, vol. 146, p. 102551, Feb. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581920301531

[138] “Sample size calculator.” [Online]. Available: https://homepage.univie.ac.at/robin.ristl/samplesize.php?test=wilcoxon

[139] E. Veas, R. Grasset, E. Kruijff, and D. Schmalstieg, “Extended Overview Techniques for Outdoor Augmented Reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 4, pp. 565–572, Apr. 2012, conference Name: IEEE Transactions on Visualization and Computer Graphics.

[140] C. Conati, G. Carenini, E. Hoque, B. Steichen, and D. Toker, “Evaluating the Impact of User Characteristics and Different Layouts on an Interactive Visualization for Decision Making,” *Computer Graphics Forum*, vol. 33, no. 3, pp. 371–380, 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12393

Appendix A

User Requirements

A.1 Interview Questions

A.1.1 Main structured interview question set

1. What are the potential environmental benefits and impacts of surveying?
2. How important is it for you to contribute to sustainability?
3. Which environmental data is captured by each surveyor application which is relevant to the topic of sustainability?
4. How is the environmental data measured? (Where, When, By whom, Which unit/scale)
5. Which environmental aspects are currently reported?
6. How are they reported? (Where, When, By whom, Which unit/scale)
7. How is this environmental data visualized? (Units, grading)
8. What are special environmental regulations (for example from your company or the government) imposed on the gas industry?
9. Can you think of any environmental factors which are not yet captured through the surveying applications which could help in becoming more sustainable?
10. How could a new AR surveying application contribute to more sustainability?

A.1.2 More in-depth unordered interview question set

1. What is the relationship between the gas concentration and the flux rate (meters/min or kg/min)?

2. How does the gas concentration relate to the size of a leak?
3. Is the methane concentration or flux rate important for the prioritization of a leak?
4. How does the flux rate relate to the carbon footprint? How is the carbon footprint calculated based on the flux rate?
5. What is ethane?
6. What is the ethane threshold which can indicate if a gas is biogenic or thermogenic?
7. How does a leak affect soil and water? (Contamination)
8. How is H2S relevant for the environment? (Toxicity, Contamination)
9. Do you have any emission grading scheme that you can share with us for the methane concentration or leak rate?

A.2 Affinity Map

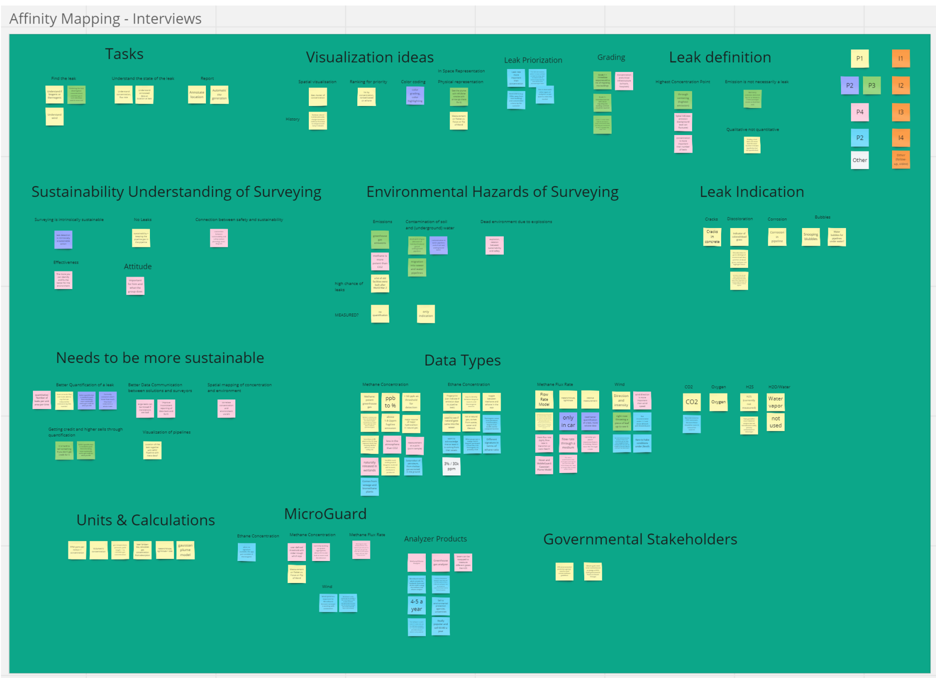
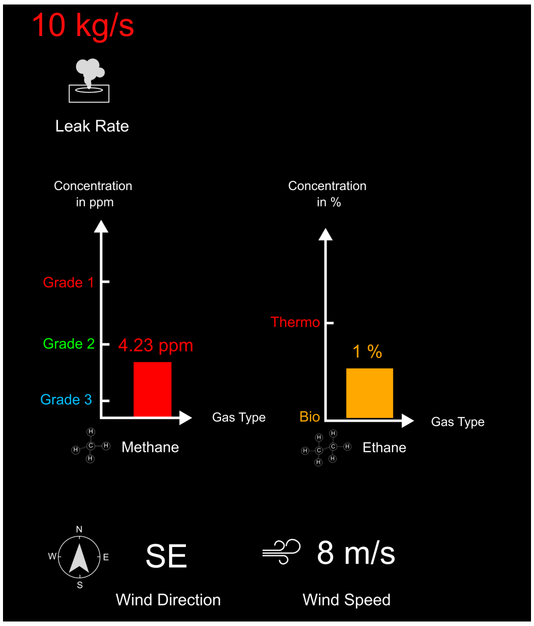


