University of Namur

DOCTOR OF SCIENCES

Determining Design Guidelines for Interface Elements for Immersive Augmented Reality

Cauz, Maxime

Award date:
2024

Awarding institution:
University of Namur

Determining Design Guidelines for Interface Elements for Immersive Augmented Reality

Maxime Cauz

Jury

Prof. Marcos Serrano
University of Toulouse (Toulouse-III-Paul-Sabatier), France
Prof. Thierry Dutoit
University of Mons (UMons), Belgium
Prof. Wafa Hammedi
University of Namur (UNamur), Belgium
Prof. Anthony Cleve
University of Namur (UNamur), Belgium
Prof. Wim Vanhoof (President)
University of Namur (UNamur), Belgium
Prof. Bruno Dumas (Supervisor)
University of Namur (UNamur), Belgium

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the subject of Computer Science

Supervised by Prof. Bruno Dumas

University of Namur
PReCISE Research Center

UNIVERSITÉ DE NAMUR
FACULTÉ D'INFORMATIQUE

Cover design: ©Presses universitaires de Namur

Rue Grandgagnage 19
B-5000 Namur (Belgium)
pun@unamur.be - www.pun.be

Registration of copyright: D/2024/1881/5
ISBN: 978-2-39029-195-4
Printed in Belgium.

“Perhaps you are the hero of another story, your own, that a reader reads in a world more real than yours.”

Christian Grenier, Virus L.I.V. 3 ou la Mort des livres

ABSTRACT

The fourth industrial revolution tries to respond to the need for innovative and specific products and services in a competitive market by integrating digital technologies [Barata, 2021; Davies, 2015]. Other fields have also undergone a transformation similar to the industrial world, such as the art and culture, medical and administrative sectors. Nevertheless, the voluminous nature and intricate structure of data necessitate the adoption of novel methodologies for visualization and interaction, commonly referred to as the challenges associated with Big Data. In addition, the context in which the information occurs becomes essential in understanding it. This context can be partly or totally real. Then, one challenge is to reduce the virtual and the real boundary to provide the information at the right place and time [Barata, 2021]. To this end, **Augmented Reality (AR)** technologies have taken part in these transformations to combine the virtual and real worlds. It consists of overlaying a virtual world over the real world seen by the user to display additional information in relation or not to their environment. In other words, it allows immediate access to the user's context information by placing information directly in the real world.

A major challenge exposed by the immersive AR technologies is the **User Interface** with the switch of paradigm in terms of workspace [Billinghurst, 2021; Kim et al., 2018; Ong et al., 2008]. Where on desktop and mobile, the workspace is in 2D and limited to the size of the screen, in immersive AR, the workspace is in 3D and limited to all the environment (i.e., unlimited in terms of space, but limited by the walls and other elements that compose the environment) [Chandler et al., 2015]. Although that presents some advantages, such as reducing the cognitive load on 3D visualization, it means rethinking the place and usefulness of 2D and 3D interfaces, including new constraints related to the real environment, such as depth, occlusion, and situated and embedded visualizations [Ens et al., 2021]. Indeed, the principle of AR is to place information directly onto the real environment. So, there is a vital need to identify good and bad design practices to prevent cognitive and visual overload, as well as the generation of overwhelming environments.

Within research possibilities offered by this challenge, the research presented in this doctoral thesis will focus on visualization as a support to the user's task through immersive AR with Head-Mounted Display (HMD). This main objective was pursued by delving into four design levels, aware of the contextual factors (e.g., users, environment, tasks) intrinsic to the application's deployment. At the asset level (e.g., text, image, video, and 3D models), we conducted a literature review to identify text parameters, contextual constraints impacting these parameters, and, thus, guidelines to respect to ensure text readability. At the canonical tasks level (i.e., tasks indivisible into further subtasks), we conducted a second literature review to determine visualization idioms, tasks to evaluate them, and, thus, the best visualizations idioms by tasks for navigation tasks to off-screen real and virtual point of interest (POI). At the application level, we studied two use cases: on the one hand, the support of a web application for text analysis with experts in humanities using metaphors and, on the other hand, a fully immersive remote collaboration application in industrial maintenance. Last, at the cross-application level, we explored strategies to mitigate information clutter in a Pervasive Augmented Reality (PAR) scenario. Ultimately, this thesis makes fifteen contributions to research and various practice domains. In addition, this thesis is written from and extends seven publications and five student works.

Keywords: Augmented Reality, Mixed Reality, Text, Readability, Navigation, Metaphors, Collaboration, Information Clutter, Pervasive Augmented Reality.

RÉSUMÉ

La quatrième révolution industrielle tente de répondre au besoin de produits et de services innovants et spécifiques dans un marché concurrentiel en intégrant les technologies numériques [Barata, 2021; Davies, 2015]. D'autres domaines ont également connu une transformation semblable à celle du monde industriel, tels que les secteurs de l'art et de la culture, de la médecine et de l'administration. Néanmoins, la nature volumineuse et la structure complexe des données nécessitent l'adoption de nouvelles méthodologies de visualisation et d'interaction, communément appelées les défis associés au "Big Data". En outre, le contexte dans lequel l'information se produit devient essentiel pour la comprendre. Ce contexte peut être partiellement ou totalement réel. Par conséquent, l'un des défis consiste à réduire la frontière entre le virtuel et le réel afin de fournir l'information au bon endroit et au bon moment [Barata, 2021]. À cette fin, les technologies de la **Réalité Augmentée (RA)** ont participé à ces transformations pour combiner les mondes virtuel et réel. Il s'agit de superposer un monde virtuel au monde réel vu par l'utilisateur afin d'afficher des informations supplémentaires en relation ou non avec son environnement. En d'autres termes, elle permet un accès immédiat aux informations contextuelles de l'utilisateur en plaçant des informations directement dans le monde réel.

L'un des principaux défis posés par les technologies de RA est l'**interface utilisateur**, avec le changement de paradigme en termes d'espace de travail [Billinghurst, 2021; Kim et al., 2018; Ong et al., 2008]. Alors que sur les ordinateurs de bureau et les téléphones portables, l'espace de travail est en 2D et limité à la taille de l'écran, dans l'environnement immersif de la RA, l'espace de travail est en 3D et limité à l'ensemble de l'environnement (c'est-à-dire illimité en termes d'espace, mais limité par les murs et les autres éléments qui composent l'environnement) [Chandler et al., 2015]. Bien que cela présente certains avantages, comme la réduction de la charge cognitive sur la visualisation 3D, cela implique de repenser la place et l'utilité des interfaces 2D et 3D, en incluant de nouvelles contraintes liées à l'environnement réel, comme la profondeur, l'occlusion, et les visualisations situées et embarquées [Ens et al., 2021]. En effet, le principe de la RA est de placer l'information directement dans l'environnement réel. Il est donc essentiel d'identifier les bonnes et les mauvaises pratiques en matière de conception afin de prévenir la surcharge cognitive et visuelle, ainsi que la création d'environnements accablants.

Dans le cadre des possibilités de recherche offertes par ce défi, la recherche présentée dans cette thèse de doctorat se concentrera davantage sur la visualisation en tant que soutien à la tâche de l'utilisateur par le biais de la RA immersives avec des casques montés sur tête. Cet objectif principal a été poursuivi en examinant quatre niveaux de conception, conscients des facteurs contextuels (par exemple, les utilisateurs, l'environnement, les tâches) intrinsèques au déploiement de l'application. Au niveau des entités (par exemple, texte, image, vidéo et modèles 3D), nous avons procédé à une revue de la littérature afin d'identifier les paramètres du texte, les contraintes contextuelles ayant un impact sur ces paramètres et, par conséquent, les lignes directrices à respecter pour garantir la lisibilité du texte. Au niveau des tâches canoniques (c'est-à-dire les tâches indivisibles en sous-tâches), nous avons effectué une deuxième revue de la littérature pour déterminer les idiomes de visualisation, les tâches pour les évaluer et, par conséquent, les meilleurs idiomes de visualisation par tâches pour les tâches de navigation vers des points d'intérêt réels et virtuels hors-écran. Au niveau applicatif, nous avons étudié deux cas d'utilisation : d'une part, le soutien d'une application web pour l'analyse de texte avec des experts en sciences humaines utilisant des métaphores et, d'autre part, une application de collaboration à distance totalement immersive dans le domaine de la maintenance industrielle. Enfin, au niveau inter-applications, nous avons exploré des stratégies pour réduire l'encombrement de l'information dans un scénario de RA permanente. En fin de compte, cette thèse apporte quinze contributions à la recherche et à divers domaines de pratique. En outre, cette thèse est rédigée à partir de sept publications et de cinq travaux d'étudiants, qu'elle prolonge.

Mots clés : Réalité Augmentée, Réalité Mixte, Texte, Lisibilité, Navigation, Métaphores, Collaboration, Encombrement de l'information, Réalité Augmentée permanente.

ACKNOWLEDGEMENTS

First, I would like to thank Bruno for allowing me to join his research team, first for my Master's thesis and then as a doctoral student. We have worked together for over four years and enjoy confronting our ideologies. They are so similar and yet so different. I have the framework, support, and advice to complete this thesis successfully. For this and your thoughtfulness, I would like to thank you.

Second, I thank Antoine C., Xavier, and Paul for their support and guidance during this thesis. You played your role as senior supervisors to perfection, even though you had no obligation to do so. Antoine, I do not think I could have deserved a better research mentor. Your efficiency and eccentricity meant that I learned a great deal from you. I would also like to thank you for all those meals together when you always took the time to listen to me in the best and worst times. Xavier, I know you are underestimating the mentoring role you have played for me since the start of my higher education until now. You paved my way into research, and I thank you for giving me a sense of fulfillment that I would not trade for anything. Thanks also for our meetings at the Sanga. Paul, you came to my rescue when my passion for mathematics could have engulfed me. Even after your return to Brittany, you continued to support me in this project. I also enjoyed our noodle meals and discussions on the difference between a sea and an ocean. Thank you, Paul.

Next, I would like to thank the jury for taking the time to read and allowing me to present my thesis. I thank Wim Vanhoof, Anthony Cleve, Wafa Hammedi, Marcos Serrano, and Thierry Dutoit. Wim, thank you for sharing your research path to help me see my future more clearly.

What would four years of research be like if it were not for my office and team colleagues? Julie, Julien, Jérôme, Thibaut, Elise, Liesbet, Santiago, Robin, Juliette and Sean. Our discussions and deliriums kept me smiling and happy to come to the office. Our meetings and collaborations have also enabled me to progress on this thesis project, and I hope I have been able to support you in yours, too. Julie, thank you for your advice and for showing me what it is like to hold two theses and a hundred projects simultaneously. Julien, you arrived on the effata project a little before I left, but what a pleasure to have someone to share code with. I still laugh at our faces after working out the Shock Wave algorithm. It is a good thing your office is always across the hall from mine. Jérôme, I would like to thank you for the Jérôme's sound box that misses me and for the kind words you have exchanged with me across the wall that now separates us. Thibaut, thank you for giving me the strength to explore Pervasive Augmented Reality. Mastering immersive technologies is no easy task, and we have seen the proof. We have learned so much, and our discouragements about immersive technologies are aligned. Elise, thank you for this calm temperament in the office and for the occasional peaks of "irritation" that make me laugh like the whole office.

Of course, I must also thank colleagues with whom I could work or who supported me throughout my research. I am thinking, in particular, of Babette, Fanny, Sophie, and Manelle. Babette, the mother of the faculty, we can thank you for a thousand things; you give so much of yourself for us. Thank you for this complicity and our long discussions to complain, which gives us a world of good. Fanny, I loved working with you as a scientific secretary. Your kindness and organization made the job easy to handle. Manelle and Sophie, thank you for all our discussions and encouragement for our theses.

Not least, I have to thank my close friends who put up with me for hours and days talking about research: Antoine J., Timothée, Mathieu, Christophe E., Hugues, Guillaume, Romain, Christophe A., Nicolas, Sixtine, Denis, Florian, Jérémie, Jérémy, Pierre-Alexandre, Gaspard, Alois, Titouan, Loreline, Louis, Agata and all those I might have forgotten. A special dedication to Antoine for proofreading my thesis.

To conclude, I would like to thank my family. My family-in-law with Jean-Luc, Catherine, Benoit, Madeleine, Marie, Loïc, Elisabeth, Arthur, Simon, and Valère. But above all, my family with Jean-Marie, Carole, Jérôme, Amélie, Alice, Adèle, Marine, Raphaël, Zoé, and my niece Alba. Mom and Dad, you have given me so much over the years that I will never know how to thank you enough. Finally, Hélène, I would like to thank you for being in my life and for supporting me in this project that came with me. Our relationship survived those difficult moments of writing when I had to work a lot, but I am convinced it came out even stronger.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Societal and Historical Context

The fourth industrial revolution is the representation, since late 2011, of society's digital evolution in the industrial sector [Barata, 2021; Davies, 2015]. It tries to respond to the need for innovative and specific products and services in a competitive market by integrating digital technologies, such as Cloud Computing, Autonomous Robots, and the Internet of Things. These changes involve real-time monitoring that results in the accumulation of data intended for improving sustainability, such as accident prevention and continuous business improvements [Barata, 2021]. Moreover, this revolution is also trying to support the personalization of products and services for customers [Davies, 2015], which increases the complexity of existing chains and related documentation. Furthermore, other media are added to the usual text and image documentation [Gattullo et al., 2017], such as video for step-by-step guidance or 3D models, to give readers freedom of handling and a better understanding of how parts fit together. Of course, all these transformations are part of continuous and long-term processes involving various stakeholders [Barata, 2021]. First, there is a need for specialized expertise to install and maintain such systems that may force a company to subcontract externalizing skills. Second, it introduces a personnel turnover attributable to the prolonged nature of this transformative undertaking [Davies, 2015].

Today, other fields have also undergone a transformation similar to the industrial world. For instance, the art and culture sector is pursuing more immersive and interactive experiences. The medical sector strives to deliver better healthcare by improving equipment and helping medical staff make decisions through artificial intelligence and interactive visualizations updated in real-time with patient data. Administrations are digitalizing their services for greater traceability and access to data. Nevertheless, the voluminous nature and intricate structure of data necessitate the adoption of novel methodologies for visualization and interaction, commonly referred to as the challenges associated with Big Data. In addition, the context in which the information occurs becomes essential in understanding it. This context can be partly or totally real. Therefore, one challenge is to reduce the virtual and the real boundary to provide the information at the right place and time [Barata, 2021].

Mobile and wearable technologies partially meet the needs of these transformations. They allow access to up-to-date data, update it when necessary, visualize complex representations of information such as graphs, take photos and videos, etc., without moving from the site. Nonetheless, these technologies present some limitations in common with conventional computing on a desktop station. First, users need to switch focus between the virtual content and the real environment. In the case of an assembly task, for instance, workers must successively turn their attention to the screen to learn about the task to complete, make an effort to associate the information with the current state of the environment, perform the task, and then turn their attention back to the screen and interact with it to go to the next step. Second, some visualization idioms also present significant constraints on a 2D screen. This is the case, for example, with 3D visualizations, which can induce a significant cognitive load, conscious attention, or time costs due to occlusion, perspective distortion, shadows and lighting, familiar size, stereoscopic disparity, and others [Munzner, 2015]. Third, with so much information to display, the virtual workspace quickly becomes too small, and the interactions are poorly adapted to efficient and intuitive navigation. To this end, **Mixed Reality (MR)** (a.k.a. **Extended Reality (xR)**) technologies have also taken part in these transformations to combine the virtual and real worlds in one. These can be divided into two sub-categories: on the one hand, **Virtual Reality (VR)** which consists of immersing one or more user's senses in a virtual world built by a computer, and on the other hand, **Augmented Reality (AR)** which consists in overlaying a virtual world over the real world seen by the user to display additional information in relation or not with their environment. The second allows immediate access to the user's context information by placing information directly in the real world. Both extend the virtual workspace to all the space around the user. While mobile computing and the Internet of Things lead the **everywhere, every time** and **all-connected** paradigms, **Augmented Reality (AR) breaks the frontier between both virtual and real worlds**. Although this is less the case with hand-held smartphones, Head-Mounted Displays (HMDs) will allow, with time, users to forget that they use a device to consult virtual information. The community named this approach immersive AR, where the real world and the virtual environment tend to be more and more indistinguishable.

According to Joshi et al. [2019], the origins of VR can be traced back to the work of Heilig [1961] with his invention, the “Sensorama Simulator.” However, within the scientific community, the more widely accepted starting point is Sutherland [1965]'s essay and, a few years later, his "Sword of Damocles" (see Figure 1.1) [Sutherland, 1968], which is considered the earliest example of HMD. However, the term VR itself was coined by Lanier [1992], as noted by Joshi et al. [2019]. Additionally, the term AR was introduced by Caudell and Mizell [1992], as stated by Sereno et al. [2022] and Van Krevelen and Poelman [2010].

AR is applied to various domains. Azuma [1997] determined, in his survey, its use in the medical, industry (manufacturing, repair, robot path planning, and military

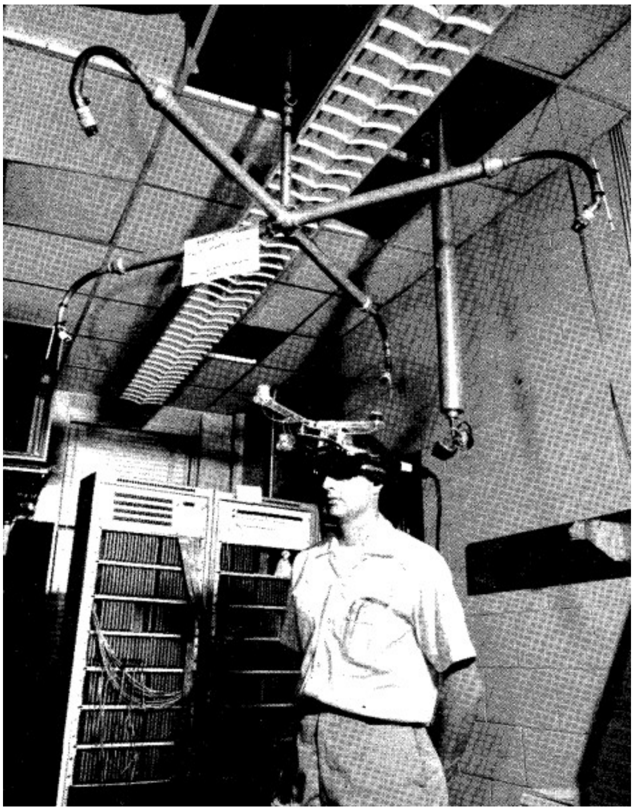


Figure 1.1: Illustration of the "Sword of Damocles" by Sutherland. (Source: Sutherland [1968])

aircraft), and entertainment domains. A few years later, Ong et al. [2008] extended the usage in the industry to the assembly, service, inspection, product development, manufacturing layout, and telerobotics. From their side, Van Krevelen and Poelman [2010] included personal information (assistance, advertising, navigation, and tourism), collaboration, education and training. From there, AR finds interest in most of the major domains of society. However, it was only around 2016 that AR was making its mark on the public with a diversification of smartphone applications [Vyas and Bhatt, 2017]. Overall, what is certain is that the industrial sector is today's favorite use case for experiencing AR applications [Bottani and Vignali, 2019; Cárdenas-Robledo et al., 2022; Dey et al., 2018; Schlosser et al., 2021]. In addition, research on this subject is relatively recent, following a hardware boom about ten years ago with the release of the Oculus Rift DK1 in 2013.

1.2 General Challenges of Immersive Augmented Reality

From the perspective of immersive technologies, taking an interest in AR means taking an interest in the technologies used for VR by extending them, on the one hand, with additional device technologies, and on the other, by adding a non-negligible set

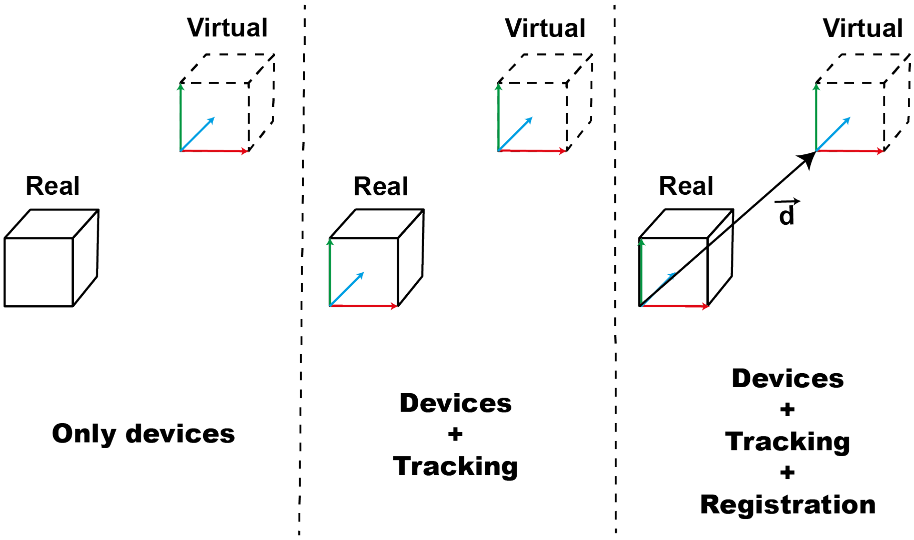


Figure 1.2: The difference between a device without and with tracking and registration in AR. The red, blue, and green axes indicate the object's known position. The d vector indicates the three-dimensional distance between the two objects.

of constraints represented by the user's real context (e.g., background texture, shape of the room, etc.). As explained above, in VR, the developer has full control of the environment presented to the user, as the real environment is completely occluded. In contrast, in AR, the developer must consider the real environment in which the application will be deployed, in the sense that there is a pre-existing universe over which he does not necessarily have control. Moreover, it must be able to respond to changes in or combine the real world with the virtual world, managing, for example, the occlusion of virtual elements by real ones. These constraints represent three of the major general challenges in the field (see Figure 1.2) [Billinghurst, 2021; Kim et al., 2018; Ong et al., 2008; Van Krevelen and Poelman, 2010; Vyas and Bhatt, 2017]. First, **design of devices**, worn on the head or at least in front of the eyes, to enable the two worlds to be displayed. Second, **tracking and sensing** to perceive real-world details and change, thus detecting the surfaces and depths necessary for correct immersion and fusion of the two worlds. Third, what is known as **registration**, which consists of the precise alignment of elements with one another, enabling a virtual element to be positioned in relation to a real one (e.g., superposition). Naturally, the immersive nature of these technologies also leads to a desire for interactivity that is more intuitive and closer to what we do in the real world (e.g., grasping and manipulating an object with our hands) [Billinghurst, 2021; Kim et al., 2018; Ong et al., 2008]. This issue of **Interaction** represents the fourth major challenge of immersion. Since it involves capturing gestures and other modalities, it is intrinsically linked to the Tracking challenge and Multi-Modal Interaction field.

Another major challenge exposed by the immersive MR technologies is the **User Interface** with the switch of paradigm in terms of workspace [Billinghurst, 2021; Kim et al., 2018; Ong et al., 2008]. Where on desktop and mobile, the workspace is in 2D and limited to the size of the screen, in immersive MR, the workspace is in 3D and limited to all the environment (i.e., unlimited in terms of space, but limited by the walls and other elements that compose the environment) [Chandler et al., 2015]. Although that presents some advantages, such as reducing the cognitive load on 3D visualization, it means rethinking the place and usefulness of 2D and 3D interfaces, including new constraints related to the real environment, such as depth and occlusion. This reflection should be carried out by addressing not only the **understanding of user's behaviors**, by verifying the applicability of the traditional "overview first, zoom and filter, then details on demand" mantra of Shneiderman [1996] prevalent in conventional information visualization due to the limitation of screen size [Chandler et al., 2015; Ens et al., 2021], but also the exploration of the placement and design of **situated and embedded visualizations** [Ens et al., 2021]. Indeed, the principle of AR is to place information directly onto the real environment. So, there is a vital need to identify good and bad design practices to prevent cognitive and visual overload, as well as the generation of overwhelming environments. Additionally, if AR can place information in the real world, it must also be able to extract context knowledge to propose context-aware solutions. Thinking even further ahead, since the displays are not the same as those used in more conventional computing, there is also **a need to re-evaluate asset** (e.g., text, image, video) property guidelines (e.g., font size, font type, color). HMD are closer to the eyes and, for some, use a transparent screen where the environment luminosity is combined with (thus impact) the virtual projection on the screen.

Finally, another major challenge is the definition of **ethical and legal scope** of application scenarios regarding privacy, data security, and cognitive overload [Ens et al., 2021]. Breaking the barrier between virtual and real worlds opens the door to more excellent visual and cognitive overload. To preserve the physical, psychological, and environmental health of the user [Agati et al., 2020; Vi et al., 2019], it is essential to keep a watchful eye on this emerging field before it explodes at the consumer level, implying the risk of losing control over it.

1.3 Research Context and Goal

Of the challenges presented above, the research presented in this doctoral thesis undertakes the **User Interface** challenge, with a specific emphasis on establishing guidelines for the design of immersive AR applications. More specifically, within research possibilities offered by this challenge, we will focus more on visualization as a support to the user's task. Eventually, these guidelines will be integrated into tools to facilitate their accessibility and applicability in practice. An example of a tool under construction is presented in Appendix B in the form of a designer input form. The designer is asked to describe the application context, and then corresponding guidelines are offered.

This main objective was pursued by delving into four design levels, aware of the contextual factors (Rauschnabel et al. [2024] highlighted four major factors in MR: consumer, content, context, and computing device) intrinsic to the application's deployment. On the asset level (i.e., text, image, video, 3D models, etc.), our focus entailed examining asset properties contingent on contextual factors and the specificities of immersive devices. On the canonical task level (i.e., tasks indivisible into further subtasks), our focus concentrated on assessing the influence of context on visualization choices. Finally, on the application and cross-applications levels, attention was directed towards, on the one hand, probing cross-cutting concerns applicable to all applications, such as information clutter, and on the other hand, studying user behavior in the utilization of immersive AR either as a complement to conventional applications or for collaborative engagement. It is noteworthy that collaborative aspects were recognized as a major challenge in the domain of immersive MR by Billinghurst [2021]. The inquiry is centered around designing an efficient collaborative environment, encompassing both in-situ and remote collaboration, as discussed by Chandler et al. [2015] and Ens et al. [2021]. The allure of MR lies, in part, in its capacity to enhance collaboration among two or more users across diverse contexts. Preliminary observations in this domain discerned that these technologies contribute to enhancement through heightened feelings of presence, immersion, involvement, engagement, and awareness [Sereno et al., 2022].

To this end, we benefited from two distinct research initiatives that fund the research. The Evocative Framework For Text Analysis - Mediality Models (EFFaTA-MeM) project supported the research during the initial year and a half of the doctoral program. Subsequently, the Federated Learning and Augmented Reality for Advanced Control Centers (FLARACC) founded by RW/Pôle Mecatech project provided financial support for the remainder of the doctoral research period.

The EFFaTA-MeM project aimed to propose novel models and methods to support text analysis. Its transdisciplinary approach adopted a view of the medium of text informed by both mediality studies and human-machine interaction studies, thereby simultaneously offering the best of both worlds. It used theories from human sciences as ways of sifting through textual data to extract several facets of the implicit knowledge integrated into texts. This project presents a perfect use case to study user behavior and expectations regarding a mix of conventional and immersive functionalities.

The FLARACC project aims to explore, in addition to the federated learning part, the application of AR in maintenance use cases. The project involves three companies, a research center, and two universities. This project has led to interesting exchanges at different levels of our research, thanks to the use cases that have emerged from the needs of the three partner companies: technician guidance, remote collaboration for technician or customer assistance, and training in maintenance practices.

1.4 Outline

The remainder of this thesis is divided into four parts.

Part 1 presents the background of the research fields related to this thesis. Chapter 2 begins by giving a scientific definition of AR and associated concepts. Subsequently, it delineates the diverse displays and interactions inherent to the augmented reality experience. Next, Chapter 3 begins by providing key concepts of the visualization field before reviewing the literature on the User Interface challenge in the immersive AR field.

Part 2 presents the research design followed to address the various objectives exposed in this introduction. Chapter 4 refines the objectives of specific research questions. Next, Chapter 5 develops the methodology followed to address the distinct research questions.

Part 3 presents the results obtained for each research question in dedicated chapters. Chapter 6 addresses the question of readability of text. Chapter 7 is interested in visualizations to guide the user to points of interest in a three-dimensional environment. Chapter 8 tackles the combination of immersive AR visualizations with a web application. Chapter 9 addresses the question of remote collaboration in a context of maintenance. Chapter 10 is interested in the information clutter issue and giving back control to the user in such a scenario.

Part 4 reflects on the cross-chapter results presented in the previous part. Chapter 11 summarizes the contributions of this thesis and discusses their implications for research and practice as well as the limitations of this thesis. Chapter 12 elaborates on four major leads for further research. These propose new contributions for which the research presented in this thesis can serve as a basis. Finally, Chapter 13 concludes the thesis.

Part I

Background

CHAPTER 2

MIXED REALITY: DEFINITIONS, DISPLAYS AND INTERACTIONS

This chapter starts by giving, in Section 2.1, a more in-depth definition of Mixed Reality (MR) as well as positioning the concept among related research fields. Next, Section 2.2 details the various possibilities in terms of displays for implementing Augmented Reality (AR) solutions, while Section 2.3 discusses the diverse modalities of interaction used in AR applications. These last two sections aim to give the reader an overview of the range of possibilities when it comes to implementing MR solutions. Finally, Section 2.4 focuses these input and output modalities in relation to the research objectives of the thesis.

2.1 Definitions

According to Schmalstieg and Hollerer [2016], the concept of **ubiquitous computing** can be traced back to Weiser [1991] and his essay on the 21st-century computer. In his work, Weiser envisions a world where computers seamlessly integrate into the physical environment, rendering users unaware of their interaction with computing devices. Ubiquitous computing is characterized by the aspects of spatial, temporal, communicative, and interactive dimensions. Essentially, it envisions a vast interconnected system where devices communicate based on their spatial and temporal contexts, supporting individuals with passive and context-aware interactions. While recent scientific advancements, such as the Internet, Mobile Computing, and the Internet of Things, have played pivotal roles in realizing ubiquitous computing's principles by enabling universal access to digital information at any time and location, AR is the field that aligns most closely with the core principle of ubiquitous computing [Schmalstieg and Hollerer, 2016].

The predominant definition of AR is credited to Azuma [1997] and is decomposed

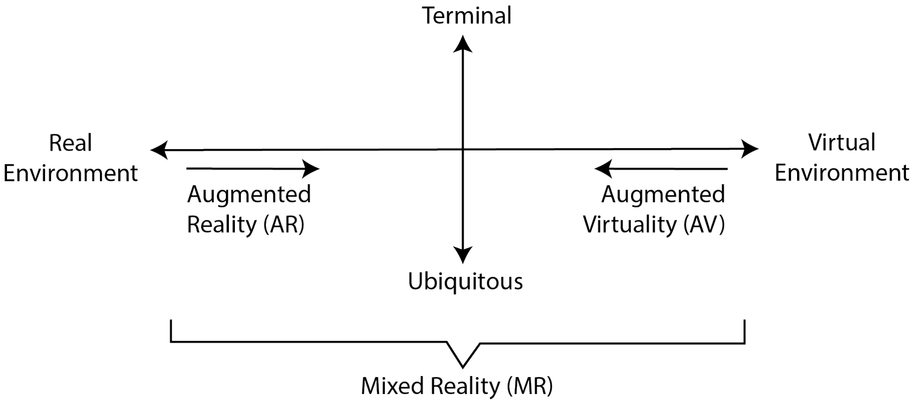


Figure 2.1: Application space for ubiquitous computing according to Newman et al. [2007]. The vertical axis represents the ubiquity that quantifies the extent to which information access transcends fixed physical locations. The horizontal axis represents the virtuality, based on the Reality-Virtuality Continuum of Milgram and Kishino [1994]. The Augmented Reality (AR) starts from the real environment to add virtual elements. The Augmented Virtuality (AV) starts from the virtual environment to add real elements. The Mixed Reality (MR) encompasses the two.

into three fundamental components: 1) it must combine the **real** and the **virtual**, 2) it must be **interactive** in **real-time**, and 3) it must be **registered** in three dimensions (i.e. real-time alignment between both virtual and real worlds). AR technologies reside in the broader field of **Virtual Reality (VR)** technologies, that emphasizes the creation of a world of its own inside the computer by occluding the real. This correlation ensues fundamentally from the intrinsic interplay between AR and its technological underpinnings, which draw upon the foundational technologies of VR while extending their capabilities to encompass the requisite tools for real-world data acquisition. Notwithstanding, these concepts are different, as illustrated by Milgram and Kishino [1994]'s continuum. Milgram's continuum positions the **real environment** at one end and the **virtual environment** at the other end. In the former, there are no virtual elements, while in the latter, there are no real elements. In contrast, **MR** encompasses intermediate worlds with, on one side, the AR, and, on the other side, the **Augmented Virtuality (AV)** (see horizontal axis in Figure 2.1). AR gradually introduces virtual elements into a predominantly real world, whereas AV does the reverse.

Paradoxically, Weiser [1991] considered VR as a divergent concept to ubiquitous computing since it excludes the real world. Nonetheless, Newman et al. [2007] proposed a dual-axis perspective to classify applications in the ubiquitous computing space (see Figure 2.1). The first axis, ubiquity, quantifies the extent to which information access transcends fixed physical locations (terminals). The second axis, virtuality, aligns with Milgram's continuum, emphasizing the varying degrees to which virtual and real environments are mixed.

However, Milgram's continuum is concerned with the visual display, thus focusing only on the sight sense. Yet, as stated by Skarbez et al. [2021], there is a need to extend this vision to the other senses. While there have been efforts to incorporate additional exteroceptive senses, such as taste, into VR devices alongside visual sensory input [Cheok and Karunanayaka, 2018; Schmalstieg and Hollerer, 2016], a fundamental challenge persists. Achieving an ideal VR experience remains unattainable as long as the user's connection to the physical body is maintained, primarily due to the influence of interoceptive senses like the vestibular and proprioceptive systems, which maintain the user in the physical world. In response to this concern, Skarbez et al. [2021] proposed to extend Milgram's continuum beyond VR with a "Matrix-like" Virtual Environment, inspired by the famous movie of the same name. This new step defines a world where all the senses, exteroceptive and interoceptive, are encompassed. In contrast, VR becomes more similar to the concept of AV, with the difference that visual elements are all virtual. This integration of VR into MR also redefines the latter, which in Skarbez et al. [2021]'s proposal no longer focuses solely on sight. This approach solves a second nomenclature problem, as the definition of MR is a real problem right now. While some adhere to Milgram's definition, others contrast it with AR, using either immersion, spatialization, or interaction as a point of comparison. Sometimes the difference between the two is driven by a strong marketing desire to distinguish immersive AR devices from the more conventional approaches on smartphones. According to them, MR no longer augments the world like a smartphone but blends it. Moreover, this vision is certainly accentuated by the concept of **Extended Reality (xR)**, instead of MR, used in practical domains to mention software enabling AR and VR application development. Notwithstanding, in the context of this thesis, we will adhere to Skarbez et al. [2021]'s definition of MR that encompasses AR, AV, and VR concepts.

In addition to the ubiquity axis, Milgram's continuum can also be associated with a modularity axis. **Mediated Reality** [Mann, 1999; Psychoactive agency, 2023] associates Milgram's concepts to **Modified Reality (MfR)** and **Diminished Reality (DR)** (see Figure 2.3). MfR presents an altered world to the user, while DR hides or removes elements from the real world [Cheng et al., 2022]. These concepts are positioned along a modularity axis orthogonal to Milgram's, forming a two-dimensional classification space for applications (see Figure 2.2). At the end of this modularity axis is the severe DR that consists in removing all elements of the real world. In this new space, the real environment is at the origin of the reference frame, while the virtual environment is at the opposite. If we follow the modularity axis, we end up with a world of nothing. However, if we follow Milgram's axis, we end up with a complete fusion of both real and virtual worlds.

Finally, a last concept to introduce is **Pervasive Augmented Reality (PAR)**. Grubert et al. [2017] distinguish PAR from **Conventional Augmented Reality (CAR)**, the predominant focus in AR research. CAR applications are context-specific and accessed as needed, whereas a comprehensive PAR experience is continuous and multipurpose-oriented. This change of paradigm opens the door to new challenges that are, at present, still largely unexplored. A more in-depth study of PAR will be carried out in Chapter 10, where it will be used as a context to study information clutter.

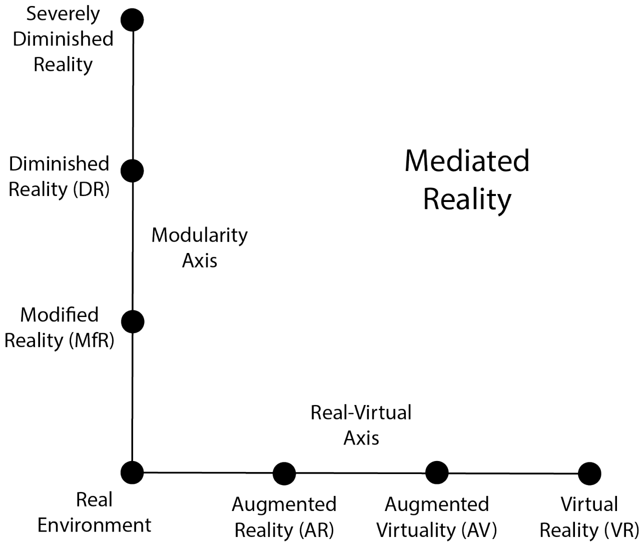


Figure 2.2: Mediated Reality space by Mann [1999]. The modularity axis and the real-virtual axis of Milgram and Kishino [1994] are orthogonal, offering the possibility of going from a worldless experience (top left) to one with two merged worlds (bottom right). The real world remains at bottom left, while the virtual world is at top right.

Figure 2.3: Link to the website of the Psychoactive agency that illustrates the Mediated Reality [Psychoactive agency, 2023]: https://mediatedreality.info/

2.2 Displays

The definition of AR by Azuma [1997] does not require the use of specific hardware devices. Instead, an AR system should comprise sensors for capturing real-world data, a computational device for managing the generation of the virtual world and processing information from the sensors, and a display output for presenting the combination of these two to the user. AR displays can be classified in five distinct categories presented in the rest of this section: **Spatial Augmented Reality (SAR)**, **Stationary Augmented Reality**, **Hand-Held Display (HHD)**, **Head-Mounted Display (HMD)**, and **Head-Up Display (HUD)**.

2.2.1 Stationary Augmented Reality and Spatial Augmented Reality (SAR)

Both systems employ external sensors for real-world data capture and a computer generating the augmented view. In the context of Stationary AR, the rendering occurs on a dedicated screen. Conversely, Spatial AR systems project rendering directly on the physical surfaces, making a notable distinction in terms of immersion with Stationary AR. Furthermore, both approaches are bound by a spatial fixity constraint but offer the benefit of not requiring users to wear any additional equipment.

An example of a SAR application is the sandbox, illustrated in Figure 2.4a, which has found interest, for instance, in both teaching [Kundu et al., 2017] and military strategy [Amburn et al., 2015]. The user can modify in real-time a landscape by modeling mountains and ravines. Then, virtual content projected onto it is adapted to the terrain generated.

2.2.2 Hand-Held Displays (HHD)

A solution that has garnered widespread popularity due to its user-friendly nature involves the use of smartphones or analogous handheld devices. These gadgets, stowed in individuals' pockets, come equipped with sensors adapted for AR applications. As such, they possess the advantageous qualities of being all-in-one and cost-effective. Nevertheless, they present constraints, notably a restricted working interface and having to hold the smartphone at the right height.

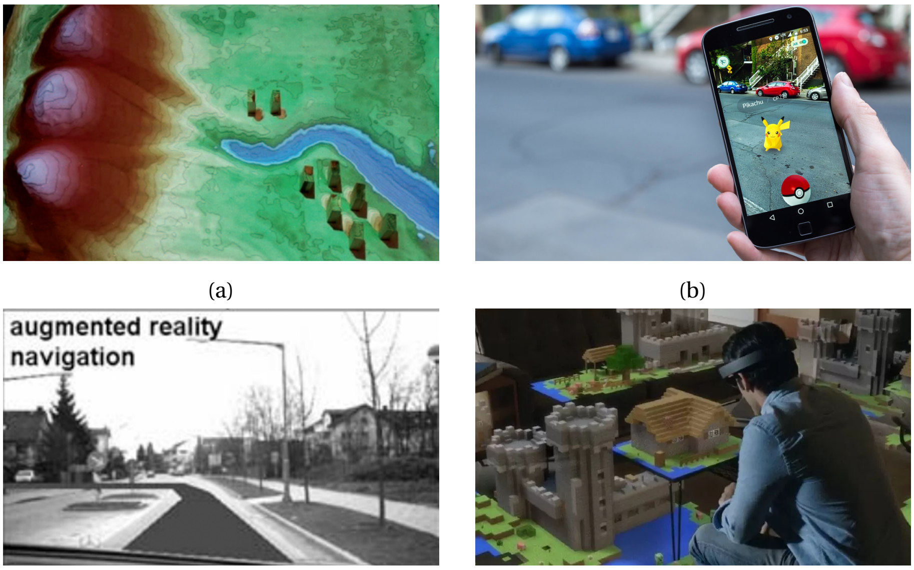
An example would be the "Pokémon Go" application, illustrated in Figure 2.4b, which has been popular in the media. The game consists in catching the creatures all over our world through the smartphone and competing in arenas scattered across cities.

2.2.3 Head-Mounted Displays (HMD)

Also known as smart glasses or Head-Worn Displays (HWDs), HMDs directly enhance human visual perception. These devices, irrespective of being of Video See-Through (VST) or Optical See-Through (OST) design, position a visual display mere centimeters from the user's eyes. A VST display captures the real-world environment through a camera, followed by a computational process that merges this real-world input

---

[1] https://www.alamyimages.fr/photo-image-montreal-ca-le-11-aout-2016-libre-d-un-homme-jouant-de-pokemon-rendez-sur-un-telephone-intelligent-pokemon-go-est-un-jeu-de-realite-virtuelle-sorti-en-juillet-2016-133627370.html



(a) (b) (c) (d)

Figure 2.4: Illustration of Augmented Reality (AR) applications using various devices: (a) the sandbox with a Spatial Augmented Reality (SAR) solution (Source: Kundu et al. [2017]), (b) Pokémon Go using a Hand-Held Display (HHD) (Source: Alamy[1]), (c) the road indication using a Head-Up Display (HUD) (Source: Narzt et al. [2005]), and (d) AR version of Minecraft using a Head-Mounted Display (HMD) (Source: Microsoft).

with virtual content before presenting the result on an opaque screen. These headsets encase the user's visual field within a confined space, thereby immersing them entirely within the presented digital content and restricting peripheral vision, resulting in a narrower field of view than that afforded by unaided human vision [Rolland and Fuchs, 2000]. Conversely, an OST display operates by generating virtual content in the form of projected light upon a transparent screen, where it subsequently reflects and combines with ambient light from the user's surroundings before being perceived by the user's visual sense. These headsets are considered less obtrusive due to their utilization of a transparent screen, which in no way encumbers the user's field of view. Overall, these two approaches possess distinct advantages and drawbacks, as extensively discussed by Rolland and Fuchs [2000] and Schmalstieg and Hollerer [2016].

First, the OST approach slightly degrades the perception of reality due to the transparent screen used, in contrast to the VST approach, which relies on both screen resolution and the camera's real-world capture quality. Thus, VST displays are more susceptible to inducing motion sickness compared to OST displays [Kaufeld et al., 2022; Munafo et al., 2016; Yuan et al., 2018]. Motion sickness occurs when the visual motion perceived by the eyes contradicts the motion sensed by the inner ear.

Second, software latency has dissimilar implications for the two approaches. In OST systems, latency only affects the virtual content, leading to misalignment between the virtual and real worlds. For example, in a surgical scenario where virtual organs aid in incision planning, latency could cause a misalignment between the displayed virtual organs and the actual positions of the real ones. Conversely, in VST systems, latency affects both real and virtual worlds simultaneously, resulting in a disruptive application freeze for the user. For instance, Microsoft [2023d] recommends a minimum frame rate of 60 frames per second (fps) for both approaches, with a preferable frame rate of 90 fps for VST applications.

Third, in OST systems, a distinction must be made between the overlay field of view (representing the area suitable for displaying virtual content) and the peripheral field of view (comprising the remaining visual space). This differentiation is unnecessary in VST setups, as both the real and virtual views are inherently linked.

Last, OST displays are susceptible to substantial variations in contrast between virtual- and real-world elements. This is because the human eye directly perceives ambient luminance, unlike VST devices where luminance is regulated by the camera. Consequently, in OST displays, the discernible spectrum of virtual colors depends on the background illuminance level. In brighter environments, virtual colors must possess higher intensity to be distinguishable, whereas dark colors may appear transparent.

HMD can also be categorized based on their immersive capabilities. On one hand, light smart glasses deploy a compact screen in front of a single eye, exemplified by devices like Google Glass[2], thereby embodying a stationary augmented reality device mounted on the user's head. On the other hand, immersive headsets encompass the broadest possible visual field, effectively erasing the boundary between the physical and virtual worlds. In this latter scenario, virtual elements are superimposed upon specific locations within the real environment, creating the illusion of tangibility.

In evaluating the merits and demerits of light smart glasses, it is important to note that differences in focal planes between the eyes can potentially induce visual fatigue. Nonetheless, this class of devices offers the advantage of freeing the user's hands, a distinction from HHD. In contrast, immersive displays may pose challenges related to factors such as device weight, limited battery capacity, and issues discussed previously in relation to OST and VST devices, including a strong motion sickness, which can vary in intensity among individuals. Furthermore, these devices face the hurdle of social acceptance depending on the place, time and person using it [Schwind et al., 2018; Van Krevelen and Poelman, 2010]. Moreover, in the case of VST displays, these devices obscure the wearer's eyes, a crucial component of human communication [Rolland and Fuchs, 2000; Schmalstieg and Hollerer, 2016].

At first, VST headsets were mainly used for VR video games and training. But recent developments have seen them equipped with frontal cameras to enable AR as well. In terms of applications, we can cite as an example the immersive version of the famous video game Minecraft on the Hololens (the Microsoft's OST HMD), which places this cubic universe in the center of the user's living room, as illustrated in Figure 2.4d.

---

[2] https://blog.google/products/devices-services/glass-enterprise-edition-2/

2.2.4 Head-Up Displays (HUD)

HUD technology refers to the system that integrates AR solutions in vehicles. This implementation involves the installation of sensors within the vehicle, establishing a connection to the on-board computer, which subsequently projects visual information directly onto the windshield. The solution derives is name from the specific positioning of the windshield, aligning it precisely with the user's line of sight. Thus, this approach inherits the advantages and disadvantages of OST display.

As an example, Narzt et al. [2005] used a HUD display to give route indications to a car driver. Their solution colorized the correct road depending on the user's destination (see Figure 2.4c).

2.3 Interactions

An interface must react and evolve according to the user's interactions, and the user's needs. When a button is clicked, a visual change is expected. Similarly, it is because there is a visual change that we know we have clicked on the button. It is also because the button is presented in a certain way, different from the other elements making up the application, that we know it is a button. This section presents the different interaction modalities offered by the various devices mentioned in the previous section, based on the works of Hertel et al. [2021], Grubert [2021], and Schmalstieg and Hollerer [2016].

2.3.1 Devices and Tactile Modality

The most fundamental human-computer interaction tools, specifically the keyboard and mouse, can also serve as interfaces for interacting with AR applications. These interactions are predominantly employed within Stationary AR systems, as they present augmented content on workstations adapted for their utilization. Nevertheless, their applicability extends beyond stationary setups, as exemplified by Maiti et al. [2017]'s research, which leverages a real keyboard instead of a virtual one projected by an HMD. In a broader context, the interaction devices intrinsic to the various displays used in AR can help control the experience. Nonetheless, not all input devices are universally adaptable to all AR setups. A device like a pen, for instance, can be employed on displays equipped with touch-sensitive screens, aligning more with HHD. The same is true for tactile interactions.