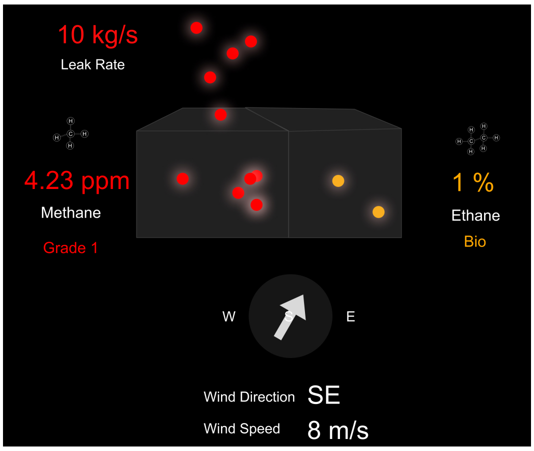
Figure A.1: Affinity map which displays the themes discovered through the initial expert interviews, from the used data types to a sustainability understanding, needs and environmental hazards.

Appendix B

Design

B.1 Visualization Designs in Figma

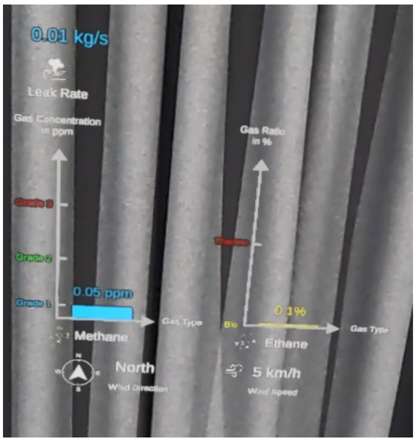




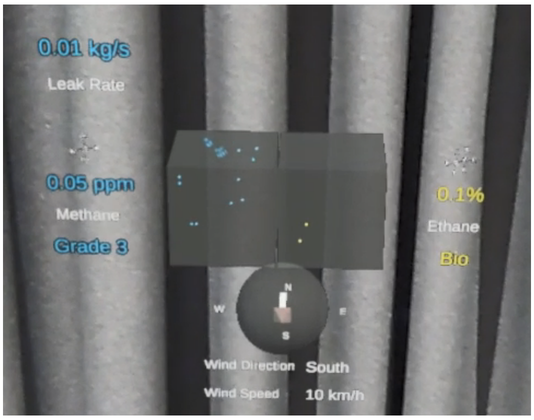
(a): statistical, iconographic 2D **(b):** physical 3D

Figure B.1: 2D (left) and 3D (right) visualizations designed in Figma.