In fact, a pivotal consideration revolves around whether the interaction device can establish a connection with the given display. This was the case for the wireless keyboard in the study of Maiti et al. [2017]. It is also the case of joysticks commonly employed in gaming consoles. The latter are also widely used in VR, where the user holds controllers in each hand, tracked by beacons and equipped with buttons that allow both control of virtual hands and access to actions in game configured by each application. Additionally, it is worth noting that even automotive components like the steering wheel and dashboard can be assimilated into this category, functioning as joystick-like interfaces for drivers within the context of augmented reality applications.

A second difference between the various input devices is the Degrees of Freedom (DoF) they offer in terms of control. This level indicates the number of properties captured by the device and sent to the application. A 6-DoF, for example, will allow controlling the translation and rotation of a virtual element. A 3-DoF, on the other hand, allows controlling either translation, rotation or certain axes of each transformation. Indeed, the mapping between each DoF and their impact on the application is left to the designers' discretion. The virtual reality controllers, for instance, have 6-DoF, since they are tracked by beacons and equipped with gyroscopes.

Finally, interaction interfaces that make an exception are the Brain-Computer Interaction (BCI) interfaces. Although still at an early development stage, they are destined to enable users to use their thoughts to control what happens in the application. Si-Mohammed et al. [2020], for instance, studied the use of a BCI to control a robot while receiving visual output and indications through the use of an HMD. Of course, the conceptualization of such interfaces would fit in perfectly with the ubiquitous computing principle underlying AR. More generally, however, this technological evolution would revolutionize the field of Human-Computer Interaction. If the HMD may be considered as the screen of tomorrow, the brain chip can be considered as the interaction of tomorrow.

2.3.2 Gesture, Haptic, and Tangible Modalities

Although gestures are not a novelty, as they can be captured on traditional computers and home consoles through supplementary sensor devices, the utilization of HMDs has notably intensified the adoption of this interaction modality. These HMDs effectively transcend the boundary between the physical and virtual realms, thereby stimulating a transformation in the nature of Human-Computer Interaction. In essence, there is a keen interest in facilitating users' interactions with virtual entities in a manner akin to their interactions with physical objects. For example, pressing a virtual button with the index finger should mimic the tactile experience of pushing a switch. To achieve this objective, HMDs are equipped with various sensors capable of tracking the positional data of various hand joints of the user. This approach has its inherent limitations, such as potential occlusion of certain parts of the hand, yet it does permit the mapping of a virtual hand onto the user's physical hand. Subsequently, this mapping enables the system to differentiate the specific gestures being executed and respond accordingly.

Nevertheless, one area of research that remains at a research stage is the implementation of haptic feedback mechanisms, which are essential for providing the sensory cues necessary for a realistic tactile experience during these interactions. In physical interactions, even the gentlest touch or scratch on an object provides sensory feedback, imparting valuable information about the object's material properties. Regrettably, in the context of virtual elements, users often encounter the issue of passing through them, leading to a disruption in their immersive experience. Presently, the most prevalent solution involves the deployment of haptic feedback gloves, which generate electrical impulses that simulate the tactile sensations one would feel when interacting with real objects. Nevertheless, it is important to note that this technology remains in a prototype stage, with ongoing developments required for commercialization. Another avenue of exploration revolves around the use of brain-computer interfaces, although it is important to emphasize that research in this domain has not yet matured sufficiently for practical deployment as discussed above. However, these challenges do not preclude the application of haptic feedback for informing users of specific events within the virtual environment, such as a watch that vibrates to indicate the receipt of a new notification.

Moreover, the fusion of gestures and haptic feedback with tangible elements presents an intriguing avenue for interaction. According to Grubert [2021], This concept was initially introduced to AR by Kato et al. [2000], who used printed markers to anchor virtual models in physical space. Manipulating these real-world markers also affected the position of the corresponding virtual elements. Another illustrative example of tangible interaction is demonstrated by Henderson and Feiner [2008], who incorporate virtual buttons on physical surfaces to provide haptic feedback when users interact with them.

2.3.3 Gaze and Voice Modalities

The final two modalities encompass gaze and voice. In the context of gaze, an indicator in the form of a colored line and/or a cursor is employed to denote the orientation of the user's head, the point of eye visual focus, or where the user points with their fingers. This system facilitates user input, potentially enabling actions like indicating or even initiating selections. However, as elucidated by Plopski et al. [2022], using gaze as a mean of interaction is problematic. It is normally used as a mean of observation, and it is therefore difficult to distinguish a user intended interaction. This problem is more commonly known as the Midas-touch problem. According to Duchowski [2018], there exist multiple solutions to get around the Midas-touch problem, such as eye gestures, multimodality, boundary crossing, etc. However, the most widely accepted solution is the use of a dwell time.

On its side, the voice modality is a well-established feature within mobile computing, frequently used to activate and give instructions to voice assistants. Within the context of augmented reality, it is more often employed with other sensory modalities to validate user intent through the recognition of specific keywords. It is susceptible to factors such as ambient noise, variations in pronunciation, and fatigue as talking implies a conscious process from the user. Consequently, the voice modality is typically integrated into a multimodal fusion framework, where its inclusion duplicates the functionality of another modality. As an illustrative example, an alternative interaction method to physically clicking on a button involves directing one's attention towards it and vocally expressing the desired command.

2.4 Research Focus

As delineated by Milgram and Kishino [1994], the spectrum of MR is extensive, rendering comprehensive coverage unattainable within the confines of a singular doctoral thesis. Similarly, the preceding sections underscore the myriad input and output modalities inherent to the MR domain. Each display category introduced affords a distinctive approach to the field, yielding scientific insights that may not universally apply across other categories. This observation extends to diverse interaction modalities, which are contingent upon both the display type and the ensuing influence on visualization idioms. In the context of this doctoral research, emphasis has been placed on immersive AR employing HMD, specifically on visualization as a support to the user's task. Therefore, the interaction modalities that will be related to our research are those traditionally associated with this type of device, in other words, gestures, gaze and voice. This approach leads to a total fusion of both virtual and real worlds, presenting novel research challenges, including the domain of Visualization, which will be expounded upon in the subsequent chapter.

CHAPTER 3

VISUALIZATION IN IMMERSIVE AR

This chapter presents, in Section 3.1, a definition of subfields from the visualization field. Next, Section 3.2 provides existing heuristics and guidelines on User Interface (UI). Section 3.3 presents context-centered visualization frameworks to address the visualization in Immersive Augmented Reality (AR). Finally, Section 3.4 situates the concepts discussed in this chapter in relation to the objectives of the research.

3.1 Definitions

The literature identifies three subfields of the visualization field [Munzner, 2015; Rhyne, 2003, 2008]. The first two are defined by Gershon and Eick [1995], at the first IEEE Symposium on Information Visualization, by the following statement: *"**Information Visualization** is a process of transforming data and information that are not inherently spatial into a visual form, allowing the user to observe and understand the information. This is in contrast with **Scientific Visualization**, which frequently focuses on spatial data generated by scientific processes."*. In other words, the difference between the two is on whether there is a physical measures associated to the spatial property inherent to the data. In contrast, the third subfield, called **Visual Analytics**, combines *"the science of analytical reasoning facilitated by interactive visual interfaces."* [Cook and Thomas, 2005]. In other words, *"Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets."* [Keim et al., 2008]. Visual Analytics combined with immersive technologies gives birth to **Immersive Analytics** that use engaging, embodied, and immersive technologies to support analytical reasoning [Chandler et al., 2015; Marriott et al., 2018]. As stated by Marriott et al. [2018], "its goal is to remove barriers between people, their data, and the tools they use for analysis". What is more, AR brings two important new concepts underlying the field with, on one hand, **situated visualization**, and, on the other hand, **embedded visualization**. The first encompasses all visualizations that take their full meaning when situated in their context and environment. The second is situated information that is deeply integrated in their context and environment [Marriott et al., 2018].

3.2 Heuristics and Guidelines

Hartson and Pyla [2012a] and International Standards Organization [2018] define **Usability** as *"the pragmatic component of user experience, including effectiveness, efficiency, productivity, ease-of-use, learnability, retainability, and the pragmatic aspects of user satisfaction"*. This is the criterion used to determine whether an interface is good or not. To this end, Forsell and Johansson [2010] determine 10 heuristics that encompass the highest coverage of common and important usability problems (see Table 3.1). These heuristics highlight the need to consider all the context around the visualization, all the more in Mixed Reality (MR) applications where spatial and long-term visual overload constraints are reinforced due to inherent characteristics. In addition, other authors have proposed similar guidelines for UI and system design in AR. LaViola JR. et al. [2017] have written a book on 3D interfaces in general, but including many references to AR. Vi et al. [2019] propose a series of guidelines specific to Head-Mounted Displays (HMDs). Finally, Funk et al. [2016] and Agati et al. [2020] provide guidelines for industrial assembly tasks, respectively, based on acquired experience and by reviewing works proposing guidelines. The guidelines extracted from these works will be presented according to the classification of Agati et al. [2020] restricted to those related to the design of UI.

Regarding the usability guidelines on visual aspects, the recommendations indicate to keep the elements simple and direct, i.e. presenting only essential information, at the right place and at the right time [Agati et al., 2020; Funk et al., 2016; Vi et al., 2019]. Placing the information directly where the action takes place reduces the cognitive effort required to move from information to task and vice versa. In addition, Funk et al. [2016] raises the problem of tasks that can make reading a text complicated, or the fact that an image or video often shows too much information that is not useful for the current task. Therefore, the idiom of visualization used depends also on the context around the elements and the user. More generally, interfaces should always be analyzed, designed and evaluated with human factors in mind [LaViola JR. et al., 2017]. However, while it is advisable to display instructions and feedback directly at the action location, it is also advisable to take full advantage of the space offered by the third dimension [Vi et al., 2019], maintaining control over the amount of information displayed at any time to limit information clutter. To this end, 3D elements are not always the solution to 2D elements, the application functionalities must be well-structured (e.g., hierarchical menus), and the user must be guided through the application [LaViola JR. et al., 2017]. Furthermore, Funk et al. [2016] and Agati et al. [2020] recommended providing monitoring information (e.g., elapsed time, error, mission completion) to the user to help them assess themselves. Last, it is also suggested to let as much as possible, without becoming cumbersome and restrictive, users personalize their experience, since all users are different [Funk et al., 2016; Vi et al., 2019].

Information coding	Perception of information is directly dependent on the mapping of data elements to visual objects. This should be enhanced by using realistic characteristics/techniques or the use of additional symbols.
Minimal actions	Concerns workload with respect to the number of actions necessary to accomplish a goal or a task.
Flexibility	Flexibility is reflected in the number of possible ways of achieving a given goal. It refers to the means available to customization in order to take into account working strategies, habits and task requirements.
Orientation and help	Functions like support to control levels of details, redo/undo of actions and representing additional information.
Spatial organization	Concerns users' orientation in the information space, the distribution of elements in the layout, precision and legibility, efficiency in space usage and distortion of visual elements.
Consistency	Refers to the way design choices are maintained in similar contexts, and are different when applied to different contexts.
Recognition rather than recall	The user should not have to memorize a lot of information to carry out tasks.
Prompting	Refers to all means that help to know all alternatives when several actions are possible depending on the contexts.
Remove the extraneous	Concerns whether any extra information can be a distraction and take the eye away from seeing the data or making comparisons.
Data set reduction	Concerns provided features for reducing a data set, their efficiency and ease of use.

Table 3.1: General heuristics for common and important usability problems. (Source: Forsell and Johansson [2010])

Concerning the usability guidelines on interactive aspects, the recommendations favor the use of multimodal input [Agati et al., 2020; LaViola JR. et al., 2017; Vi et al., 2019] and output as well as using both natural and non-natural techniques [LaViola JR. et al., 2017] depending on the necessary realism of the application. LaViola JR. et al. [2017] insist on assessing the true potential of the human body to design alternative interfaces. All in all, what is important is to keep intuitive interactions [Funk et al., 2016; Vi et al., 2019] that limit the unnecessary movements [LaViola JR. et al., 2017]. Last, similarly to the visual aspect, the user must be able to personalize their experience [Funk et al., 2016; Vi et al., 2019].

For cognitive guidelines, the first recommendation is, of course, to prioritize user's comfort by ensuring the physiological and environmental user's aspects [Agati et al., 2020; Vi et al., 2019]. The second recommendation is allowing users to feel like they are controlling the experience [Agati et al., 2020; Funk et al., 2016; Vi et al., 2019]. In other words, the user must feel in charge of what is happening in the application. For instance, the user must be able to move forward or backward at will through the steps that define the task. A third recommendation is allowing trial and error as much as possible to let the user explore the application [Vi et al., 2019]. The last recommendation of this category is relying on what already exists, including 2D visualization and interaction works, to help users familiarize themselves with the application in the early stages [Agati et al., 2020; LaViola JR. et al., 2017].

Regarding the ergonomics guidelines, the application design must prioritize the physical user's comfort [Agati et al., 2020; Vi et al., 2019] by, for instance, equipping as much as possible the environment rather than the user. Not only does this remove weight and inconvenience, it also makes it easier to share the system between users [Agati et al., 2020; Funk et al., 2016]. Additionally, Funk et al. [2016] and Agati et al. [2020] suggest a design for hands-free usage to not restrict the user interactions within the real world. To conclude, the last recommendation is to design according to the hardware capabilities and limitations [LaViola JR. et al., 2017; Vi et al., 2019].

In summary, an immersive AR experiences must be simple and consistent. UI must be integrated with the real world by presenting only essential information at the right place (i.e., where it is the most valuable), whether in 2D or 3D, and by offering intuitive control through multimodal interactions, personalization and guiding elements to prevent the user from getting lost or confused. All in all, **a key element is to place context and human factors at the heart of the design phase**.

3.3 Context and Human Factors Approaches

Munzner [2015] introduces a task-centered framework for visualization, articulated around three fundamental questions. The **What** pertains to the characterization of data accessible within the application and their inherent properties, encompassing factors such as data type and static/dynamic behavior. The **Why** concerns the rationale behind presenting data to the user by probing the intended purpose. The **How** defines the final visualization used to represent data and associate interactions, leveraging insights derived from both the What and Why levels. Additionally, this framework has found extensions, particularly in the context of immersive environments [Aguilar et al., 2018]. The **When** queries the temporal component, determining the appropriate moment at which data should be made visible. This temporal aspect is pivotal in mitigating information clutter and alleviating cognitive burden, achieved by aligning data presentation with the contextually relevant temporal junctures and spatial positioning prescribed by the **Where** level. It is noteworthy that, while this approach is inherently task-centric, it does not preclude the exploration of broader inquiries, such as the intended audience for the data, encapsulated within the **Who** level. This framework allows the selection of an optimal visualization strategy tailored to the specific task that the user intends to undertake with the data.

In addition, in the field of computer science, a frequently employed strategy is known as the "divide and conquer" approach [Gonzalez-Calleros et al., 2009]. In a way, research for interface design, in terms of visualization, can be seen on four different levels. The first level concerns **assets** (i.e., text, images, videos, 3D models, etc.), the fundamental elements of all visualizations. At this level, the focus is on studying their personal characteristics in relation to the environment in which they will be used, and the devices used to display them. A simple example concerns text legibility, which will vary from one setup to another depending on factors such as the resolution of the screen. The second level concerns the **canonical tasks** that are the smallest possible division of any complex task. Focusing on canonical tasks enables the use of existing literature with diverse application cases that may not necessarily align globally due to the context variations [Palmarini et al., 2017], but could be pertinent at the canonical task level. Moreover, it facilitates a systematic division of the various facets of the application, permitting the early exclusion of solutions that are incongruent with the current context [Radkowski et al., 2015], as well as a standardization of UI [Gattullo et al., 2019; Scurati et al., 2018; Tainaka et al., 2023]. The third level concerns the **applications** themselves, in regard to how combining all the elements to build a homogeneous and intuitive application that respects the heuristics discussed in the previous section. Finally, the last level addresses the **cross-applications concerns**, such as information clutter, security, etc.

Concerning canonical tasks, Carter [1986] studied the frequency of verbs used to perform actions such as creating a file or assigning a password. Gonzalez-Calleros et al. [2009] proposed a list of 87 canonical tasks for UI in general. From their side, M. Stanney et al. [1995] proposed a list of 18 canonical tasks for Virtual Reality (VR), while Tainaka et al. [2023] proposed 31 of them for AR. The exhaustive enumeration of these canonical tasks is accessible in Table 3.2. However, an important difference lies in the source from which these classifications were derived. Whereas those of Gonzalez-Calleros et al. [2009] and M. Stanney et al. [1995] are more focused on UI, that of Tainaka et al. [2023] is more concerned with actions in general within AR applications, including canonical tasks for maintenance and assembly tasks.

The underlying principle of this approach entails a two-step procedure. Initially, the objective is to identify a set of effective visualizations and interactions for each of the canonical tasks. Subsequently, from this set, the optimal visualization and interaction methods are chosen with due consideration to the specific requirements of the given task. Tainaka et al. [2023] proposed 42 distinct visualization types based on factors such as primary information, visual elements, and coordinate systems. They consistently considered non-textual elements to be displayed simultaneously with text. The correlation between these canonical tasks and visualization types was established through a set of 17 specific questions, which are detailed in Table 3.3. According to them, the first three questions are considered essential, while

Category	Canonical Tasks
Gonzalez-Calleros et al. [2009]	
Convey	Communicate, Transmit, Call, Acknowledge, Respond, Answer, Suggest, Direct, Instruct, Request
Perceive	Acquire, Detect, Search, Scan, Extract, Identify, Discriminate, Recognize, Locate, Examine, Monitor, Scan, Detect
Create	Input, Encode, Enter, Associate, Name, Introduce, Insert, Assemble, Aggregate, Add
Modify	Change, Alter, Transform, Tuning, Rename, Segregate, Resize, Collapse, Expand
Delete	Eliminate, Remove, Cut, Ungroup, Disassociate
Trigger	Initiate/Start, Play, Search, Active, Execute, Function, Record, Purchase
Stop	End, Finish, Exit, Suspend, Complete, Terminate, Cancel
Move	Relocate, Hide, Show, Position, Orient, Path, Travel
Mediate	Analyze, Synthesize, Compare, Evaluate, Decide
Filter	Segregate, Set aside
Reinitialize	Wipe out, Clear, Erase
Toggle	Activate, Deactivate, Switch
Select	Pick, Choose
Duplicate	Copy
Navigation	Go to
M. Stanney et al. [1995]	
Perception	Depth, Surface/Segregation, Displacement, Dynamics, Path Tracing, Visual Search/Detection, Identification, Comparison, Spatial Judgment
Cognition	Problem Solving, Categorization, Classification, Divided Attention, Focused Attention, Integrated Attention
Interaction	Self-Movement, Navigation, grab/Select
Tainaka et al. [2023]	
Check	Compare with a sample, Read data, Check status, Pick up/Lift, Remove/Detach, Install/Joint, Insert/Thread, Arrange/Align, Move/-Transfer, Rotate/Turn, Pour/Fill, Drain/Empty, Fold/Bend, Push/-Press, Pull/Tighten
Operate	Open, Close, Cut/Saw, Drill/Punch, Tap/Clap, Hold/Clutch, Wipe/Polish, Tie/Knot, Wrap/Pack, Apply/Paint, Sharpen/Scrape, Spray/Blow, Draw/Write
Locate	Identify/Find
Others	Keep, Wait

Table 3.2: Comparison of canonical tasks classifications.

the remaining ones are considered sub-necessary. The former were employed to exclude visualization methods that might introduce errors or reduce work efficiency, whereas the latter were found to facilitate task execution.

3.4 Research Focus

In this doctoral research, we aimed to enhance and specialize the general guidelines and heuristics outlined in Section 3.2, transforming them into tangible rules applicable throughout the design process. In this respect, we have focused our research across the entire design spectrum, divided into four levels detailed in Section 3.3. In addition, special emphasis was placed on addressing the challenges arose by the concepts of situated and embedded information in the context of Immersive AR visualizations. Consequently, our approach heavily relied on the extended framework proposed by Munzner [2015], incorporating the essential What, Why, Where, When, Who, and How questions. These considerations formed the core of our methodology in developing experimental prototypes and guidelines.

	Question	Visualization Type
1.	Is the work easily understood without specific assistance?	If yes, use text only or with an image. Otherwise, do not use text only.
2.	Is the object flexible?	If yes, use text with an image or a video.
3.	Does it need accuracy within 1 cm?	If yes, do not use object coordinate system.
4.	Is it necessary to identify as a point?	If yes, use a pointing asset in object coordinate system.
5.	Is it necessary to identify as an area?	If yes, use an area pointing asset in object coordinate system.
6.	Is it necessary to show hand motions?	If yes, use a video or a video of a first-person perspective with a transparent background.
7.	Is direction necessary?	If yes, use a direction asset in world or object coordinate system.
8.	Is an auxiliary line necessary?	If yes, use a line asset in world or object coordinate system.
9.	Is it necessary to look around the object from different perspectives?	If yes, use a static or animated product model in world coordinate system.
10.	Does the CG fit with the effective field of view of the HMD?	If no, do not use a static or animated product model in the object coordinate system.
11.	Is the operation complex?	If yes, use a video, a video in first-person perspective with a transparent background, or an animated product model.
12.	Are both of operation and the goal of the object necessary?	
13.	Is it hidden behind something?	If yes, use a direction asset in world coordinate system, or a pointing asset or a static product model in object coordinate system.
14.	Is it dark?	
15.	Are there objects similar to the target object around?	
16.	Is it important to focus attention on the object at hand?	If yes, use object coordinate system.
17.	Does the object get lost while moving?	If yes, use an area pointing asset, a line asset, or a direction asset in a head or object coordinate system.

Table 3.3: Questions by Tainaka et al. [2023] for the correlation between canonical tasks and visualization types. According to them, the first three questions are necessary, while the others are sub-necessary. For each question, the right column presents the recommended visualization type depending on the question answer. The text is considered as always associated with the other assets.

Part II

Research Design

CHAPTER 4

RESEARCH QUESTIONS

The research goal of this thesis is to study how to design, according to a context-centered process, the visual elements of a User Interface (UI) to maximize usability in an immersive Augmented Reality (AR) experience, in particular with a Head-Mounted Display (HMD). Due to the wide scope of this goal, the focus of this thesis was on five research questions described in this chapter. Figure 4.1 situates each research question to their dedicated chapter and design level.

The first research question focused on text at the asset level of the design spectrum. Text presents numerous parameters (e.g., font, size, color, position), making it challenging for designers. In addition, text is one of the most widely used assets in AR applications. It can be used to label objects or places, to give instructions or details about real and virtual elements. Text is a central element of human language and a common element of UI. In a study devoted to maintenance, assembly, and training, Gattullo et al. [2022] determined that text represents 26% of the assets used in the 122 papers they reviewed. It will be interesting to see if the design guidelines that apply to conventional devices are still relevant for HMDs. Additionally, it is important to consider how to modify text design based on whether the HMD is an Optical See-Through (OST) or a Video See-Through (VST), as they have different capabilities, such as managing ambient brightness. Therefore, the first research question is formulated as follows:

RQ1: What text parameters, including the associated contextual constraints affecting these parameters, can designers tune to improve text readability?

The second research question focused on the navigation into an Immersive AR environments, considered as a subtask in the classifications detailed in Chapter 3. The navigation consists of guiding the user to a point of interest, involving either movement or rotation of the head, or techniques to clear the view, or a combination of these. This subtask was chosen because it appears in the vast majority of applications. The problem of guiding the user through the interface was already

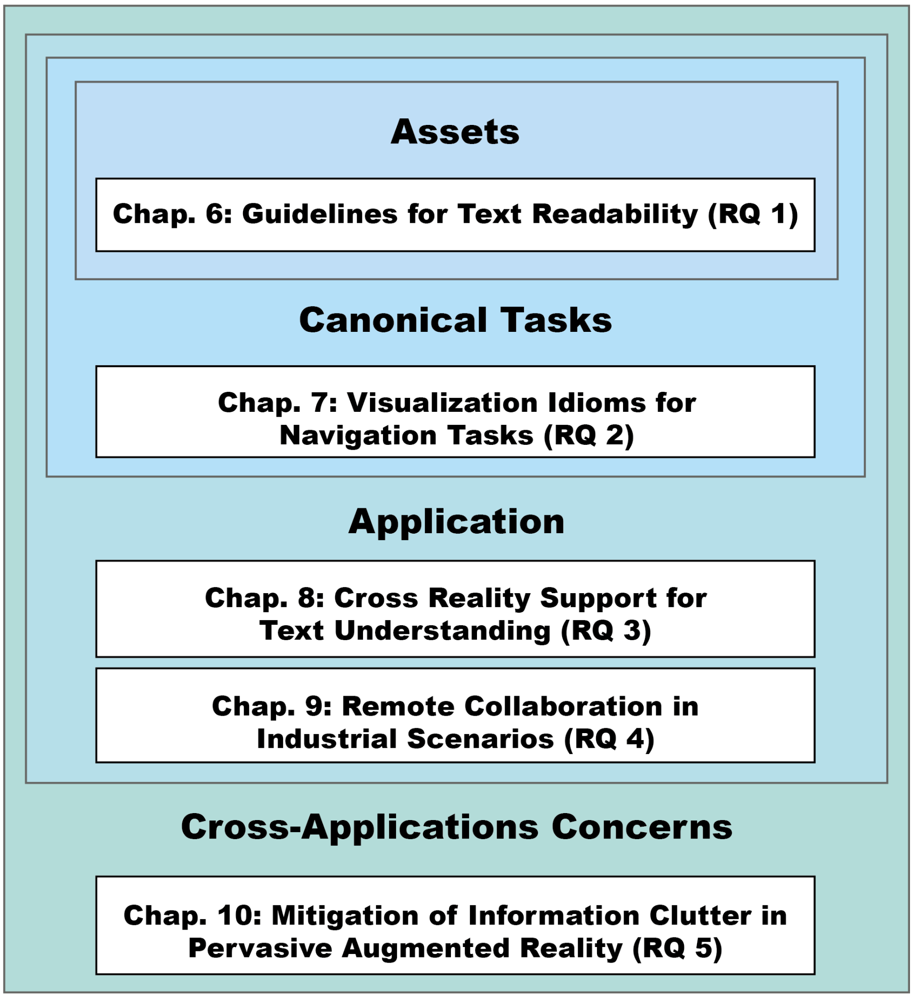


Figure 4.1: Plan of the doctoral focus with the four levels of design, the positioning of each chapter of contributions, and their association with the research questions.

present in 2D interfaces, but it has been reinforced with the transition to the 3D environment, which means that the elements of interest are no longer necessarily in the user's view. Now, with situated and embedded information, the user must move within the environment. The objective is to define a set of viable solutions from which to choose once at an upper level of design (i.e. application level). For that, each solution must be described, on one hand, by the sequence in time and space of the various assets and interactions necessary to complete the subtask, and, on another hand, the constraints induced by the selected assets and interactions. Therefore, the second research question is formulated as follows:

RQ2: What are the best visualization idioms that can be used to guide the user to different points of interest?

The third research question focused on the support for existing conventional applications. Combining conventional and immersive applications leads us to cross-reality systems. In other words, systems designed on different levels of Milgram and Kishino [1994]'s continuum, introduced in Chapter 2. At present, the technical and financial characteristics of HMD do not yet allow long-term use. However, this kind of solution may be of interest in the support it can provide for existing applications on computers or smartphones. This is the case, for example, with the Evocative Framework For Text Analysis - Mediality Models (EFFaTA-MeM) project, which offers a web application for text analysis for human science experts. Based on this use case, we looked at how 3D visualizations can effectively complement the understanding of text already offered by the web application. Based on these lessons learned in a specific use case, we tried to draw conclusions that could be generalized to other contexts of use. Therefore, the third research question is formulated as follows:

RQ3: How can Immersive Augmented Reality support existing conventional applications for text understanding?

The fourth research question focused on the remote collaboration at the application level. In other words, how to propose to the local and remote user the best experience to achieve the local task. This concern is considered as an important challenge in Mixed Reality (MR), as discussed in Chapter 1. In pursuing this use case, our objective is dual-fold. Firstly, we aim to contribute to the resolution of the identified problem, and concurrently, we seek to investigate the design principles inherent in a comprehensive application. Therefore, the fourth research question is formulated as follows:

RQ4: Which approaches best enable remote collaboration in industrial scenarios?

The last research question focused on information clutter challenges, a pervasive issue that arises in numerous applications whenever the volume or complexity of information exceeds manageable thresholds. Therefore, the fifth research question is formulated as follows:

RQ5: How can the issue of information clutter be mitigated through active control from the user in Pervasive Augmented Reality scenarios?

The reciprocal influence among various design levels, spanning from asset considerations to cross-application concerns, is paralleled by interconnections within the associated research questions. The study on text in RQ1 has served as a foundational element for the EFFaTA-MeM project, selected as a use case in RQ3. In addition, RQ2 exhibits universality across applications, given that not all pertinent information aligns within the user's visual field. In the context of RQ3, this phenomenon manifests in the web application, where a mini-map of the graph is presented when zoomed in. RQ4 mirrors this circumstance when tasks split into disparate segments of the physical environment or are obscured by other elements. Furthermore, RQ5 tackles issues potentially shared between RQ3 and RQ4, contingent upon the volume of information necessitating display.

CHAPTER 5

RESEARCH METHODOLOGY

This chapter deals with the general methodology adopted for the doctoral research. Section 5.1 presents the methodologies adopted to review the literature and elaborate prototypes. Then, Section 5.2 explains how these methodologies were applied to each research question.

5.1 Methodologies

5.1.1 Survey Literature

To survey literature, we followed the **Systematic Literature Review (SLR)** methodology recommended by Kitchenham et al. [2015]. The main steps consist of 1) establishing the need for a review, 2) defining the research questions, 3) developing the protocol to identify relevant works, 4) conducting the review, 5) assessing the coverage of the study, 6) analyzing the results and 7) reporting the results. However, sometimes, it could be interesting to include, in addition, the Grey Literature [Garousi et al., 2019] in a **Multivocal Literature Review (MLR)** study. Overall, the protocol remains the same, except that in addition to targeting academic libraries (e.g., IEEEXplore, ACM DL digital libraries), those containing Grey Literature (e.g., Google) are also considered.

Garousi et al. [2019] noted that the most widely accepted definition of Grey Literature is the Luxembourg definition of Schopfel and Farace [2010]: *"<Grey Literature> is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body"*. According to them, integrating Grey Literature in a survey permits to *"close the gap between the academic research and professional practice"*. In addition, Lawrence et al. [2014] argued that a part of the Grey Literature equally follows a peer-review process, ensuring its quality.

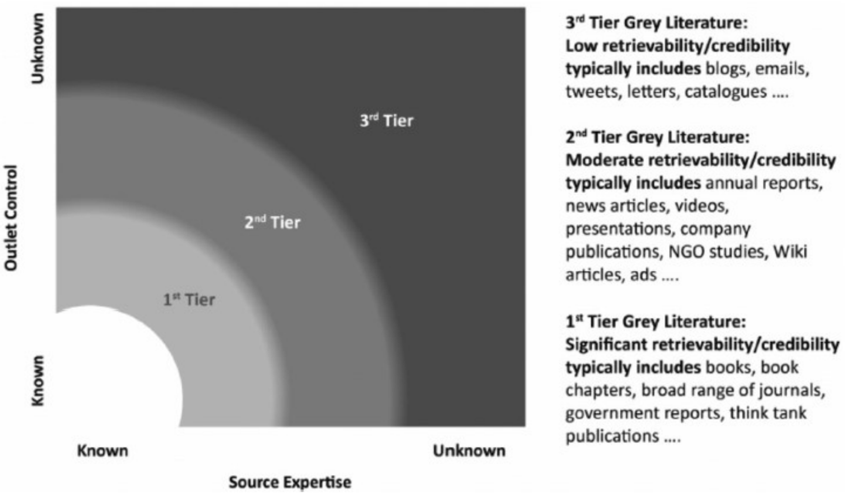


Figure 5.1: Shades of Grey Literatures. (Source: Adams et al. [2017])

To this end, Adams et al. [2017] divided the Grey Literature in three tiers based on the source expertise and outlet control dimensions (see Figure 5.1). The first tier consists of books, magazines, or even reports with a significant control and credibility. The second tier consists of question/answer forums, wikis, videos, etc. with a moderate control and credibility. Finally, the last tier consists of blogs, tweets, letters, etc. with low or no control and credibility.

5.1.2 Design Science Research Framework

The **design science methodology**, well-established in information system [Hevner et al., 2004; Perea et al., 2017], offers an efficient approach to prototyping. The design science methodology is structured around three core components (see Figure 5.2): 1) the **environment** encompasses the identification of pertinent business needs originating from individuals, organizations, and technologies, 2) the **knowledge base** is built upon rigorous and applicable knowledge drawn from prior foundations and methodologies, and 3) the **Information System (IS) Research** involves an iterative process that requires both the development or enhancement of an artifact and its subsequent evaluation. These three components interact in a reciprocal manner. The environment and knowledge base serve as inputs to the IS research phase, providing essential insights into the artifact's functionalities and the methodology for its evaluation. Simultaneously, the IS research phase enriches the environment by offering feedback on the application of the artifact in real-world settings. Additionally, it contributes to the creation of new knowledge in the process.

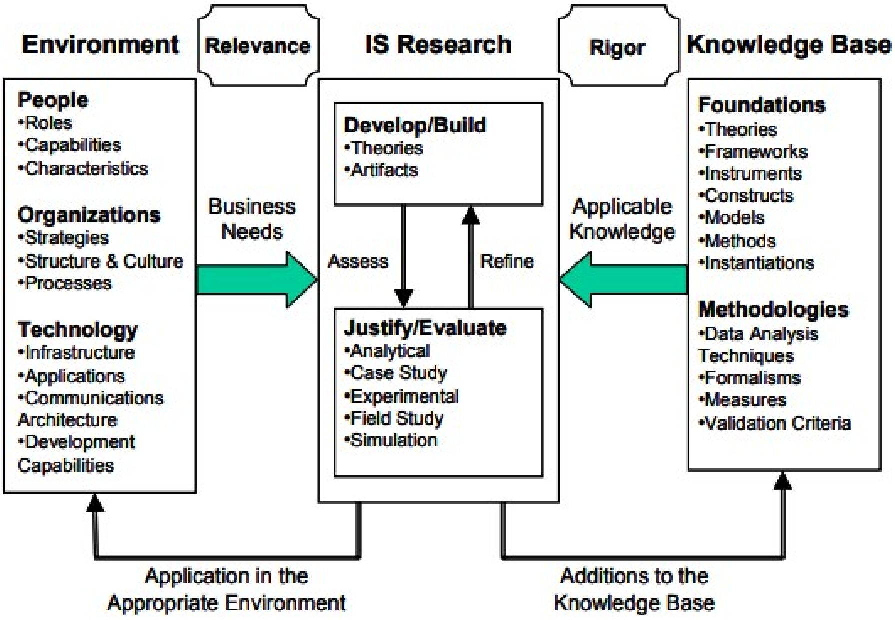


Figure 5.2: Design Science Research Framework. (Source: Hevner et al. [2004]).

	SLR	Prototypes	Preliminary Evaluation	Full Evaluation
Chapter 6	1			
Chapter 7	1			
Chapter 8		3 + Evoq tool	3	
Chapter 9		1	1	
Chapter 10		2	1	2

Table 5.1: Methodologies, prototypes, and evaluations by chapter.

5.2 Methodology by Research Questions

This section presents the methodology adopted for each research question. Table 5.1 summarizes the methodologies and artifacts of each chapter.

5.2.1 Methodology for Research Question 1

What text parameters, including the associated contextual constraints affecting these parameters, can designers tune to improve text readability?

The objective behind RQ1 is to define guidelines on the presentation of text assets that make up Augmented Reality (AR) interfaces. To this end, although numerous studies have investigated various parameters in controlled experiments, we could not find any work that would give us a comprehensive overview. Therefore, we conducted a MLR to gather existing works and define future works. As stated by Kitchenham et al. [2015], the roles of a systematic review are, among others, comparing practices, evaluating the adoption of techniques in practice domains, and defining the benefits of using tools in a specific context, which are our objectives with this research question. In addition, we included the Grey Literature because we may determine the difference in guidelines between the scientific and practical domains. We consider that device manufacturers and game engines must have conducted studies or based their documentation on the literature.

5.2.2 Methodology for Research Question 2

What are the best visualization idioms that can be used to guide the user to different points of interest?

The objective behind RQ2 is the exploration of navigation techniques to guide users in the environment. The starting point was to define the current state-of-the-art of this domain. We found works summarizing the visualization of occluded on-screen elements. However, despite the numerous techniques proposed in the literature, we did not find any survey for the off-screen points of interest that gives clear guidelines on which techniques to use depending on the context. We therefore decided to carry out this study in the form of a SLR, listing existing techniques and comparing them according to the tasks on which they were evaluated. According to Kitchenham et al. [2015], one of the roles of a systematic review in software engineering is to determine whether specific techniques or practices are more effective than others, and if so, under what circumstances this will be true, which confirms that a SLR is the right choice to achieve our objectives.

5.2.3 Methodology for Research Question 3

How can Immersive Augmented Reality support existing conventional applications for text understanding?

The objective behind RQ3 is the support that immersive AR can bring to a web application. The immersive solution was presented in the form of 3 artifacts, each based on a different immersive metaphor supposed to help the analyst. We chose to develop three different metaphors, each with different functionalities, in order to better study how experts reacted when using them. All artifacts' conception followed the Design Science methodology explained above, and the same internal process of validation. The project team of the Evocative Framework For Text Analysis - Mediality Models (EFFaTA-MeM) project, which serves as use case, was divided into two groups, on one hand, the researchers and developers that conducted and implemented the research, and on the other hand, the project supervisors composed of interdisciplinary university professors providing initial expertise in both user experience and the analysis process underlying the project. When the former finished exploring the literature, the latter validated or invalidated the various leads proposed. Similarly, when the former finished implementing a version of the artifact, the latter validated or invalidated the new features. These choices were always based on a team discussion, but allowed us to take a fresh look at the proposed solution, especially as some of the professors came from the target audience. In addition,

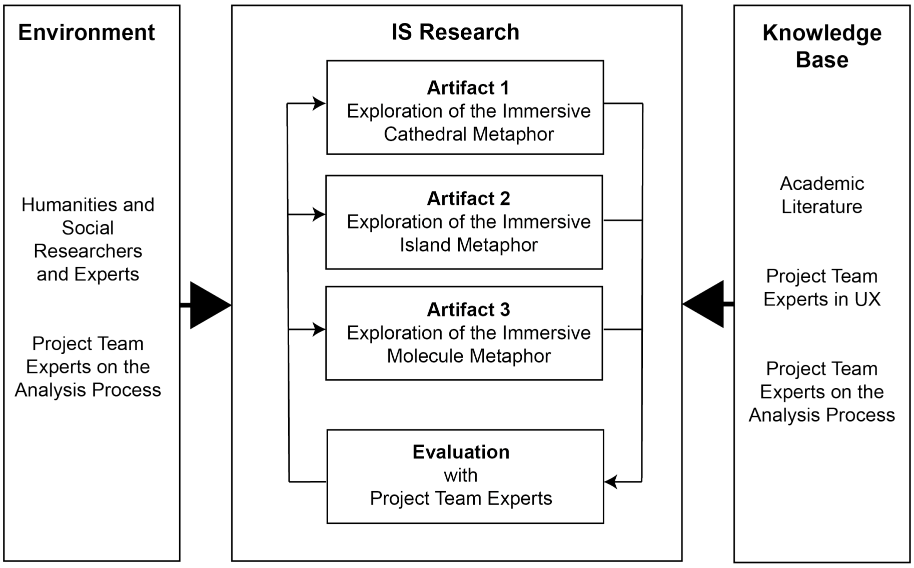


Figure 5.3: Design Science Research framework applied to RQ3.

certain complications made it difficult to keep the project going. This means that only the web application and the Shock Wave graph algorithm have been tested by users outside the team. The proposed immersive visualizations have only been tested internally, and therefore present only preliminary results.

5.2.4 Methodology for Research Question 4

Which approaches best enable remote collaboration in industrial scenarios?

The objective behind RQ4 is the exploration of efficient visual cues for the remote collaboration. An artifact was then designed according to the Design Science Research Framework. Business needs were determined thanks to literature, and interviews with an expert from the Federated Learning and Augmented Reality for Advanced Control Centers (FLARACC) project. The knowledge base was constituted from a review of the academic literature. During the interviews, a preliminary evaluation was conducted with an existing application on the market. Figure 5.4 illustrates the application of the framework to the research question.

5.2.5 Methodology for Research Question 5

How can the issue of information clutter be mitigated through active control from the user in Pervasive Augmented Reality scenarios?

The objective behind RQ5 is to mitigate the effect of the information clutter issue. For that, we adhered to the Design Science methodology with the development of one artifact. The study conducted an exploration of strategies to facilitate the user to keep control of the experience. The formulation of business requirements was

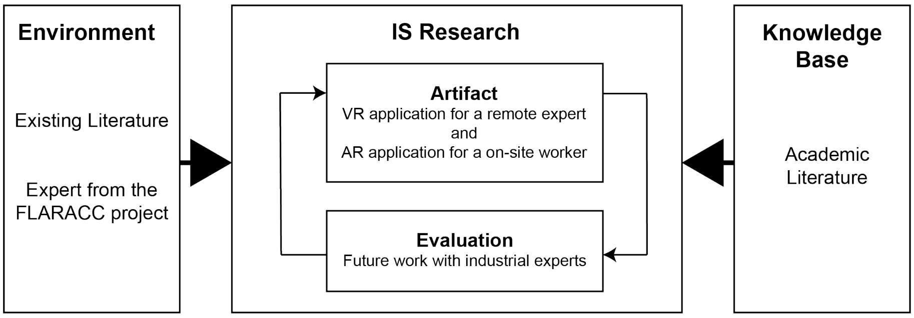


Figure 5.4: Design Science Research framework applied to RQ4.

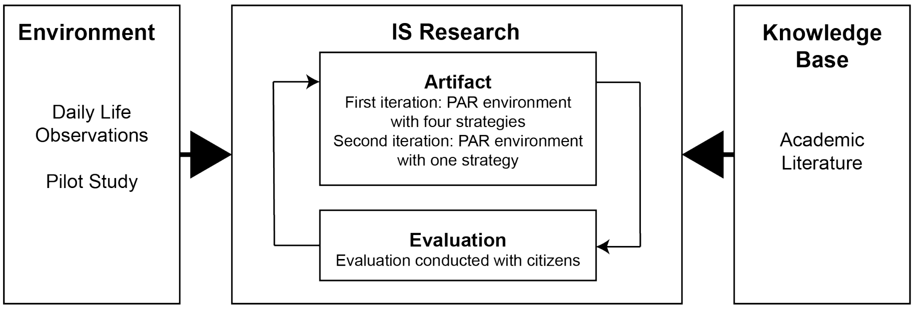


Figure 5.5: Design Science Research framework applied to RQ5.

influenced by empirical observations of daily life behaviors and a pilot study (i.e. interviews) designed to ascertain the genuine challenges inherent to a Pervasive Augmented Reality (PAR) environment. In contrast, the foundation of our knowledge base was meticulously constructed upon a robust framework of prior peer-reviewed scientific literature, specifically in the domain of, or relevant to, PAR, which was used as a use case. A PAR environment was selected because it naturally offers a wealth of information suitable for this kind of study, and its ability to tackle a subject that is still under-explored, despite the societal impact it can have. Figure 5.5 illustrates the application of the methodology to the research question.

Part III

Results

CHAPTER

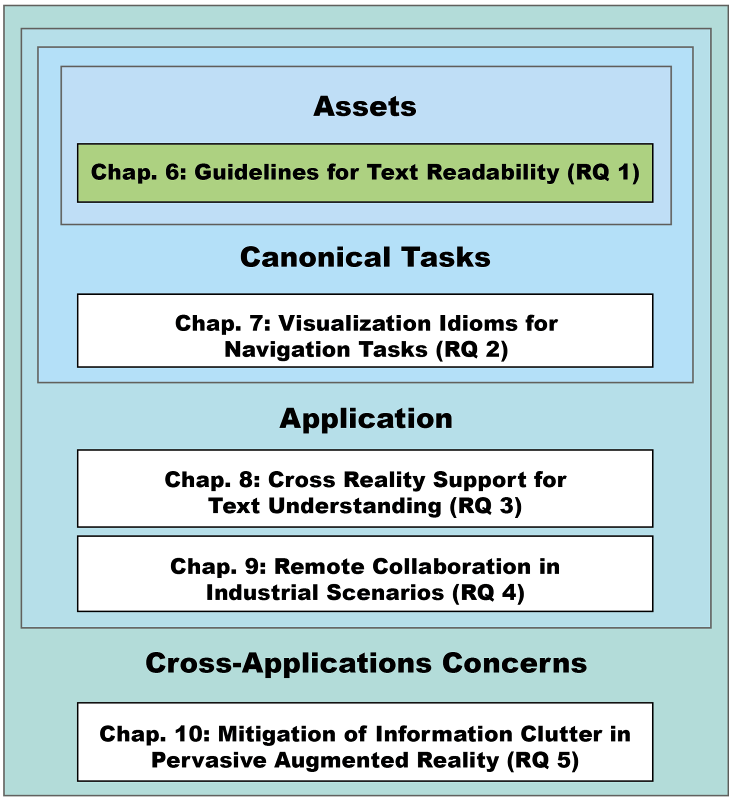
6

GUIDELINES FOR TEXT READABILITY

6.1 General Introduction

This chapter is dedicated to research question 1: **What text parameters, including the associated contextual constraints affecting these parameters, can designers tune to improve text readability?**. As explained in Chapter 5, we did not find any work that would give us a comprehensive overview of the existing literatures. As far as we found, there exists only one recent survey on the topic. Erickson et al. [2020a] reviewed the literature to explore user's perception of text on Optical See-Through (OST) Head-Mounted Displays (HMDs). Their results summarized the literature they found on text drawing style, text color, background texture, environment lighting, text position, and text size. However, they restrained their search to the IEEEXplore and ACM DL digital libraries and to OST displays. In addition, Korkut and Surer [2023] recently reported a need to extend text-related guidelines to Virtual Reality (VR), thus to Video See-Through (VST) displays.

We conducted a Multivocal Literature Review (MLR) that extends the work of Erickson et al. [2020a] by (1) opening the search beyond the IEEEXplore and ACM DL digital libraries, (2) including works on VST HMDs to facilitate the comparison between the two types of displays and expand the applicability of our results, and (3) including official documentations from manufacturers or game engines as grey literature [Cauz et al., 2024]. Ultimately, the aim of our research is to identify the text parameters designers can tune to improve text readability, as well as the associated contextual constraints, in order to propose an up-to-date state of knowledge and a designing guide grounded in both academic and grey literature. The guide is of interest both to researchers wishing to explore the readability of text, and to designers wishing to set up functional applications adapted to their users, their tasks, and their environment. In no case, however, are we going to derive new rules, which positions this chapter as a review and not a theory synthesis [Jaakkola, 2020].



Therefore, we formulated the following research questions concerning Augmented Reality (AR) application on HMDs:

RQ1.1 What are the different text parameters that can be tuned and how to tune them to maximize the readability of a text?

RQ1.2 What are the contextual constraints to consider when designing an application including texts?

RQ1.3 What is the gap between the academic and grey literature in terms of the guidelines proposed on text readability?

Publications and supervised works

The content of this chapter is based on one peer-reviewed publication in a scientific journal:

Cauz, M., Clarinval, A., and Dumas, B. (2024). Text Readability in Augmented Reality: A Multivocal Literature Review [Article under Review Process after Minor Revisions]. **Virtual Reality**

- This paper presents the results of the MLR that compose this chapter.

Outline

The organization of this chapter adheres to the subsequent structure. Section 6.2 presents the methodology guiding the conduct of the MLR, following the steps of Garousi et al. [2019]. The need for a review and the research questions were already established in this introduction and Chapters 4 and 5. The next main steps consist of developing the protocol (Section 6.2.1, 6.2.2, 6.2.3 and 6.2.4), conducting the review (Section 6.2.6), and assessing the quality of the coverage (Section 6.2.5). In Section 6.3, we report the results of the selected papers for each text parameter we identified. In Section 6.4, building on these results, we develop our designing guide in the form of decision trees built from those parameters. Finally, in Section 6.5, we discuss the implications of our results, limitations, and future works before concluding in Section 6.6.