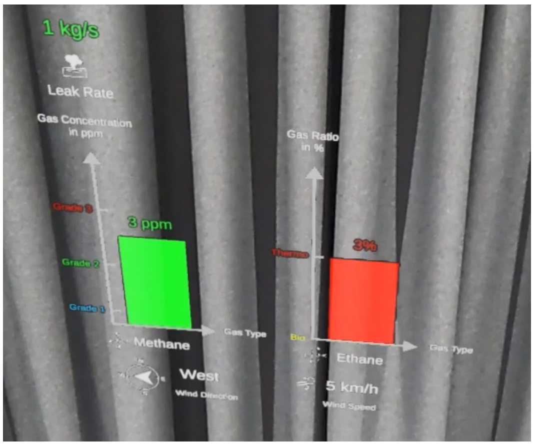
B.2 Threshold Levels Through the HoloLens 2



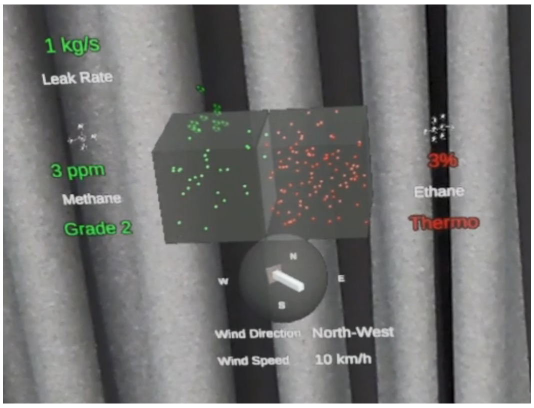
(a): 2D - Low Data Level



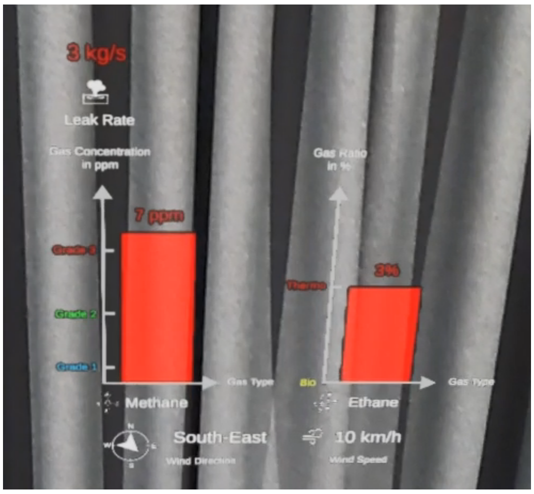
(b): 3D - Low Data Level



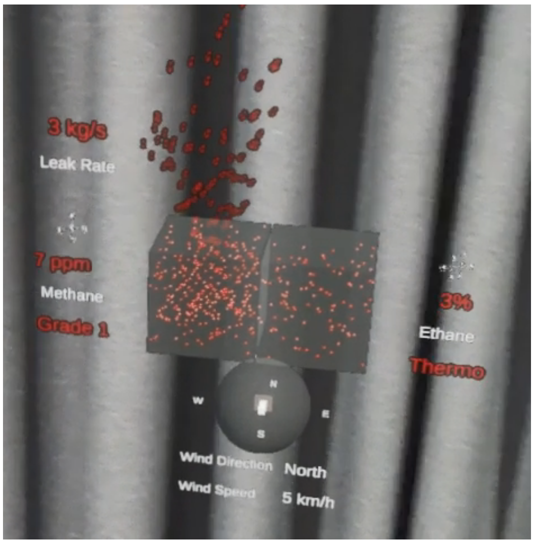
(c): 2D - Medium Data Level



(d): 3D - Medium Data Level



(e): 2D - High Data Level



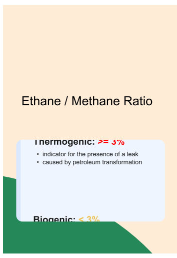
(f): 3D - High Data Level

Figure B.2: Threshold levels of both 2D (left) and 3D (right) visualizations through the HoloLens 2.

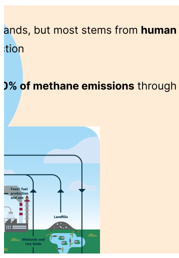
Appendix C

Survey and Usability Study

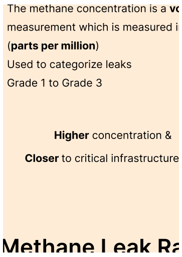
C.1 Surveying Rules



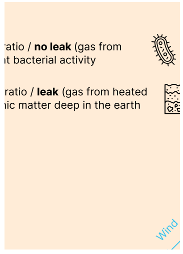
(a): Surveying rules for five data types: Methane concentration, leak rate, ethane / methane ratio, wind direction and wind speed.