6.2 Methodology

We conducted the MLR by following the guidelines from Garousi et al. [2019]. They adapted the work of Kitchenham et al. [2015], which proposed guidelines to conduct Systematic Literature Review (SLR) in software engineering, to present their MLR protocol that includes Grey Literature. Overall, the protocol remains the same, except that in addition to targeting academic libraries (e.g., IEEEXplore, ACM DL digital libraries), those containing Grey Literature (e.g., Google) are also considered.

6.2.1 Grey Literature Sources

According to Garousi et al. [2019], integrating Grey Literature in a survey allows to "close the gap between the academic research and professional practice", which is needed to answer RQ3, and benefits from the knowledge of both academia and practice, which we need to answer RQ1 and RQ2. In order to ensure a certain quality of documents, the search for Grey Literature targets device manufacturers or game engine websites. They correspond to the first tier of Grey Literature defined by Adams et al. [2017] (i.e., high control and credibility). Indeed, manufacturers inevitably asked themselves the question of text legibility when faced with the task of designing their device's operating system and applications. In addition, we assume that manufacturers and game engine teams whether conducted experiments to validate their documentation before publishing it or relied on scientific literature. Then, we decided to search for guidelines on the official documentation of Microsoft, Google Fonts (named Google in the rest of the paper), Valve (HTC Vive), Meta (Oculus), Lynx, Varjo, Magic Leap, Unity, and Unreal Engine. The first seven are major device manufacturers, while the last two are the two main game engines often used to develop Mixed Reality (MR) applications. We conducted the search manually by screening the hierarchy of their documentation for developers. Note that manufacturers focus on the documentation of their devices. In the same way, game engines focus on the documentation of devices they support.

6.2.2 Digital Libraries and Search Terms

According to the online Cambridge Dictionary, legibility is defined as “the fact of being easy to read, or the degree to which something is easy to read”[1]. In contrast, readability is defined as “the quality of being easy and enjoyable to read”[2]. In other words, readability extends the more restricted legibility concept with User Experience (UX) considerations. Nevertheless, we observed that the two terms are sometimes interchanged in the literature. In addition, the definition of readability is sometimes extended to include the ability to understand text as with language research. In this survey, we rely on the definition of text readability as given by the Cambridge Dictionary, but we included legibility in our search as well since they are used interchangeably in the literature.

Regarding White Literature, we focus our initial search on the IEEEXplore, ACM DL, and ScienceDirect digital libraries. All three are popular among researchers related to computer science domains [Kitchenham et al., 2015]. In addition, the papers selected in the work of Erickson et al. [2020a] were also included. Three concepts are at the center of our research: the asset, the studied property, and the application domain. Except for the last concept, we decided to restrain our search terms to as few as possible to prevent digital libraries from returning too many non-related papers. As discussed by Erickson et al. [2020a], the terms related to the asset and studied property are common to various domains, which rapidly increases the number of articles yielded by an automated search. For example, the term “visual” can take on different meanings depending on the terms with which it is associated. However, we mitigated this limitation by complementing the search query with other literature retrieving techniques (i.e., snowballing and coverage assessment). The search terms were applied to the title, abstract, and keywords of papers in the literature. All articles published before the end of 2022 have been considered. The complete search query is given below:

- Text AND
- (Readability OR Legibility) AND
- (“Augmented Reality” OR AR OR “Mixed Reality” OR MR OR “Virtual Reality” OR VR OR “Head-mounted display” OR HMD OR “Head-worn display” OR HWD)

6.2.3 Inclusion and Exclusion Criteria

To assess the relevance of the identified sources, we defined inclusion and exclusion criteria. A document is relevant if it satisfies the inclusion criteria and none of the exclusion criteria. The only inclusion criterion is defined as follows: the document must focus on text readability in AR or VR on HMDs. The exclusion criteria and their rationale are as follows:

1 Not written in English: The standard language for white literature is English.
2 Duplicated: We consider two documents as duplicate only if they present the same research published at the same venue (i.e., our search may return the

---

[1] https://dictionary.cambridge.org/dictionary/english/legibility?q=Legibility (Accessed 26 November 2023)

[2] https://dictionary.cambridge.org/dictionary/english/readability?q=Readability (Accessed 26 November 2023)

same article multiple times due to cross-references in digital libraries) or if the authors clearly indicate that two papers are duplicate.

3 [For White Literature]Not peer-reviewed: We consider only documents published in conferences or journal outlets that mandate peer-reviewing.

4 [For White Literature]No evaluation of the text parameters: By requiring documents to include an evaluation, we ensure that they present novel research based on findings validated with users.

5 [For Grey Literature]No official documentation of an HMD manufacturer or a game engine used to develop MR applications: As discussed in Section 2.1, we restrain our search to sources with high control and credibility.

6.2.4 Snowballing phase

We conducted a phase of reverse and forward snowballing on the relevant papers yielded by the search query. The inclusion was made again based on the criteria. For the forward method, we obtained the citations on Google Scholar and Scopus. We did not repeat the snowballing phase since the majority of relevant papers returned by this phase were already obtained at the beginning of the process. The longer the process went on, the fewer papers appeared that matched our criteria. Additionally, we ended the survey with a coverage check phase, which assesses the completeness of our protocol.

6.2.5 Coverage check

To assess the completeness of the review, we conducted a keyword search on three popular indexers, namely Google, Google Scholar, and Scopus. The first one assesses coverage for the Grey Literature, while the last two assess coverage for the White Literature. To this end, we updated our search query to include selected works that do not match it. For each group of terms associated with the same concept that a paper does not match in its title, abstract or keywords, we searched the terms it uses to discuss on the concept. Based on the terms obtained, we selected those which best present the concept and appear the most. However, we would have liked to add the words “reading” and “visual”, but for the same reason explained previously, they yield too many non-related works. Therefore, the search query used on Google indexer combined the name of each company with a synonym of the word “Text” or “Readability” (see terms used below) or “UI interface” or “Guidelines” or “Best practices”. The last three are common terms in developer documentations. The final search query used to conduct the coverage check on Google Scholar and Scopus indexers was:

- (Text OR Typography OR Typeface OR Characters OR Fonts) AND
- (Readability OR Legibility) AND
- (“Augmented Reality” OR AR OR “Mixed Reality” OR MR OR “Virtual Reality” OR VR OR “Head-mounted display” OR HMD)

Nevertheless, considering the important number of results that Google Scholar usually returns, we stopped after 100 consecutive papers were excluded. For the search on Google, we limited to five pages.

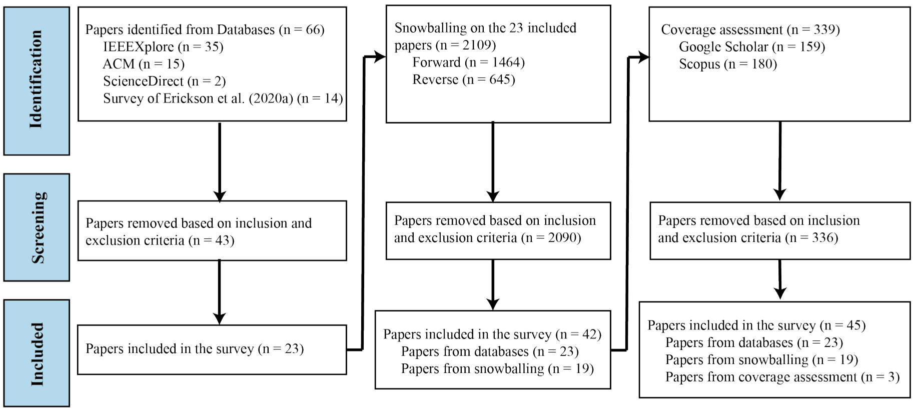


Figure 6.1: Summary of the White Literature collection process.

6.2.6 Conducting the review

We began by collecting papers from the White Literature. The process is illustrated in Figure 6.1. Based on the initial search query, we obtained 52 results published before the end of 2022: 35 from IEEEXplore, 15 from ACM DL, and 2 from ScienceDirect. To these must be added the papers selected in the review by Erickson et al. [2020a], resulting in a total of 66. We applied the inclusion and exclusion criteria and kept 23 of these papers for our survey. Then, we conducted a phase of snowballing as explained in Section 6.2.4. We collected 2,109 papers in total, 645 papers from the reverse snowballing and 1,464 from the forward. We applied the inclusion and exclusion criteria and included 19 additional papers. Finally, we conducted an additional coverage check on two general academic digital libraries to capture potentially missed papers. This process returned 339 papers, 3 of which were added to the set. The final set included in the survey contains 45 papers.

Regarding Grey Literature, we found guidelines on text rendering only in the Microsoft [2023a,b,c], Google Fonts [2023]; Google for Developers [2017], and Oculus Oculus Developers [2023a,b] documentation. The additional search on the Google indexer returned no additional results except websites that do not match our criterion of a device manufacturer or game engine.

6.3 Results by Text Parameter

This section summarizes the guidelines coming from the selected papers and pages of Grey Literature. We recorded 24 articles on OST, 17 on VST, and 4 on both types of screens. Table 6.1 lists the 45 papers by indicating their domain, their types of devices, and the parameters of the text studied. In this section, we classified them according to the text parameter they focus on. We carefully read each paper and each page of Grey Literature to determine the text parameters that were being studied. Once the list was obtained, we tagged each paper as shown in Table 6.1. We defined 7 parameters: the color, the text drawing style, the font-style, the font-size, the anchor, the position (including depth), and the text appearance and segmentation. The last parameter is a category that includes all parameters less frequently discussed in

	Augmented reality	Virtual reality	Optical see-through	Video see-through	Color and/or contrast	Text drawing style	Font type	Font size	Anchoring	Position	Text appearance and segmentation
Gabbard et al. [2006]	X		X		X	X				X	
Gabbard et al. [2007]	X		X		X	X					
Renkewitz et al. [2008]	X		X		X			X			
Gabbard et al. [2010]	X		X		X						
Fiorentino et al. [2013]	X		X		X	X					
Sridharan et al. [2013]	X		X		X						
Hincapie-Ramos et al. [2014]	X		X		X						
Hincapie-Ramos et al. [2015]	X		X		X						
Borg et al. [2015]	X		X					X	X		
Zhao et al. [2017]	X		X		X		X	X			
Rzayev et al. [2018]	X		X							X	X
Blanc-Goldhammer et al. [2018]	X		X		X						
Chang et al. [2019]	X		X					X			
Klose et al. [2019]	X		X						X	X	
Kim et al. [2019a]	X		X		X						
Gabbard et al. [2019]	X		X					X		X	
Woodward et al. [2020]	X		X						X	X	
Fukushima et al. [2020]	X		X						X		
Rosilius et al. [2021]	X		X		X						
Erickson et al. [2021]	X		X		X					X	
Falk et al. [2021]	X		X		X	X					X
Koide et al. [2022]	X		X							X	
Arefin et al. [2022]	X		X							X	
Lee et al. [2023]	X		X						X	X	
Gattullo et al. [2014]	X		X	X	X	X					
Debernardis et al. [2013]	X		X	X	X						
Gattullo et al. [2015]	X		X	X	X	X					
Kruijff et al. [2019]	X		X	X	X			X			
Orlosky et al. [2013]		X		X						X	
Grout et al. [2015]		X		X		X					
Dingler et al. [2018]		X		X	X		X	X		X	X
Hoffman et al. [2019]		X		X				X			
Tsunajima et al. [2020]		X		X				X			X
Wang et al. [2020]		X		X			X	X			X
Wei et al. [2020]		X		X		X					X
Kojic et al. [2020]		X		X	X			X			
Erickson et al. [2020b]		X		X	X						
Buttner et al. [2020]		X		X						X	
Kobayashi et al. [2021]		X		X	X		X	X	X		
Shimizu et al. [2021]		X		X						X	
Rzayev et al. [2021]		X		X					X		X
Dewitz et al. [2021]		X		X				X			
Agić et al. [2022]		X		X	X		X	X		X	
Kojic et al. [2022]		X		X	X	X				X	X

Kobayashi et al. [2022]		X		X			X				
Total	28	17	28	21	24	9	6	14	7	15	8

Table 6.1: The environment, the devices, and the text parameters studied in each selected paper.

the literature such as text length, line space, and text weight. Finally, we read all the papers and pages of Grey Literature to extract atomic guidelines and grouped those that are the same. Nevertheless, although contrast is not a text parameter since depending on the color, it will be discussed in first to bring up some important key points for understanding the text drawing style and color parameters. The information of each section is presented in the following order. First, useful concepts and results shared between the two types of displays and coming from the White Literature are discussed. Second, results limited to one type of display are described. Last, the information coming from the Grey Literature is addressed.

6.3.1 Contrast

Since AR consists of the combination of the real and virtual worlds, two of the environmental factors that affect readability are the background texture and the background illuminance. Leykin and Tuceryan [2004] demonstrated that background variations only affect readability when the text contrast is low. Debernardis et al. [2013] and Gattullo et al. [2015, 2014] observed the need for a minimal contrast ratio between the text and the background on the two display types. However, because of their intrinsic rendering characteristics, the two display types react differently to the contrast requirement between the text and the background. On VST displays, the background illuminance is normalized by the camera. Therefore, the background illuminance strongly affects OST displays but not VST displays [Gattullo et al., 2015, 2014]. On OST displays, the brighter the background, the less the colors will appear.

Specific to VST displays Kojic et al. [2020] recommended a minimal ratio of 7:1 as proposed by the Web Content Accessibility Guidelines 2.0 (WCAG2.0)[3] and to avoid using a too high contrast ratio. They hypothesized that bright color on a screen so close to the eyes will emit too much light. The formula they used to calculate the contrast ratio was the one from the WCAG2.0[4], based on the relative luminance of the foreground and background colors (i.e., calculated on the quantity of red, green, and blue in the linear RGB color space).

Specific to OST displays Blanc-Goldhammer and MacKenzie [2018] suggested a minimal ratio of 1.6:1 calculated as the ratio between the text luminance and the background luminance (i.e., the real luminance emitted by the two elements).

---

[3] http://www.w3.org/TR/2008/REC-WCAG20-20081211/#visual-audio-contrast7 (Accessed 26 November 2023)

[4] https://www.w3.org/TR/WCAG20-TECHS/G17.html (Accessed 26 November 2023)

(a) Billboard style (b) Outline style

(c) Shadow style (d) Plain style

Figure 6.2: Examples of text drawing styles: (a) billboard style, (b) outline style, (c) shadow style, and (d) plain style.

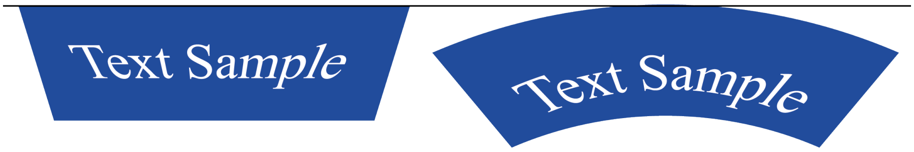
From Grey Literature As expected, the three companies recommend a minimal contrast on the two types of display and note the impact of background illuminance on OST displays. They also suggest preventing too bright experiences.

6.3.2 Text Drawing Style

Through the included papers, four text drawing styles were evaluated: the billboard (see Figure 6.2a), the outline (see Figure 6.2b), the drop shadow (see Figure 6.2c), and the plain text styles (see Figure 6.2d). The billboard style employs a plain color box behind the text to isolate it from the influence of the background. Since the occlusion created by the box must not be avoided, it is recognized as the most effective style [Fiorentino et al., 2013; Gabbard et al., 2006, 2007; Gattullo et al., 2015]. In comparison, the outline (resp. drop shadow) style draws a border (resp. shadow) around the letters of the text. Lastly, for the plain text styles on both display types, the efficiency depends on the chosen color, the contrast between the color and the background, and, for OST displays, the color illuminance (i.e., prefer a color with a high illuminance) [Gabbard et al., 2006, 2007; Gattullo et al., 2015, 2014].

Specific to VST displays Kojic et al. [2022] observed the users' preference for the flat billboard (see Figure 6.3a) but a better performance on the curved one (see Figure 6.3b). Grout et al. [2015] identified a positive correlation between the interest in the curvature and the size of the text sample used. Wei et al. [2020] determined that the curve angle must be at maximum between 50° and 60° around one axis rather than two. Concerning the outline style, adding a minimal outline (e.g., 1px) improves readability, but increasing it too much does not improve readability further [Gattullo et al., 2015, 2014].

Specific to OST displays Falk et al. [2021] recommended the use of a solid billboard rather than a semi-transparent one. Regarding the outline and shadow styles, Gabbard et al. [2007] presented results that seem to indicate an equivalence between the two. Additionally, Gattullo et al. [2015, 2014] observed the inefficiency of the



(a) Flat billboard (b) Curved billboard

Figure 6.3: Examples of flat billboard (a) and curved billboard (b) seen from the top. The horizontal black line helps to demonstrate the curve.

outline style even with the largest outline. Fiorentino et al. [2013] determined that combining the outline style and billboard style is inefficient since the two styles are better used alone than together.

From Grey Literature Microsoft [2023a,b,c] recommends the use of a billboard to improve the readability of text. Google Fonts [2023] and Google for Developers [2017] suggest the curvature of the UI elements, such as billboards, on a circle centered behind the user for a smooth presentation.

6.3.3 Color

In addition to color itself, we must begin by clarifying the concept of color polarity. Positive polarity (i.e., light mode) consists in using dark text on a light background, and negative polarity (i.e., dark mode) consists in using light text on a dark background [Buchner et al., 2009].

Specific to VST displays Dingler et al. [2018] and Kojic et al. [2022] recommended respectively the negative and positive polarities both based on users' preference. Erickson et al. [2020b] made an in-between by advising the negative polarity in case of dark lighting and, inversely, the positive polarity in case of light lighting. They also suggest preferring a bright lighting (i.e., white = RGB(255, 255, 255)) rather than a dim lighting (i.e., white = RGB(25.5, 25.5, 25.5)). For the outline and plain style, Gattullo et al. [2015, 2014] recommended the positive polarity. Regarding color itself, Debernardis et al. [2013], Gattullo et al. [2015], and Kruijff et al. [2019] advised a white text on a blue billboard. If the color has a semantic, they suggested assigning the color to the billboard with a white text or a black text.

Specific to OST displays The impact of background illuminance implies a preference for the negative polarity [Debernardis et al., 2013; Erickson et al., 2021; Falk et al., 2021; Gattullo et al., 2014; Kim et al., 2019a; Zhao et al., 2017]. A light text allows a more important contrast with the background and, inversely, a too light background does not allow any color to create the necessary contrast. Then, for a plain text style, the colors to be favored are for example cyan, green, and white [Fiorentino et al., 2013; Gabbard et al., 2006, 2007; Zhao et al., 2017]. In the case of the billboard style, the contrast between the text and the box is more important than the one between the box and the background [Debernardis et al., 2013; Fiorentino et al., 2013; Gabbard et al., 2007; Rosilius et al., 2021]. Inversely, for the outline and drop shadow styles, the contrast between the text and the drawing style is less important than the one between the drawing style and the background [Gabbard et al., 2007]. In any case, black text must be avoided on a billboard since the black color is transparent on OST displays. The texture of the background will appear, making the readability correct only in the case of a sufficiently low background illuminance, i.e., when the background texture may be considered as black [Debernardis et al., 2013; Erickson et al., 2021; Kim et al., 2019a; Rosilius et al., 2021]. Regarding color itself, Debernardis et al. [2013], Gattullo et al. [2015], and Kruijff et al. [2019] advised a white text on a blue billboard. If the color has a semantic, they suggested assigning the color to the billboard with a white text or to the text with a white billboard.

Additionally, Gabbard et al. [2006, 2007] explored three algorithms to adapt the color to the background: the complement, the maximum HSV complement, and the maximum brightness contrast. The first is the inverse color in the RGB color space. The second is the saturated complement in the HSV color space. The third acts on the Y component (i.e., the spectral luminous efficiency function) of the color in the XYZ color space. The maximum brightness contrast performs better than the other two. However, their efficiency to improve the readability depends on the background. Thus, they were evaluated as globally less practical than the billboard style and the green plain text. In addition, they can be effective with the outline, drop shadow, and plain text styles, but not with the billboard style. As explained above, for the latter, it is the contrast between the text and the box, and not between the text and the background, that is important. In the same way, Gabbard et al. [2007] and Fiorentino et al. [2013] reported that the billboard style must not be mixed with the outline style.

Finally, Gabbard et al. [2010], Sridharan et al. [2013] and Hincapie-Ramos et al. [2014, 2015] studied the blending of colors. For their part, Gabbard et al. [2010] analyzed the hue movements of several colors depending on different backgrounds. They concluded that white backgrounds seem to have less impact on the hue than color backgrounds. Sridharan et al. [2013] and Hincapie-Ramos et al. [2014, 2015] used similar analyses to propose a correction algorithm that defines the color to assign to the text to obtain a specific color once blending with the background. The algorithm consists in searching step-by-step for a color which will produce the desired color in a binned color gamut. Hincapie-Ramos et al. [2014, 2015] proposed three implementations: the algorithm itself, the algorithm limited to the bins being a sufficient contrast with the background, and the algorithm associated with a billboard displayed only when the contrast is too low.

From Grey Literature Microsoft [2023a,b,c] recommends the use of white text on dark or colored billboards. If the text must be black, they suggest using a bright billboard. Additionally, they advise the use of RGB(235, 235, 235) and RGB(16, 16, 16) respectively for the white and black color to prevent too bright experiences and to make the black visible on OST displays.

6.3.4 Text Appearance and Segmentation

Three reading methods will be discussed in this section: the paragraph presentation, the scrolling method, and the Rapid Serial Visual Presentation (RSVP). The former is the text displayed at once or per paragraph, but the studies rarely used more than one paragraph. The second is a presentation line-by-line. The latter is a method introduced by Forster [1970] that quickly presents the text word-by-word.

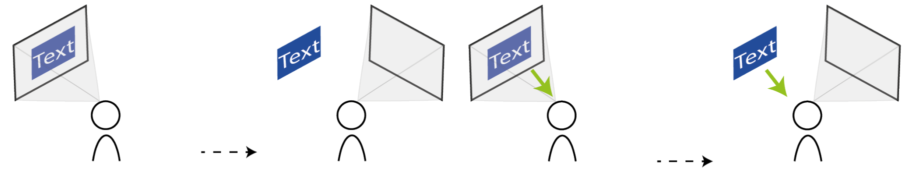
Specific to VST displays Rzayev et al. [2021] advised the use of RSVP for short texts when the user must move in and interact with the world. In contrast, for a primary task without movement or a long text, they suggested the paragraph presentation. In addition, Kojic et al. [2022] recommended one or two maximum columns while Wang et al. [2020] recommended a minimal line spacing of 1.5. Finally, Tsunajima and Nishiuchi [2020] determined that the number of characters in one line is an important factor of readability, too many characters lead to the user being unable to read those on the sides. One effect of this is that more frequent head movements are required to scan the text, implying more fatigue and higher reading time. This issue was also observed by Dingler et al. [2018], who proposed 40 ± 6.6 characters in width and 7.3 ± 1.7 lines. Wei et al. [2020] fixed the field of view of the view box between 25.4° and 28.1° according to users' preference.

Specific to OST displays Rzayev et al. [2018] evaluated two reading methods when walking and sitting. The results showed a preference for the scrolling method when walking and for the RSVP method when sitting. Falk et al. [2021] recommended writing the text in several short lines rather than a single long line.

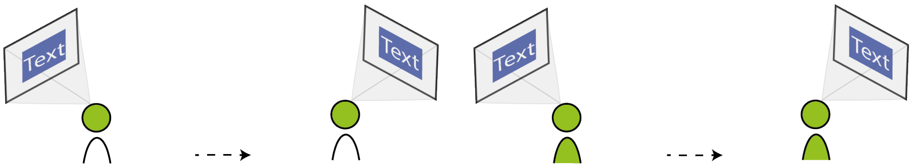
From Grey Literature Google Fonts [2023] and Microsoft [2023a,b,c] recommend using 2D text rather than 3D text because the extrusion deteriorates the readability of the text. Additionally, at a small font-size, they do not suggest thin or light weight text as it is more sensible to flickers and vibrations. Conversely, huge bold fonts do not allow proper discernment of the letters. In terms of letters recognition, Google Fonts [2023] advises ensuring sufficient space between them and to not use a halo that causes blur. Finally, Oculus Developers [2023a,b] recommends left-aligning the text without justifying it, making text lines short, and subdividing text into sections.

6.3.5 Anchor

An anchor is the element to which another element is attached to define its position and rotation within the virtual world. We found four different anchors in the included papers: the world-anchor (see Figure 6.4a), the edge-anchor (see Figure 6.4b), the screen-anchor (see Figure 6.4c), and the body-anchor (see Figure 6.4d). The first fixes the position and rotation of the text at a specific location in the world. The second is equivalent to the first, with the text oriented in permanence towards the user. The third fixes the position and rotation of the text relatively to the user's head. The last fixes the position and rotation of the text relatively to the user, but contrarily to the previous one, it is related to the body and not the head.



(a) World-anchor (b) Edge-anchor



(c) Screen-anchor (d) Body-anchor

Figure 6.4: Illustration of the four anchors: (a) world-anchor, (b) edge-anchor, (c) screen-anchor, and (d) body-anchor. The dotted arrow indicates that the user has moved. The green arrow in edge-anchor (b) indicates that the text is always oriented toward the user. The green color on the user's body (c, d) indicates which part of the user's body the text follows.

Specific to VST displays Rzayev et al. [2021] recommended the edge-anchor or screen-anchor for the RSVP method and the edge-anchor or world-anchor for the paragraph presentation. However, Kobayashi et al. [2021] observed a preference for the world-anchor rather than the edge-anchor.

Specific to OST displays Woodward et al. [2020] and Klose et al. [2019] recommended the world-anchor or body-anchor for non-essential information and everyday use, while the screen-anchor should be used for more important information or dual tasks (e.g., walking and reading simultaneously) needing permanent visual monitoring. Fukushima et al. [2020] suggested, for a walking user, the world-anchor when head movement is not critical and the screen-anchor when head movement is critical. Borg et al. [2015] explained that when walking, the human body activates some mechanisms to counterbalance movements such as those of the head and stabilize the user's view. According to them, it is a reason why the world-anchor should be favored over the screen-anchor. The latter generates retinal slip detrimental to the readability. Lee et al. [2023] recommended the body-anchor rather than the screen-anchor for textual notifications since it is better for comprehension and walking performance.

From Grey Literature Google Fonts [2023] recommends the use of screen-anchor only for crucial information or short texts.

6.3.6 Position

Specific to VST displays Shimizu et al. [2021] tested different methods to adapt the position of the text depending on the user point of attention. The results showed that

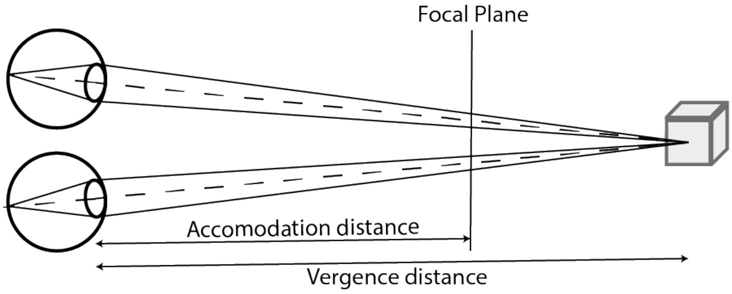


Figure 6.5: Illustration of vergence and accommodation distances.

playing on the vertical and horizontal axis is better than playing on the depth. In their paper, the vertical and horizontal axis represent the surface of a sphere centered on the user while depth corresponds to the sphere radius. Orlosky et al. [2013] observed a user tendency to place text 5.6° below screen center. Similarly, Dingler et al. [2018] observed a text position on a billboard at -1 ± 2.6° from the horizontal. They also define two formulas for minimal and maximal distance vergence (see Equations 6.1 and 6.2) to be in acceptable range of vergence-accommodation conflict. The vergence-accommodation conflict occurs when the vergence distance (i.e., distance to the object) mismatches the accommodation distance (i.e., focal distance) such as on augmented and virtual reality headsets (see Figure 6.5). d_v is the vergence distance, d_f is the focus distance, m_{near}, m_{far}, T_{near}, and T_{far} are the slopes and intercepts of the upper and lower bounds of vergence-accommodation comfort when considering vergence and focus on diopters. Nevertheless, they observed that users often chose a distance beyond $d_v$max and then proposed a distance between 1 meter and 10 meters with an ideal distance of 3 meters on an HMD with a focal distance of 1.3 meters. Kojic et al. [2022] recommended a depth of 6.5 meters rather than 1.5 meters for a panel of long text. Finally, Buttner et al. [2020] determined that a rotation of the text above 60° around the vertical axis (left, right) impacts significantly the readability.

$$d_v\text{min} = \frac{d_f m_\text{near}}{1 - T_\text{near} d_f} \tag{6.1}$$

$$d_v\text{max} = \frac{d_f m_\text{far}}{1 - T_\text{far} d_f} \tag{6.2}$$

Specific to OST displays Research on OST displays is focused on the screen-anchor position. Klose et al. [2019] and Rzayev et al. [2018] determined that, for secondary information or too much complexity in the background, the text should be placed at the top-right, and, for primary information or dual-task needing a permanent monitoring, the text should be placed at the bottom-center. Lee et al. [2023] recommended placing notifications at 10° under the horizontal line. Koide et al. [2022] advised the use of a vertical alignment based on the head-movement for smaller depth (0.45 meter) when walking. However, they observed that its utility decreased

Type of text	Distance to user	Viewing angle	Text state	User's state	Examples
Text in HUD	0.5-1.5m	Fixed	Stationary, moving	Still, moving	Time, notifications, text, updates, music track name
Text for long reading	0.5-5m	Fixed, variable	Stationary	Still	Descriptions, articles, long-form reading
Sticky info text	0.5-5m	Variable	Stationary	Still	Info over real-world objects
Signage text	Beyond 5m	Variable	Stationary	Still, moving	Signages, billboards
Responsive text	0.5-beyond 5m	Variable	Stationary, moving	Still, moving	Navigation markers, responsive ads
Ticker text	0.5-beyond 5m	Fixed, variable	Moving	Still, moving	News, long-running info, ads

Table 6.2: Propositions of text placement according to the type of text. (Source: Google Fonts [2023])

with increasing depth (0.75 meters and 1.5 meter). For the depth, Woodward et al. [2020] advised against a too long depth (0.5 meter) because the information is then considered secondary and does not need to remain fixed to the user view. So, it is recommended to use the world-anchor in this case. Gabbard et al. [2019] recommended a small distance (e.g., 2 meters) to prevent context and focal switching. Arefin et al. [2022] concluded that context switching had no effect on task performance but increases fatigue.

From Grey Literature Google Fonts [2023] defines the comfortable area of view between 30-35° and 120° when rotating the head. The center of the view is ± 6° under the horizontal line. In Table 6.2, they propose the placement of text according to the type of information. The viewing angle defines whether the text is always facing the user even if they move around it. The text state determines whether the text moves with the user. The user's state indicates whether the user remains still or is moving. The distance to the user is also indicated in the table. On this topic, Microsoft [2023a,b,c] recommends a minimum of 0.4 meters and a comfortable area between 1.25 meters and 5 meters. Oculus Developers [2023a,b] indicates that an arm length of depth is sufficient.

6.3.7 Fonts-Type

Specific to VST displays Dingler et al. [2018] recommended the Arial (sans-serif) font for English text rather than Times New Roman (serif) font. Kobayashi et al. [2021, 2022] focused on Japanese characters. In their first work, they observed that Yu Mincho (serif) font is significantly better than Yu Gothic (sans-serif) font. Their later work highlighted a tendency for the Antigothic font family to be more readable and less fatiguing than the Mincho and Gothic font families.

Specific to OST displays Zhao et al. [2017] suggested the use of Verdana (sans-serif) font compared to Times New Roman (serif) for people with moderate or severe low vision.

From Grey Literature Oculus Developers [2023a,b] indicates that sans-serif fonts are more legible than serif or stylized fonts. However, Google Fonts [2023] agrees with that statement only for high-contrast serif typefaces. Moreover, they recommend low to moderate-contrast typefaces that require fewer pixels for presenting details. In general, they recommend using fonts with wider letters and that are not condensed due to the important impact of the depth on the readability. For that, they advise against Helvetica and Univers fonts. Yet, Microsoft [2023a,b,c] recommends the former with the Segoe UI and Arial fonts. Additionally, Microsoft [2023a,b,c] recommends limiting to two fonts family. Finally, Oculus Developers [2023a,b] suggest using OpenDyslexic, Arial or Comic Sans fonts.

6.3.8 Fonts-Size

Whether on OST or VST displays, Kruijff et al. [2019] observed that users preferred a font-size larger for text at the edge of the field of view than for text at the center. To describe writing sizes, we have converted all paper suggestions to the angular unit named Distance-Independent Millimeters (dmm) [Google for Developers, 2017]. dmm consists of x millimeter at one meter of distance where x is the coefficient of the unit (e.g., 1dmm = 1 millimeters at one meter or 2 millimeters at two meters [Google for Developers, 2017]).

Specific to VST displays In regard to the Latin alphabet, Dingler et al. [2018] recommended a font-size of 32dmm ± 11dmm for a capital letter and 41dmm ± 14dmm for a body text. Similarly, Hoffman et al. [2019] suggested a comfortable font-size of 26dmm for label (capitalized) and body text. Additionally, they observed a correlation with the contrast of the text. The more the contrast decreases under the minimal value, the bigger the font-size must be to compensate. As for Kojic et al. [2020], they obtained similar results for short texts (2 words) with a font-size of 30dmm, but for longer texts between 21 and 51 words, the needed font-size is reduced to 17dmm. The explanation given by the authors was that participants consider the short text as a title and, therefore, gave it a larger preferred size. Dewitz et al. [2021] aligned themselves with these results, while they found a comfortable font-size of 15dmm. Additionally, Agić et al. [2022] proposed a font-size of 32-40pt at 5 meters, 36-56pt at 10 meters, and 56-68pt at 15 meters on an HTC Vive.

Considering logographic alphabet, Kobayashi et al. [2021] recommended a visual angle between 19dmm and 21dmm while Wang et al. [2020] proposed a font-size of 12pt or more at 0.5 meters on a Samsung Gear VR.

Specific to OST displays Erickson et al. [2021] observed a positive correlation between the depth and the font-size. Borg et al. [2015] recommended a font-size between 5dmm and 17dmm when standing. If the user must walk, the lower limit becomes 9dmm. They also observed that the font-size of screen-anchor text was more affected at the extremity of the range than world-anchor text. Renkewitz et al. [2008] observed a minimal font-size of 16dmm to reach a maximal 2-second recognition time. They also concluded that recommendations on font-size for desktop may be applied to HMDs. Gabbard et al. [2019] observed a minimal font-size of 5dmm for text at 6 meters of depth. Finally, Zhao et al. [2017] recommended a

Type of text	Weight	Font-size
Headline	Regular	40dmm
Title	Medium	32dmm
Subheading	Regular	28dmm
Body 2	Medium	24dmm
Body 1	Regular	24dmm
Caption	Regular	20dmm
Button	Medium	24dmm

Table 6.3: Font-sizes recommended depending on the type of text. (Source: Google Fonts [2023])

minimal font-size of 100px at 3 meters on an Epson BT-200 for people with moderate or severe low vision.

Considering logographic alphabets, Chang et al. [2019] observed a mean font-size of 4dmm, but to reach the 99th percentile legibility threshold, the font-size must be to 7dmm. However, they observed, on the one hand, that the more strokes in the character, the more the font-size must be large, and, on the other hand, the more components in the character (i.e., separated groups of strokes), the easier to read the character is.

From Grey Literature First, Microsoft [2023a,b,c] indicates not to rely on computer font-size because of the impact of depth. Second, Oculus Developers [2023a,b] suggests beginning with a font-size of 10% of the screen and adapt to the requirements. Microsoft [2023a,b,c] defines a comfortable font-size between 11dmm and 14dmm at 0.45 meters and 11dmm and 13dmm at 2 meters. Finally, the font-sizes recommended by Google Fonts [2023] are presented in Table 6.3.

6.4 Results by Aggregation of Text Parameters

We aggregated the parameters that were intrinsically related. On the one hand, these are the text drawing style and the color (see Figure 6.6), and, on the other hand, these are the anchor and the position (see Figure 6.7). As for the other parameters, they do not depend on a sequence of decisions to be taken, are not sufficiently studied, or are subject to too many contradictions in the literature to propose a clear decision tree. In addition, while it is true that contrast and font-size are related, as observed by Gattullo et al. [2014], it is a question of balancing the two to reach a sufficient readability threshold. Similarly, font-size and depth are related, since the font-size is expressed in angular units, it will automatically adapt to the choice made for the distance between the user and the text.

The two decision trees are represented in Figures 6.6 and 6.7. They start with the green node “Start” and finish with the green nodes containing the list of guidelines for the choices made according to the route. The yellow blocks represent decisions, while the blue blocks represent major choices to respect. Numbers in square brackets refer to a set of documents that are the source of the guideline. In Figure 6.6, the choice of device is the first question, as it determines the rest of the process for the text drawing style and color. On the contrary, in Figure 6.7, we hypothesized that results for both types of display may be merged since it does not seem to depend on device-related criteria. The decision trees illustrate that there is still room for further research to improve the guidelines. For example, in Figure 6.6, we observed, on a VST display, that for a billboard without semantic on the color (see bottom-right part of the figure), the designer must choose which literature they follow among three exclusive possibilities: (1) [Debernardis et al., 2013; Dingler et al., 2018; Gattullo et al., 2015; Kruijff et al., 2019], (2) [Erickson et al., 2020b], and (3) [Kojic et al., 2022].

6.5 Discussion

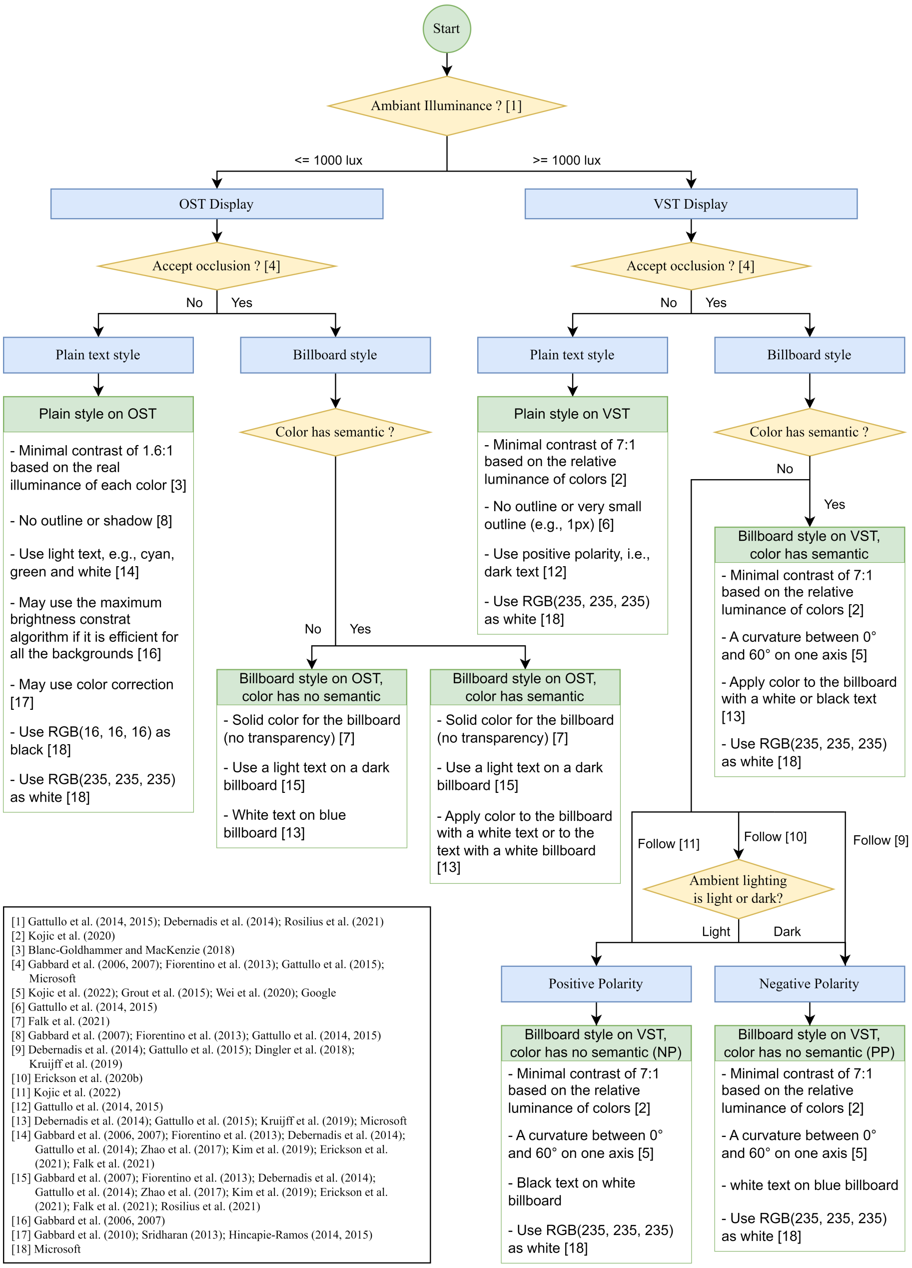
In this section, we provide a detailed discussion on the three research questions. For each, we present the results obtained from our literature review and we elaborate on future research avenues these results open.

6.5.1 Research Question 1.1: Text Parameters

The first research question is: "What are the different text parameters and how to tune them to maximize the readability of a text?". We identified seven parameters in the literature: text drawing style, color, anchor, position, font-type, font-size, and text appearance and segmentation. We observed that the text drawing style and the color, as well as the anchor and the position were respectively intrinsically related. Furthermore, we also found a relationship of balance between the contrast and the font-size. The lower the contrast, the larger the font-size needs to be. In addition, we identified an intrinsic relation between the font-size and the depth, but the use of angular units for the first allows them to be treated separately.

Concerning the drawing style of the text, a consensus was reached in the literature to favor the use of billboards (i.e., a panel box behind the text) except when the occlusion is a problem for the user. In that case, authors recommend the use of a plain text style, since the outline style and shadow style do not present better performance. On VST displays, a small outline can be added to improve readability. Regarding the shadow style, it was studied once on OST displays. In terms of color, on OST displays, the light-additive property implies the use of negative polarity. In comparison, on VST displays, the literature is more divided, advising both positive and negative polarity. However, for both displays, the colors that stand out most are white for the text and blue for the billboard. In case color has semantics, it is recommended to assign the semantically meaningful color to the billboard.

Regarding the anchor of the text, research on VST displays is more concerned with the way the text is presented, while research on OST displays is more concerned with the type of text and the simultaneous movement of the user. Generally, when it comes to short text such as notifications or dashboard-type information, it is recommended to use a screen-anchor approach. For long text or text associated with objects, previous works recommend the use of world-anchor or edge-anchor texts. Nevertheless, some studies demonstrated that the human body has a system for balancing eyesight when walking, which conflicts with the screen-anchor. In terms of position, the center of the view is located at ± 6° below the horizontal line. Text is also recommended at the bottom of the view when attached to the user's



Plain style on OST

- Minimal contrast of 1.6:1 based on the real illuminance of each color [3]
- No outline or shadow [8]
- Use light text, e.g., cyan, green and white [14]
- May use the maximum brightness constrat algorithm if it is efficient for all the backgrounds [16]
- May use color correction [17]
- Use RGB(16, 16, 16) as black [18]
- Use RGB(235, 235, 235) as white [18]

Billboard style on OST, color has no semantic

- Solid color for the billboard (no transparency) [7]
- Use a light text on a dark billboard [15]
- White text on blue billboard [13]

Billboard style on OST, color has semantic

- Solid color for the billboard (no transparency) [7]
- Use a light text on a dark billboard [15]
- Apply color to the billboard with a white text or to the text with a white billboard [13]

Plain style on VST

- Minimal contrast of 7:1 based on the relative luminance of colors [2]
- No outline or very small outline (e.g., 1px) [6]
- Use positive polarity, i.e., dark text [12]
- Use RGB(235, 235, 235) as white [18]

Billboard style on VST, color has semantic

- Minimal contrast of 7:1 based on the relative luminance of colors [2]
- A curvature between 0° and 60° on one axis [5]
- Apply color to the billboard with a white or black text [13]
- Use RGB(235, 235, 235) as white [18]

Billboard style on VST, color has no semantic (NP)

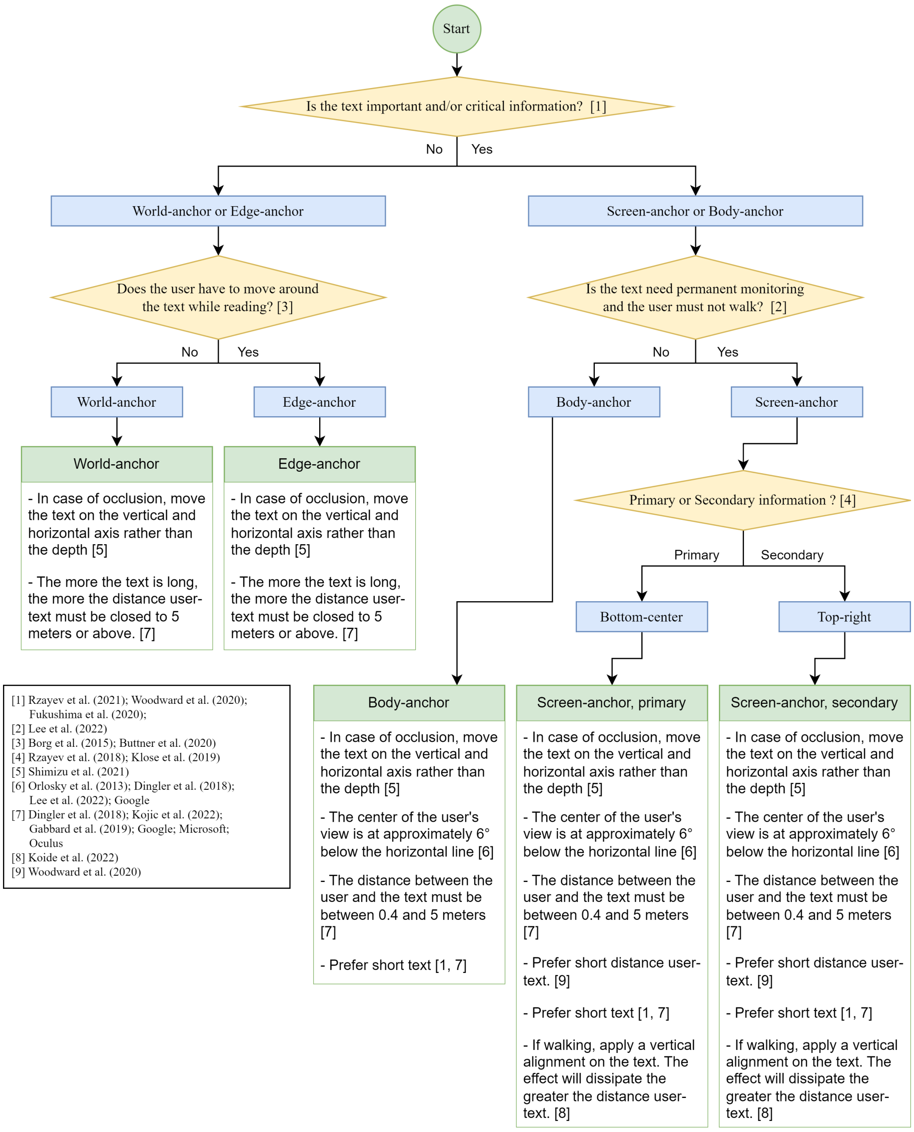
- Minimal contrast of 7:1 based on the relative luminance of colors [2]
- A curvature between 0° and 60° on one axis [5]
- Black text on white billboard
- Use RGB(235, 235, 235) as white [18]

Billboard style on VST, color has no semantic (PP)

- Minimal contrast of 7:1 based on the relative luminance of colors [2]
- A curvature between 0° and 60° on one axis [5]
- white text on blue billboard
- Use RGB(235, 235, 235) as white [18]

[1] Gattullo et al. (2014, 2015); Debernadis et al. (2014); Rosilius et al. (2021)
[2] Kojic et al. (2020)
[3] Blanc-Goldhammer and MacKenzie (2018)
[4] Gabbard et al. (2006, 2007); Fiorentino et al. (2013); Gattullo et al. (2015); Microsoft
[5] Kojic et al. (2022); Grout et al. (2015); Wei et al. (2020); Google
[6] Gattullo et al. (2014, 2015)
[7] Falk et al. (2021)
[8] Gabbard et al. (2007); Fiorentino et al. (2013); Gattullo et al. (2014, 2015)
[9] Debernadis et al. (2014); Gattullo et al. (2015); Dingler et al. (2018); Kruijff et al. (2019)
[10] Erickson et al. (2020b)
[11] Kojic et al. (2022)
[12] Gattullo et al. (2014, 2015)
[13] Debernadis et al. (2014); Gattullo et al. (2015); Kruijff et al. (2019); Microsoft
[14] Gabbard et al. (2006, 2007); Fiorentino et al. (2013); Debernadis et al. (2014); Gattullo et al. (2014); Zhao et al. (2017); Kim et al. (2019); Erickson et al. (2021); Falk et al. (2021)
[15] Gabbard et al. (2007); Fiorentino et al. (2013); Debernadis et al. (2014); Gattullo et al. (2014); Zhao et al. (2017); Kim et al. (2019); Erickson et al. (2021); Falk et al. (2021); Rosilius et al. (2021)
[16] Gabbard et al. (2006, 2007)
[17] Gabbard et al. (2010); Sridharan (2013); Hincapie-Ramos (2014, 2015)
[18] Microsoft

Figure 6.6: Decision-tree to select the right text drawing style and color for a text.