(b): Introduction into the surveying rules (page 1).



(c): Introduction into the surveying rules (page 2).



(d): Introduction into the surveying rules (page 3).

C.2 SAGAT Questions

The SAGAT questions for the sub-goal of identifying a leak can be found in table C.3. The questions for sub-goal 2 of understanding the short-term, local impact can be viewed in table C.4. Finally, table C.5 contains the SAGAT questions for the third sub-goal of understanding the long-term, global impact.

C.3 Questionnaire Questions

C.3.1 Demographic Information

1. What do you identify as?

2. What is your age?

3. How much experience do you have with gas leak surveying?

4. How much experience did you have with a HoloLens or other mixed-reality experiences before conducting the experiment?

C.3.2 Cognitive Load: NASA-TLX

(7-point Likert scale, from very low to very high)

1. How mentally demanding was the task?
2. How physically demanding was the task?
3. How hurried or rushed was the pace of the task?
4. How successful were you in accomplishing what you were asked to do?
5. How hard did you have to work to accomplish your level of performance?
6. How insecure, discouraged, irritated, stressed and annoyed were you?

C.3.3 Affective Environmental Awareness

The affective environmental awareness questions used in the evaluation can be found in table C.1.

C.3.4 Conative Environmental Awareness

The conative environmental awareness questions used in the evaluation can be found in table C.2.

C.4 Post-Interview

C.4.1 Visualization Concept 1

1. How understandable was the first visualization?
2. How mentally demanding was the first visualization?
3. How could the first visualization be improved?
4. How useful would the first visualization be in your work?
5. How much did the visualization make you aware and feel responsible for the environmental impact of a leak?

Before (from Fraj et al. [23])	**After (adjustment to surveying context)**
It frightens me to think that much of the food I eat is contaminated with pesticides.	It frightens me to think that so much methane is emitted to the atmosphere.
It genuinely infuriates me to think that the government doesn't do more to help control pollution of the environment.	It genuinely infuriates me to think that the gas industry does not do more to help control global warming.
I become incensed when I think about the harm being done to plant and animal life by pollution.	I become incensed when I think about the harm being done to the climate by leaking gas.
I get depressed on smoggy days.	OMITTED
When I think of the ways industries are polluting, I get frustrated and angry.	When I think of the ways industries are emitting methane, I get frustrated or angry.
The whole pollution issue has never upset me too much since I feel it's somewhat overrated.	The whole methane emissions issue has never upset me too much since I feel it's somewhat overrated.
I rarely ever worry about the effects of smog on myself and family.	I rarely ever worry about the effects of methane emissions on myself and my family.

Table C.1: Previous affective environmental awareness questions from Fraj et al. [23] (left) and the adjusted ones which fit the surveying context (right).

C.4.2 Visualization Concept 2

1. How understandable was the second visualization?

2. How mentally demanding was the second visualization?

3. How could the second visualization be improved?

4. How useful would the second visualization be in your work?

5. How much did the visualization make you aware and feel responsible for the environmental impact of a leak?

Before (from Fraj et al. [23])	After (adjustment to surveying context)
I'd be willing to ride a bicycle or take the bus to work in order to reduce air pollution	I'd be willing to diligently do my job to reduce methane emissions.
I would be willing to use a rapid transit system to help to reduce air pollution.	I would be willing to document the environmental impact of a leak as much as possible to help reduce methane emissions.
I would donate a day's pay to a foundation to help improve the environment.	OMITTED
I would be willing to stop buying products from companies guilty of polluting the environment, even though it might be inconvenient.	I would be willing to stop ignoring small leaks which are environmentally harmful even though they might be difficult to find.
I'd be willing to write my congressman weekly concerning ecological problems.	I'd be willing to report the emitted greenhouse gasses of every leak to the responsible person for monitoring the leaks.
I wouldn't go house to house to distribute literature on the environment.	I wouldn't go to my colleagues to share insights about the environmental impact of leaks.
I would not be willing to pay a pollution tax even if it would considerably decrease the smog problem.	I would not be willing to take more time to document the leak with pictures and videos even if it would considerably help the repair of the leak.

Table C.2: Previous conative environmental awareness questions from Fraj et al. [23] (left) and the adjusted ones which fit the surveying context (right).