World-anchor

- In case of occlusion, move the text on the vertical and horizontal axis rather than the depth [5]

- The more the text is long, the more the distance user-text must be closed to 5 meters or above. [7]

Edge-anchor

- In case of occlusion, move the text on the vertical and horizontal axis rather than the depth [5]

- The more the text is long, the more the distance user-text must be closed to 5 meters or above. [7]

Body-anchor

- In case of occlusion, move the text on the vertical and horizontal axis rather than the depth [5]

- The center of the user's view is at approximately 6° below the horizontal line [6]

- The distance between the user and the text must be between 0.4 and 5 meters [7]

- Prefer short text [1, 7]

Screen-anchor, primary

- In case of occlusion, move the text on the vertical and horizontal axis rather than the depth [5]

- The center of the user's view is at approximately 6° below the horizontal line [6]

- The distance between the user and the text must be between 0.4 and 5 meters [7]

- Prefer short distance user-text. [9]

- Prefer short text [1, 7]

- If walking, apply a vertical alignment on the text. The effect will dissipate the greater the distance user-text. [8]

Screen-anchor, secondary

- In case of occlusion, move the text on the vertical and horizontal axis rather than the depth [5]

- The center of the user's view is at approximately 6° below the horizontal line [6]

- The distance between the user and the text must be between 0.4 and 5 meters [7]

- Prefer short distance user-text. [9]

- Prefer short text [1, 7]

- If walking, apply a vertical alignment on the text. The effect will dissipate the greater the distance user-text. [8]

[1] Rzayev et al. (2021); Woodward et al. (2020); Fukushima et al. (2020);
[2] Lee et al. (2022)
[3] Borg et al. (2015); Buttner et al. (2020)
[4] Rzayev et al. (2018); Klose et al. (2019)
[5] Shimizu et al. (2021)
[6] Orlosky et al. (2013); Dingler et al. (2018); Lee et al. (2022); Google
[7] Dingler et al. (2018); Kojic et al. (2022); Gabbard et al. (2019); Google; Microsoft; Oculus
[8] Koide et al. (2022)
[9] Woodward et al. (2020)

Figure 6.7: Decision-tree to select the right anchor and position for a text.

view. Additionally, most of the text should be displayed at minimum 0.4 – 0.5 meters and maximum 5 - 10 meters. Once again, the type of text will influence the depth at which it is placed. In the results, we hypothesized that the results from both types of display may be merged, but this should be verified in future works.

Considering the font-type, the White Literature seems to indicate that fonts with too much detail, such as the serif family, should be avoided. However, while some Grey Literature sources support this idea, others reject it. Some fonts have been recommended, but again the literature does not always reach consensus on this. Yet, the Arial font remains a standard choice. In terms of font size, all surveyed studies found different numbers. Recommendations vary between 5dmm and 17dmm, with some going as high as 40dmm. This recommendation applies to both Latin and logographic alphabets. It must however be noted that in this survey we presented only comfortable thresholds instead of minimal font sizes. Additionally, some works give their recommendations in a unit that depends on the pixel size, and thus on the device used. These are complicated to compare with other works that use angular units. Last, it is not clear how distance influences the minimum and maximum bounds, even if these are expressed in angular units. Do they remain fixed, or are they dependent on a distance-based function? The vergence-accommodation conflict could potentially play a role in favor of the second proposal, but this remains to be verified in future research.

We grouped under the text appearance and segmentation label all guidelines that have not been sufficiently studied to have their own category, such as line spacing, weight, text length, number of columns, text alignment, and presentation mode. The extent of their impact on readability is unclear and needs further study. Classifying parameters by order of importance could also help designers to better control their experiences, while letting them indulge their artistic creativity. From the results, we can already conclude that contrast, depth, and font size are critically important, but the question remains open for other parameters, calling for novel research in these still shady areas.

In view of the answer to the first research question, we recommend focusing future research efforts on four different avenues: (1) better define the rules around parameters under the text appearance and segmentation label, (2) clarify the range of optimal text size, (3) define which results obtained on VST (resp. OST) displays can also be applied on OST (resp. VST) displays without repeating the study, and (4) define a taxonomy of text usage and associate readability rules to each.

6.5.2 Research Question 1.2: Contextual Constraints

The second research question is: “What are the contextual constraints to consider when designing an application including texts?”. We determined six constraints: the type of display, the background texture, the background illuminance, the user’s task, the dual task such as walking, and the user’s body. However, the type of display and the background illuminance are intrinsically related.

As already discussed, the type of display impacts the way the device renders the text, but also on which devices the guidelines were tested. Gattullo et al. [2015, 2014] and Debernardis et al. [2013] recommended using OST displays with an ambient illuminance of less than 1,000 lux and VST displays with an ambient illuminance of more than 1,000 lux. In the case of OST displays, the light-additive property does not allow the device to be exposed to too much light. In the case of VST displays, the cameras normalize brightness, making dark environments too dark to use the headset. Nevertheless, this does not prevent them from being used below 1,000 lux, but there is a preference for OST displays. Additionally, too low light (10-12 lux) is not suitable for OST displays [Kim et al., 2019a]. Concerning the background texture, it impacts the readability of text only when the contrast is low or when the texture is complex [Gabbard et al., 2006; Leykin and Tuceryan, 2004]. On one hand, as with conventional devices (i.e., computer or mobile), the color difference between the text and the background on which it is read must be sufficient to determine the outline of the letters. On another hand, Gabbard et al. [2006] determined that it was the complexity of the background in the area close to the text that had an impact on its reading, rather than the complexity of the background in general. In their experience, they observed that participants obtained good results with the brick wall because they place the text on a single brick that presents a degree of homogeneity. Therefore, the readability of a text depends on the disturbance of the background texture close to the text, generating variation in contrast between the two. However, the designer can have a control on that using either dynamic algorithms such as the ones proposed by Gabbard et al. [2010], Sridharan et al. [2013], and Hincapie-Ramos et al. [2014, 2015], or classifiers [Leykin and Tuceryan, 2004; Manghisi et al., 2017; Scharff et al., 1999; Tanaka et al., 2008].

Regarding the user's task and dual task (e.g., walking and reading simultaneously), there is no doubt about the impact on text readability. However, there is a need for more in-depth analysis, especially on the user's task. At present, the only interest is in knowing whether the occlusion will be a problem for the user. For the dual task, it was demonstrated that the more the dual task needs attention, the more the readability decreases [Klose et al., 2019]. Additionally, Wei et al. [2020] identified a trade-off between the ease of reading and an immersive reading experience. In this respect, the task and its context can help define the level of immersion required.

Last, similarly to desktop UIs, the user's body has a significant impact. First, when walking, the view is automatically balanced to avoid too much jerking that is incompatible with certain anchoring methods. Second, the vergence-accommodation conflict plays an important role in the reading of text. Dingler et al. [2018] offer formulas for determining a range of acceptability, but the problem is to find solutions that play on the focus of the device. Therefore, in future work, it would be interesting to look at solutions based on more than just the text itself. Apart from this, several studies have reported that the user does not always make appropriate choices in relation to measures such as accuracy and reading speed [Gattullo et al., 2015; Kim et al., 2019a]. Additionally, Zhao et al. [2017] concluded that *"if a sighted person can use the AR glasses while walking, it is also feasible for a low vision person to use the glasses while walking"*. This is due to a similar negative impact of walking on shape and text recognition for sighted and low-vision people.

In regard to the answer to the second research question, we recommend pursuing future research efforts on two main avenues. First, the hardware development of displays should eventually solve several of the above-mentioned problems. For instance, successfully combining the strengths of both types of devices, either with special lenses or with a transparent VST screen. Secondly, the use of algorithms combined with sensor data could help reduce environmental and/or user's metabolic problems such as those trying to counterbalance the walk or the brightness of the background.

6.5.3 Research Question 1.3: Gap Between White and Grey Literature

The third research question is: "What is the gap between the White and Grey Literature in the guidelines proposed on text readability?". Among the sites we visited, only three device manufacturers have guidelines on text readability in their documentation. On the color, contrast, text drawing style, font-type, and font-size, the literatures are similar, but for the rest, they tend to talk about one side but not the other, and vice versa. This observation strengthens the case for including Grey Literature in a literature review like ours. Additionally, it seems that the White Literature is more comprehensive, showing signs of difficulty in transferring information from the scientific side to the practical side. However, we observed that Google Fonts [2023] cites the works of Gabbard et al. [2006] and Gattullo et al. [2015]. Their documentation also appears to be written by, or at least with, an independent researcher [Yadav, 2023]. Just how much of this documentation is based on scientific sources is unclear. It is also possible that the time constraint is playing into the hands of practice ahead of academic research, with solutions being studied in private laboratories. Perhaps, if it is not already the case, documentation should be written more often in collaboration with researchers, and thus, promote popularization through media other than scientific books, articles, and conferences (or at least reference these works in documentation). Another approach would be to offer works like this one, and above all continuing to keep it up to date regularly on a platform, to facilitate access to scientific research in practices and, possibly, directly to the public. However, the question of who would be responsible for such a project remains difficult to answer. Besides, we observed a more rigorous classification of text from the Grey Literature than the White Literature (see Tables 6.2 and 6.3). As already discussed in Section 5.1, it would be interesting to use such a classification in the future, to indicate more clearly what type of information is being studied. In fact, it has been shown that requirements differ according to the role of the text.

6.5.4 Limitations

Variability in experimental protocols makes it difficult to collate and compare results from dozens of papers, since the number of parameters involved is too large. We decided not to apply quality criteria, even if some papers contradict themselves. In such cases, we chose to take the version that was presented in their discussion. It would be interesting in future work to generalize and confirm the results obtained to date. Even more so as the question arises as to whether some of the results achieved on older headsets are still relevant today, given the technological advances of today's headsets. Similarly, the question will arise in the future with recently published papers. Eventually, we will need to be able to determine more precisely the impact of hardware on results. Additionally, the nomenclature associated with text readability domains is diverse and varied, a problem also highlighted by Erickson et al. [2020b].

Furthermore, we limited our scope to papers that studied HMDs. Nonetheless, the impact of HMD's hardware on the parameters is not well-defined and necessitates additional studies. In addition, some works on other devices, such as HUDs, present results that may be applied to HMDs. According to Renkewitz et al. [2008], this is the case for font size guidelines on desktop systems, but it is a conclusion that has not been shared by all works dealing with this text parameter. It is also possible to include all the literature on classifiers that try to determine the readability of text based on images. If we study the criteria used by the best-performing models, we may discover or confirm results obtained via more conventional approaches, such as the work included in this survey.

Finally, we have included results from analyses such as reading speed or number of errors, as well as results from user preferences. However, for the latter, it is well known that many social, cultural, and other factors have a major impact on results. Additionally, Gattullo et al. [2015] and Kim et al. [2019a] demonstrated that users do not always make the choice of performance. A whole area of research remains to be carried out to verify the results and determine the relationship between readability and enjoyment.

6.6 Summary

We conducted a multivocal literature review to identify the list of text parameters that can be tuned to improve text readability, as well as the associated contextual constraints, in order to propose an up-to-date state of knowledge and guidelines. We adopted a multivocal protocol to include the Grey Literature of high control and credibility. In total, we found seven text parameters and six contextual constraints. The most important ones are the contrast (i.e., related to the text drawing style and color), the font-size, the anchor, and the position. They are impacted by the background, the role of the text and the user's body, such as the vergence-accommodation conflict. Nevertheless, the guidelines pertaining to those parameters seem complete and accepted by the community, except for some minor points. The same cannot be said for the other parameters, which are either under-researched or the subject of contradictions among scholars and/or practitioners. In addition, this research also highlighted a lack of communication of guidelines from White Literature to Grey Literature. All in all, this survey has gathered numerous guidelines for text design on OST and VST HMDs. They have been brought together to form a guide, assisting in the choice of the best solution for the context in which text is to be used. This will be of use both for researchers wishing to extend this research, and for designers wishing to ensure the readability of their texts in their applications. Finally, we conclude with the proposition of future research directions to extend the results and to promote them outside scientific publications.

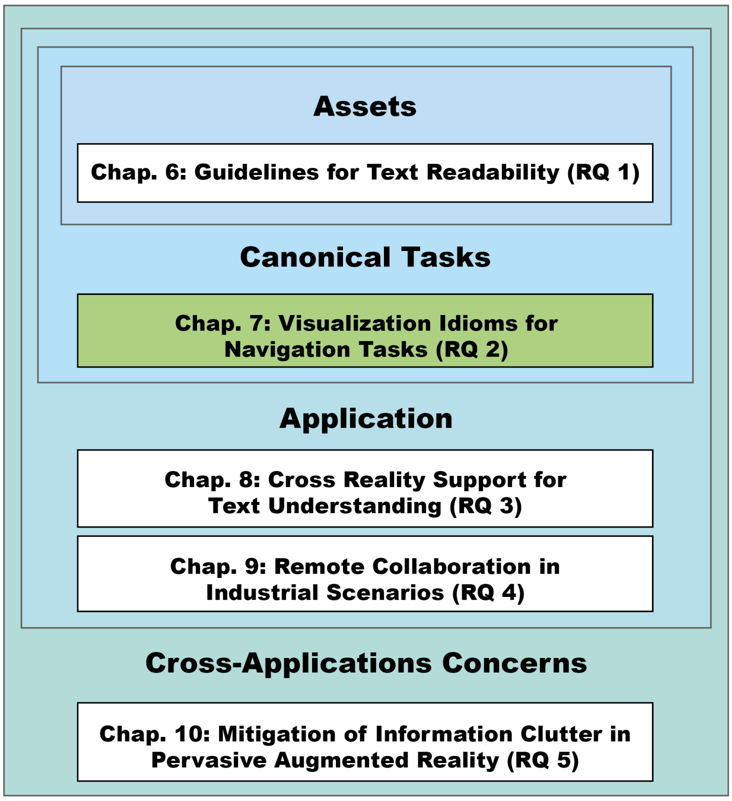
CHAPTER 7

VISUALIZATION IDIOMS FOR NAVIGATION TASKS

7.1 General Introduction

This chapter is dedicated to research question 2: **What are the best visualization idioms that can be used to guide the user to different points of interest?**. Navigation refers to traveling from an origin point *A* to a target destination *B*. The process can involve changing one's gaze direction or engaging in physical locomotion within the surrounding environment. This issue has consistently held significant importance in everyday life and computer applications [Farr et al., 2012]. In the former, humans have always exhibited mobility, leading to the development of tools facilitating navigation and the establishment of cognitive landmarks. In the latter, some applications cannot simultaneously present all the information content due to spatial constraints on display dimensions. In scenarios involving Immersive Augmented Reality (AR), the complexity of the addressed issue is heightened, on the one hand, by the inherent three-dimensional nature of the environment and, on the other hand, by the nature of points of interest (POIs): real or virtual. As a result, conventional techniques for screens, such as zooms, scrollbars, or touch-based scrolling methods, are no longer sufficient, especially when it comes to situated or embedded information.

Navigation in a 3D environment faces two major types of POIs. First, POIs can be out of the user's field of view (a.k.a., off-screen POIs). A naive approach would let the user explore the environment independently, but this would be neither efficient nor intuitive in many applications. On the contrary, numerous visual idioms exist to notify the user of POIs position in the environment. Second, POIs in the user's field of view (a.k.a., on-screen POIs) can be occluded by real or virtual elements. Determining adequately the depth relating to the user can also be a problem. For both types, in numerous applications, it is necessary to help users construct an accurate mental image of their evolving environment. However, an assessment of the strengths and weaknesses of each idiom within the user's task is absent in



the scientific literature. This chapter will focus on off-screen POIs by conducting a Systematic Literature Review (SLR) to comprehensively catalog related visual idioms and discern their efficiency within the user's task. In contrast to the previous chapter, our aim here is more of a theory synthesis [Jaakkola, 2020] than a simple review; the review must help to determine which idiom is to be used when. We formulated the following research questions:

RQ2.1 What visual idioms can support navigation tasks to off-screen POIs?
RQ2.2 What are the navigation tasks used to validate visual idioms indicating off-screen POIs?
RQ2.3 What is the best visual idiom for each navigation task to off-screen POIs?

Publications and supervised works

The content of this chapter is based on my master's thesis and the master's thesis of a student for which I was co-supervisor. In addition, an article based on this chapter is currently being written.

Cauz, M. (2019). Augmented Reality visualization and interaction for maintenance. Master's thesis, University of Namur

- My master's thesis deals with the design of immersive AR applications for industrial maintenance. One chapter was dedicated to a light review of navigation visual idioms for off-screen points of interest. A comparison based on the user's task was introduced. This work was used to settle the base of the survey literature conducted in this chapter.

Hincq, L. (2022). Application de technique de réalité augmentée dans la maintenance industrielle. Master's thesis, University of Namur

- This master's thesis is an extension of my master's thesis, and thus somewhat extended the results presented in Cauz [2019]. It will complete the initial set of visual idioms used in the survey literature conducted in this chapter.

Outline

The organization of this chapter adheres to the subsequent structure. Section 7.2 presented existing works in the navigation field supported by computing science, establishing the need for a new review. Section 7.3 presents the methodology followed in conducting the SLR, following the steps of Kitchenham et al. [2015]. The need for a review and the research questions were already established in this chapter's introduction and background sections. The next main steps consist of developing the protocol (Section 7.3.1, 7.3.2 and 7.3.3), conducting the review (Section 7.3.5), and assessing the quality of the coverage (Section 7.3.4). We report the visual idioms defined in the included papers in Section 7.4. Sections 7.5 and 7.6 report the tasks used to evaluate these idioms and which are the best for each task, respectively, for 2D screens and immersive scenarios. Finally, in Section 7.7, we discuss the implications of our results, limitations, and future works before concluding in Section 7.8.

7.2 Background

Idioms to support navigation tasks and environment awareness are generally divided into overview+detail, focus+context, and visual aids [Chittaro, 2006]. The first consists of two separate views, presenting an overview of the work environment and a detailed view of what the user is paying specific attention to. In conventional computing, these can be separated into two displays or a dedicated part of a single screen. In Immersive AR, only the second solution is available. Still, with the distinctiveness that the detail view corresponds to the user's view, the overview visualization is an element attached to this view or a specific location in the environment. Illustrative instances encompass bird's eye views, the World-in-Miniature (WiM) metaphor, and wordlets. The former, introduced by Fukatsu et al. [1998], entails the real-time streaming of a camera assuming an aerial viewpoint. The second approach encompasses a scaled-down representation of the environment, permitting user manipulation to observe the scene from an omniscient standpoint [Stoakley et al., 1995]. The last, proposed by Elvins et al. [1997], encompasses a 3D representation of small parts of the environment encapsulated within thumbnails, denoted as "wordlets".

Focus+context adds contextual information directly in the detail view, preventing the need to switch attention between two views. This category is itself divided into several possibilities: (1) elision of elements, (2) superimposition of information layers, and (3) distorted view [Chittaro, 2006; Munzner, 2015]. The first two are exclusive for on-screen POIs, as they remove elements and propose multi-levels of details in the user's view, respectively. The third is subject to comprehensive investigation, with the best-known example being the fish-eye perspective, characterized by its ability to magnify objects within the user's field of view. While applying effects to virtual elements is generally feasible, the same does not consistently hold for real elements. The frame can be pre-processed before rendering in the context of Video See-Through (VST) displays. In contrast, designers encounter limitations with Optical See-Through (OST) displays as they cannot apply analogous effects to the physical environment. In addition, the impact of such techniques on the situated and embedded information remains to be defined.

Visual aids add symbols to the user's view to indicate the direction of POIs. As a result, one limitation that becomes clear is the issue of visual overload, depending on the number of elements to be notified to the user and their superposition, whether the symbols are grouped in a single zone or not.

7.2.1 On-Screen Points of Interest

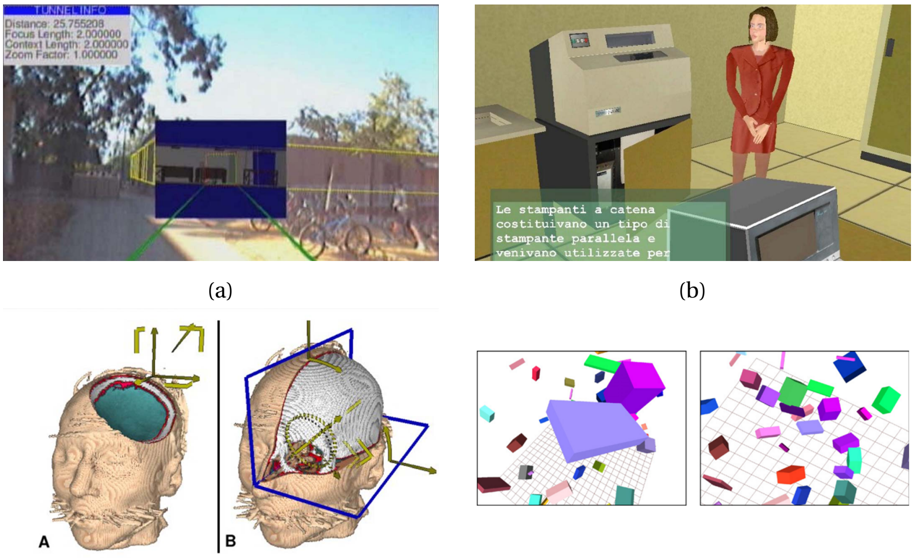
Elmqvist and Tsigas [2008] introduces a triad of factors responsible for the manifestation of occlusion within an environment. Primarily, spatial interaction among distinct elements delineates the spatial proximity existing between said elements. Subsequently, object density pertains to the number of entities present within the environment relative to their sizes and the size of the environment. Finally, the intricacy intrinsic to each element contributes to its potential to obstruct segments of its own or other configurations. In the aforementioned study, Elmqvist and Tsigas [2008] determine five main patterns of methodologies. The first, overview+detail, was already discussed above.

Virtual X-Ray

The virtual X-ray techniques transform obstructing surfaces into imperceptible or partially translucent entities. As a result, the perception of depth is diminished, concurrently leading to an overall augmentation in cognitive workload. This technique can be realized by applying volumetric configurations, as put forth in the study by Viega et al. [1996]. Within these configurations, observable surfaces enclosed by the volumetric structure are extracted, unveiling the concealed structures beneath (e.g., the skeletal composition of a hand). An alternative approach, proposed by Bane and Hollerer [2004], introduces the concept of an "X-Ray Tunnel," delineating a region where these surfaces are omitted (see Figure 7.1a). Lastly, Feiner and Seligmann [1992] investigated diverse techniques for manipulating the opacity of an element to represent it as an occluded part of another element. One approach, for example, is to draw a hole in the last element to give the impression of a missing part that reveals the object of interest.