C.4.3 Comparison of Visualizations

1. Which visualization was more helpful and why?

    (a) Which made you more aware of the situation?

    (b) Which was less mentally demanding?

2. What do you think about. . .

(a) ...the methane and ethane concentration visualizations?

(b) ...wind direction and speed visualizations?

(c) ...leak rate visualizations?

3. Do you have any final thoughts you want to share on your overall experience?

C.5 Setup

Figure C.1: Researcher carrying a physical wand prop which the participants could hold during the experiment to imagine the visualization at the tip of a surveying wand.

	Description	SA Question	SA Level 1	SA Level 2	SA Level 3
1a	medium methane biogenic ethane strong wind wind from west medium leak rate	1: What is the current wind speed? 2: Is there a chance that the main leak is further away? 3: Stay or continue moving?	1: 5 km/h 2: **40 km/h**	1: **Yes** 2: No	1: Stay 2: **Continue moving**
1b	medium methane thermogenic ethane low wind wind from west medium leak rate	""	1: **5 km/h** 2: 40 km/h	1: Yes 2: **No**	1: **Stay** 2: Continue moving
2a	medium methane thermogenic ethane low wind wind from south-east medium leak rate	1: What is the current wind direction? 2: Is there a chance that a leak is nearby? 3: Move towards the wind or orthogonal to it?	1: North-West 2: **South-East**	1: **Yes** 2: No	1: Orthogonal to the wind 2: **Towards the wind**
2b	low methane biogenic ethane medium wind wind from north-west low leak rate	""	1: **North-West** 2: South-East	1: Yes 2: **No**	1: **Orthogonal to the wind** 2: Towards the wind
3a	high methane thermogenic ethane low wind wind from north high leak rate	1: What is the current ethane level? 2: Is this gas thermogenic or biogenic? 3: Does this pipeline have a leak?	1: 0.1 % 2: **3 %**	1: **Thermogenic** 2: Biogenic	1: **Yes** 2: No
3b	low methane biogenic ethane low wind wind from north low leak rate	""	1: **0.1 %** 2: 3 %	1: Thermogenic 2: **Biogenic**	1: Yes 2: **No**

Table C.3: SAGAT Questions for sub-goal 1: Identify a leak.

	Description	SA Question	SA Level 1	SA Level 2	SA Level 3
1a	high methane thermogenic ethane medium wind wind from east high leak rate	1: What is the current methane level? 2: Is it only a fugitive emission or is it an emission with a safety risk? 3: Should this leak have a high prioritization for repair?	1: 3 ppm 2: **7 ppm**	1: Fugitive emission 2: **Emission with safety risk**	1: **High prioritization** 2: Medium prioritization
1b	high methane thermogenic ethane medium wind wind from east high leak rate	””	1: **3 ppm** 2: 7 ppm	1: **Fugitive emission** 2: Emission with safety risk	1: High prioritization 2: **Medium prioritization**
2a	high methane thermogenic ethane low wind wind from east high leak rate	1: What is the current methane level? 2: What is the grade of the leak based on its concentration? (Note: The leak is close to critical infrastructure) 3: Can this leak be approached by the repair team without any safety measurements?	1: **0.05 ppm** 2: 7 ppm	1: **Grade 3** 2: Grade 1	1: No 2: **Yes**
2b	low methane biogenic ethane low wind wind from east low leak rate	””	1: 0.05 ppm 2: **7 ppm**	1: Grade 3 2: **Grade 1**	1: **No** 2: Yes

Table C.4: SAGAT questions for sub-goal 2: Understand the short-term, local impact.

	Description	SA Question	SA Level 1	SA Level 2	SA Level 3
1a	high methane thermogenic ethane medium wind wind from south high leak rate	1: What is the current leak rate? 2: How fast is the leak rate? 3: Should this leak have a high prioritization for repair due to its environmental impact?	1: **3 kg/s** 2: 0.01 kg/s	1: Fugitive emission 2: **Emission with safety risk**	1: **High prioritization** 2: Low prioritization
1b	low methane biogenic ethane medium wind wind from south high leak rate	””	1: 3 kg/s 2: **0.01 kg/s**	1: **Fugitive emission** 2: Emission with safety risk	1: High prioritization 2: **Low prioritization**

Table C.5: SAGAT questions for sub-goal 3: Understand the long-term, global impact.