Tour Planner

The Tour Planner methodologies guarantee the existence of an accessible trajectory to every element of interest within the designated environment (see Figure 7.1b). This approach offers the benefit of maintaining environmental integrity. However, it



(a) (b)

(c) (d)

Figure 7.1: Illustration of various categories of idioms for on-screen POIs: (a) Virtual X-Ray (Source: Bane and Hollerer [2004]), (b) Tour Planner (Source: Chittaro et al. [2003]), (c) Volumetric Probe (Source: McGuffin et al. [2003]), and (d) Projection Distorter (Source: Elmqvist and Tsigas [2007]).

mandates user mobility and adopts an optimized spatial orientation strategy, such as the ones discussed in the following section on off-screen POIs.

Volumetric Probe

The Volumetric Probe encompasses the implementation of a distortion probe, subject to user control, which induces localized transformations upon objects to regulate occlusion effectively. This methodology relies on either the elimination of disruptive elements or the isolation of intended elements. Leveraging a volumetric probe holds the potential to resolve intricacies in even the most challenging scenarios. However, it's important to note that the inherent characteristics of this approach result in the loss of many aspects of the elements (e.g., spatial correlation). Instances include the research conducted by McGuffin et al. [2003], which presents a user-interaction technique enabling precision element segmentation akin to surgical procedures (see Figure 7.1c). Additionally, Kalkofen et al. [2009] contributed to this field with an approach involving explosive diagrams, where all elements of a 3D model are moved from their original location to separate them from each other. The same approach is used, for instance, in furniture assembly drawings.

Projection Distorter

The Projection Distorter pattern employs the process of amalgamating diverse perspectives of targets into a unified view to optimize the potential for uncovering insights. This strategy frequently bears semblance to a focus+context approach (i.e., the view focuses on an element while the others are in a context mode; thus, less detailed), where each viewpoint corresponds to a distinct focus. Besides the implementation facility, the resultant visual presentations may frequently evoke feelings of perturbation and disorientation within the user. This often leads to the limited retention of object properties. This pattern applies to the type of camera projection, for example, orthogonal, perspective [Elmqvist and Tsigas, 2007] (see Figure 7.1d) or fisheye [Singh and Balakrishnan, 2004] (i.e., distortion of a point of view).

7.2.2 Off-Screen Points of Interest

As explained in the introduction of this chapter, we conducted reviews on off-screen POIs techniques in two dedicated master theses [Cauz, 2019; Hincq, 2022]. However, in both cases, the process was not systematic, making the results incomplete, especially as this was not the focus of either thesis. Cockburn et al. [2009] proposes a review of overview+detail, zooming, and focus+context interfaces but does not compare them and, given the publication date, is not up-to-date with recent works. Moreover, they do not propose answers to the research questions 2.2 and 2.3. The paper of Burigat and Chittaro [2011a] is in the same case. Cheliotis et al. [2023] conducted a systematic review with a focus on the development environment (i.e., hardware, software, and methodologies) in AR navigation. Unfortunately, they do not list the visual techniques discussed in the reviewed papers. Finally, Rothe et al. [2019] addresses the topic in a review for cinematic Virtual Reality (VR). They examined 56 papers centering on gaze-guiding techniques to formulate a taxonomy of these techniques, specifically emphasizing cinematic considerations rather than the user's task. As far as we know, no survey literature answers the questions posed in the introduction of this chapter.

7.3 Methodology

We conducted a SLR by following the guidelines from Kitchenham et al. [2015]. This section presents the following methodology.

7.3.1 First Set of Identified Papers

Our first aim was to conduct a keyword search to constitute an initial set of papers. Unfortunately, the most common keywords are too generic and yield excessive results from the three chosen libraries: IEEE Xplore, ACM DL, and ScienceDirect. All three are popular among researchers related to computer science domains [Kitchenham et al., 2015]. We tried a combination of these keywords as search queries: “off-screen”, “visualization”, “technique”, “objects”, “navigation”, “points of interest”, and “wayfinding”. Therefore, we compose an initial set of papers from the references of the reviews discussed in Section 7.2.2. In other words, from Cockburn et al. [2009], Burigat and Chittaro [2011a], Rothe et al. [2019], Cauz [2019], Hincq [2022], and Cheliotis et al. [2023].

7.3.2 Inclusion and Exclusion Criteria

We defined inclusion and exclusion criteria to assess the identified sources' relevance. A document is relevant if it satisfies at least one inclusion criterion and none of the exclusion criteria. These criteria were applied to the title, abstract, and, if necessary, the text itself. The inclusion criteria are defined as follows:

1. The paper focuses on novel visual idioms supporting navigation tasks to off-screen POIs.
2. The paper presents a user evaluation to compare at least two visual idioms supporting navigation tasks to off-screen POIs.

The first inclusion criterion pertains to novel idioms, i.e., the paper must focus on the idioms rather than simply using them to achieve a more general objective. The exclusion criteria are as follows:

1. Not written in English: The standard language for white literature is English.
2. Duplicated: We consider two documents as duplicate only if they present the same research published at the same venue (i.e., our search may return the same article multiple times due to cross-references in digital libraries) or if the authors clearly indicate that two papers are duplicate.
3. Not peer-reviewed: We consider only documents published in conferences or journal outlets that mandate peer-reviewing.
4. Alter real-world perception: We consider only visual idioms that do not distort the user's field of view (e.g., blur, zoom, or fisheye) or suppress real elements (e.g., X-Ray). We consider that real-world alteration requires a study in its own right to define use cases on a more general scale, which strays too far from the focus of this review.
5. Only real or virtual POIs: We consider only visual idioms suitable for real and virtual POIs. The goal is to establish a consistent set of visual idioms that can be applied to most scenarios.
6. Not for Augmented Reality: We consider only visual idioms suitable for AR on Head-Mounted Display (HMD) as it is the focus of this research.

7.3.3 Snowballing

We conducted a phase of reverse and forward snowballing on the relevant papers selected from the initial set. For the forward method, we obtained the citations on Google Scholar. We do not search on Scopus as Google Scholar is usually more comprehensive. In addition, there is a substantial overlap between the two.

7.3.4 Coverage Assessment

As explained in section 7.3.1, a keyword search returns too many irrelevant results. It was, therefore, not possible to adopt the same approach as we had used in chapter 6; Thus, we turned to two other approaches. The first is to conduct a final snowballing phase on the references of the included papers. We did not apply this phase to citations, as we observed that it also tends to return many irrelevant results since many papers cite the included papers just because they used it without focusing on it. In addition, 52% and 19% of the included papers due to the forward snowballing phase come from the citations of Baudisch and Rosenholtz [2003] and Burigat et al.

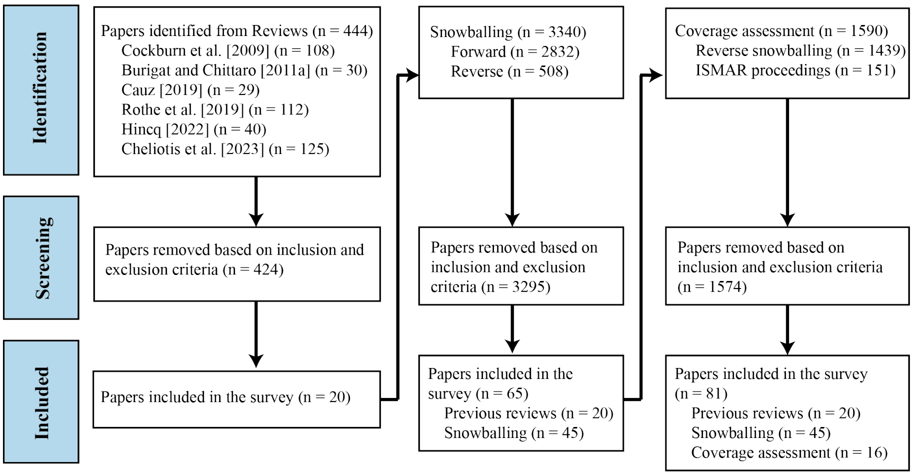


Figure 7.2: Summary of the collection process. See Table 7.1 for more details on the coverage assessment.

Steps	Total	Rejected	Included
ISMAR	151	148	3
Snowballing Round 1	1102	1091	11
Snowballing Round 2	248	246	2
Snowballing Round 3	89	89	0
Total	**1590**	**1574**	**16**

Table 7.1: Details on the coverage assessment.

[2006], respectively. This means that they are both important to the field and therefore cited by a majority of works that followed. Secondly, we applied a search to the proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) conference, one of the most important conferences in the field of Mixed Reality (MR), where several included papers were published. The search query was: *navigation OR off-screen OR wayfinding.*

7.3.5 Conducting the Review

The process is illustrated in Figure 7.2. The reverse snowballing process conducted on the six reviews resulted in the inclusion of 20 papers out of the 444 identified. Then, we applied reverse and forward snowballing on these included papers. This phase returned 508 (reverse) and 2,832 (forward) papers for a total of 3,340 papers. We applied the inclusion and exclusion criteria and included 45 additional papers. Finally, we conducted a coverage check in two steps. First, we checked the ISMAR conference proceedings that resulted in the inclusion of 3 papers out of the 151 papers identified. Second, we conducted a recursive reverse snowballing on the 45 papers included from the last snowballing phase. We obtained 1439 papers for 16 papers included. Table 7.1 details the coverage assessment. The final set included in the survey contains 81 papers published before the end of 2022.

7.4 Visual Idioms

This section presents the visual idioms determined within the included papers. Some of them work for both on-screen and off-screen POIs. We have grouped these idioms into nine categories: (1) path-based, (2) map-based, (3) border-based, (4) geometric-based, (5) arrow-based, (6) light-based, (7) picture-based, (8) text-based, and (9) guide-based. We have defined these categories based on each major approach's distinctive features. In other words, although it is not represented here, in practice, an idiom can belong to more than one category. For the names of the idioms, we have tried to keep those given by the authors of the included papers as far as possible. Unfortunately, some names were shared between papers to describe different idioms, while others were duplicates of existing approaches. Moreover, some variant approaches were proposed but did not touch on the visual aspect that interests us here. Depending on the change, these variants were either not included (if the paper focuses on other purposes) or were associated with the basic idiom. Tables 7.2 and 7.3 summarize the idioms discussed in each included paper.

Tables 7.4 and 7.5 summarize these idioms' properties. The *family* column indicates which of the three major families the idiom belongs to. The *2D/3D* column indicates whether the idiom is drawn on a surface (2D) or is modeled in 3D. The *on-screen POIs* and *multiple POIs* columns indicate whether the idiom works for on-screen and multiple POIs, respectively. The *overlap* column indicates whether some solutions to mitigate the overlap of POIs representations exist on at least one dimension. This column only makes sense if the idiom works for several POIs at once. Last, the *anchor* column indicates which anchors can be used with the idiom. A screen anchor implies that the visual elements are attached to the user's field of view (i.e., are always in the field of view like drawn on the screen). A border anchor draws the visual elements on the border of the user's field of view (i.e., a screen anchor with restricted space). A body anchor fixes the visual elements to a part of the user's body. A world anchor places the visual elements at a specific location.

Tables 7.6 and 7.7 summarize for each idiom the parameters free to play to convey additional information, such as the distance to or the type of POIs. These properties apply to the representation of POIs or the idiom in general. These are the size (i.e., free on at least one dimension), shape, border (i.e., only for lines, not surfaces), color, opacity, icon, and text. The latter two indicate whether an icon or a text can be associated with the element.

7.4.1 Path-based Idioms

This category includes visual idioms that use a thread-like path to guide the user from their starting position to the desired position. Intermediate points are often used at each intersection to indicate forks.

Ariadne's thread

The most iconic representation in this category is the **Ariadne's thread** (Figure 7.3a), named after the Greek myth of Theseus and the Minotaur. The thread is represented by a line, usually at ground or bust height, to which designers can draw icons and texts. For instance, Kraus et al. [2020] used this idiom to indicate where the user has

Paper	Idioms
Teodosio and Mills [1993]	Picture
Stoakley et al. [1995]	World-in-Miniature
Elvins et al. [1997]	Wordlets, Picture, Text
Suomela and Lehikoinen [2000]	Horizontal compass
Agrawala and Stolte [2001]	LineDrive
Chewar and McCrickard [2002]	2D map, 2D arrows, Text
Bell et al. [2002]	World-in-Miniature
Chittaro et al. [2003]	Guide
Zellweger et al. [2003]	CityLight
Baudisch and Rosenholtz [2003]	Halo, 2D arrows
Chittaro and Burigat [2004]	2D radar, 2D compass, 3D compass
Kolbe [2004]	2D arrows
Chittaro et al. [2005]	World-in-Miniature
Narzt et al. [2005]	Ariadne's thread
Khan et al. [2005]	SearchLight
Tonnis et al. [2005]	2D radar, 3D compass
Schwerdtfeger et al. [2006]	Ariadne's thread, 2D map, Text
Biocca et al. [2006]	Attention Funnel
Tonnis and Klinker [2006]	2D radar, 3D compass
Burigat et al. [2006]	2D arrows, Halo
Biocca et al. [2007]	Attention Funnel
Burigat and Chittaro [2007]	2D compass, 3D compass, 2D radar
Gustafson and Irani [2007]	EdgeRadar, Halo
Miyashita et al. [2008]	Guide
Gustafson et al. [2008]	Halo, Wedge
Buchmann et al. [2008]	Horizontal compass, 2D arrows, 2D compass
Schwerdtfeger and Klinker [2008]	Ariadne's thread, Attention Funnel
Oulasvirta et al. [2008]	2D map, World-in-Miniature
Trapp et al. [2009]	Halo, Halodot, 3D Halo
Schinke et al. [2010]	2D radar, 3D compass
Henze et al. [2010]	Halo, 2D arrows
Henze and Boll [2010]	Halo, 2D arrows
Gay-Bellile et al. [2010]	Ariadne's thread
Trapp et al. [2011]	Halo, 3D Halo, SWAVE
Jo et al. [2011]	2D radar, Aroundplot, 3D compass
Burigat and Chittaro [2011b]	2D map, Wedge, 2D arrows
Gonçalves et al. [2011a]	Halo, Halodot
Schwerdtfeger et al. [2011]	Attention Funnel, Ariadne's thread
Gonçalves et al. [2011a]	Halodot
Ghani et al. [2011]	Dynamic Insets
Dünser et al. [2012]	2D map, 2D radar
Burigat et al. [2012]	2D map, Wedge
Gonçalves et al. [2013]	2D map, Halodot, 2D arrows

Table 7.2: Visual idioms discussed in each included paper.

Paper	Idioms
Funk et al. [2014]	2D map, 2D arrows, Picture
Watanabe et al. [2015]	2D map, 2D map + Picture
Ward et al. [2016]	2D arrow, Moving Window
Miau and Feiner [2016a]	Wedge, 2D compass
Miau and Feiner [2016b]	Wedge, 2D compass
Carmo et al. [2016]	Edge radar, Edge radar + 2D map, Edge radar + 2D map + 2D Radar
Lin et al. [2017]	2D arrows, Picture
Gruenefeld et al. [2017a]	Halo, Wedge, 2D arrows
Gruenefeld et al. [2017b]	EyeSee360, Halo, 2D arrows, Wedge
Renner and Pfeiffer [2017a]	Attention Funnel, SWAVE, 2D arrows, 2D arrows + Flashing
Renner and Pfeiffer [2017b]	Ariadne's thread, 2D arrows
Renner and Pfeiffer [2017c]	2D arrows, SWAVE
Gruenefeld et al. [2018a]	HaloMR, WedgeMR
Gruenefeld et al. [2018b]	EyeSee360, 3D arrows
Bork et al. [2018]	2D radar, 3D radar, Edge radar, Aroundplot, EyeSee360, 3D compass
Renner et al. [2018]	Ariadne's thread, SWAVE
Bork et al. [2019]	Bird's eye
Perea et al. [2019]	Aroundplot, Halo, 2D arrows
Petford et al. [2019]	2D radar, Circle, Wedge, Flashing
Gruenefeld et al. [2019a]	2D radar, EyeSee360
Gruenefeld et al. [2019b]	EyeSee360
Kraus et al. [2020]	2D map
Carmo et al. [2020]	2D map, 2D radar, Edge radar
Yu et al. [2020]	2D radar, 3D radar, 3D wegde, 3D compass
Wallgrun et al. [2020]	2D radar, 2D arrows, Guide
Arntz et al. [2020]	3D arrows, Ariadne's thread
Renner and Pfeiffer [2020]	Ariadne's thread, 2D map
Osmers and Prilla [2020]	2D map, 2D compass
Hein et al. [2020]	Attention Funnel, SWAVE, 2D arrows
Tang and Zhou [2020]	Ariadne's thread, Ariadne's thread + guide, Ariadne's thread + guide + 2D map, Ariadne's thread + 2D map
Rovira et al. [2020]	2D map with and without 2D arrows and picture
Ye et al. [2021]	Guide
Hu et al. [2021]	Circle, 3D arrows, Border light
Harada and Ohyama [2021]	3D radar, Moving window, Radiation, 3D compass, Spherical Gradation
Chung et al. [2021]	Picture, 2D radar, 3D arrows
Evangelista et al. [2021]	Horizontal compass
Biswas et al. [2022]	Halo, 2D arrows, 3D arrows
Wieland et al. [2022]	Ariadne's thread, 3D halo, 3D compass

Table 7.3: Visual idioms discussed in each included paper (Continued).

Idiom's name	**Type**	**2D/3D**	**On-Screen POIs**	**Multiple POIs**	**Overlap**	**Anchor**
Path-based						
Ariadne's thread	Aid	3D	Yes	Yes	Managed	World
Attention Funnel	Aid	3D	Yes	No	/	World
Map-based						
Bird's eye	Overview+detail	2D	Yes	Yes	Managed	Screen/Body/World
2D Map	Overview+detail	2D	Yes	Yes	Managed	Screen/Body/World
Dynamic Insets	Overview+detail	2D	No	Yes	Unmanaged	Screen
LineDrive	Overview+detail	2D	Yes	Yes	Managed	Screen/Body/World
2D Radar	Overview+detail	2D	Yes	Yes	Managed	Screen/Body/World
3D Radar	Overview+detail	3D	Yes	Yes	Managed	Body/World
World-in-Miniature	Overview+detail	3D	Yes	Yes	Managed	Body/World
Wordlets	Overview+detail	3D	Yes	Yes	Managed	Screen/Body/World
Border-based						
City Lights	Focus+Context	2D	No	Yes	Unmanaged	Screen
Edge Radar	Focus+Context	2D	No	Yes	Managed	Screen
Aroundplot	Focus+Context	2D	No	Yes	Managed	Screen
EyeSee360	Focus+Context	2D	No	Yes	Managed	Screen
Horizontal Compass	Focus+Context	2D	Yes	Yes	Unmanaged	Screen
Arrow-based						
2D Arrows	Aid	2D	Yes	Yes	Managed	Screen
2D Compass	Aid	2D	Yes	Yes	Managed	Screen
3D Arrows	Aid	3D	Yes	Yes	Managed	Screen/Body/World
3D Compass	Aid	3D	Yes	Yes	Managed	Screen/Body/World

Table 7.4: Visual Idioms supporting navigation tasks for off-screen Points of Interest (POIs).

Idiom's name	**Type**	**2D/3D**	**On-Screen POIs**	**Multiple POIs**	**Overlap**	**Anchor**
Geometric-based						
Halo	Aid	2D	Yes	Yes	Managed	Screen
Halodot	Aid	2D	Yes	Yes	Managed	Screen
HaloMR	Aid	2D	Yes	Yes	Managed	Screen
3D Halo	Aid	3D	Yes	Yes	Managed	World
Wedge	Aid	2D	No	Yes	Managed	Screen
3D Wedge	Aid	3D	Yes	Yes	Managed	Body
WedgeMR	Aid	2D	No	Yes	Managed	Screen
SWAVE	Aid	3D	Yes	No	/	Body
Moving window	Aid	2D	Yes	Yes	Unmanaged	Screen
Circle	Aid	2D	Yes	Yes	Unmanaged	Screen
Radiation	Aid	3D	Yes	Yes	Managed	World
Light-based						
SearchLight	Aid	3D	Yes	Yes	Unmanaged	World
Spherical Gradation	Aid	3D	Yes	Yes	Unmanaged	World
Flashing	Aid	2D	Yes	Yes	Unmanaged	Screen
Border light	Aid	2D	Yes	Yes	Unmanaged	Screen
Picture-based						
Picture	Overview+detail	2D	No	Yes	Unmanaged	Screen
Text-based						
Text	Aid	2D	Yes	Yes	Yes	Screen
Guide-based						
Guide	Aid	3D	Yes	Yes	Yes	World

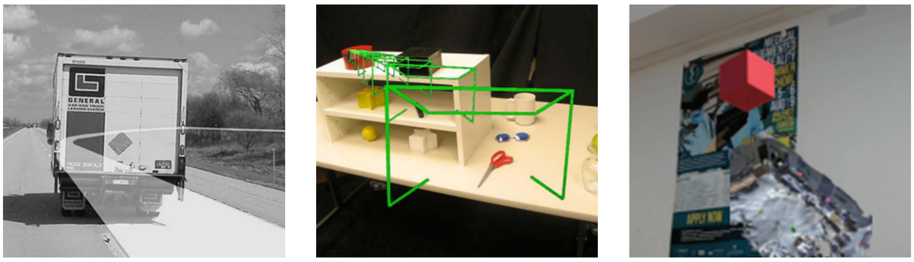
Table 7.5: Visual Idioms supporting navigation tasks for off-screen Points of Interest (POIs) (Continued).

Idiom's name	**Size**	**Shape**	**Border**	**Color**	**Opacity**	**Icon**	**Text**
Path-based							
Ariadne's thread	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attention Funnel	Yes	Yes	Yes	Yes	Yes	No	No
Map-based							
Bird's eye	Yes	Yes	No	Yes	Yes	Yes	Yes
2D Map	Yes	Yes	No	Yes	Yes	Yes	Yes
Dynamic Insets	Yes	Yes	No	Yes	Yes	Yes	Yes
LineDrive	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2D Radar	Yes	Yes	No	Yes	Yes	Yes	Yes
3D Radar	Yes	Yes	No	Yes	Yes	Yes	Yes
World-in-Miniature	Yes	Yes	No	Yes	Yes	Yes	Yes
Wordlets	Yes	Yes	No	Yes	Yes	Yes	Yes
Border-based							
City Lights	No	No	No	Yes	Yes	No	No
Edge Radar	Yes	Yes	No	Yes	Yes	Yes	Yes
Aroundplot	Yes	Yes	No	Yes	Yes	Yes	Yes
EyeSee360	Yes	Yes	No	Yes	Yes	Yes	Yes
Horizontal Compass	Yes	Yes	No	Yes	Yes	Yes	Yes
Arrow-based							
2D Arrows	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2D Compass	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3D Arrows	Yes	Yes	Yes	Yes	Yes	No	No
3D Compass	Yes	Yes	Yes	Yes	Yes	No	No

Table 7.6: Parameters available for visual Idioms supporting navigation tasks for off-screen Points of Interest (POIs).

Idiom's name	**Size**	**Shape**	**Border**	**Color**	**Opacity**	**Icon**	**Text**
Geometric-based							
Halo	No	No	Yes	Yes	Yes	No	No
Halodot	No	No	Yes	Yes	Yes	No	No
HaloMR	No	No	Yes	Yes	Yes	No	No
3D Halo	No	No	Yes	Yes	Yes	No	No
Wedge	No	No	Yes	Yes	Yes	No	No
3D Wedge	No	No	Yes	Yes	Yes	No	No
WedgeMR	No	No	Yes	Yes	Yes	No	No
SWAVE	No	No	Yes	Yes	Yes	No	No
Moving window	Yes	Yes	Yes	Yes	Yes	No	No
Circle	Yes	Yes	Yes	Yes	Yes	No	No
Radiation	No	No	Yes	Yes	Yes	No	No
Light-based							
SearchLight	No	No	No	Yes	Yes	No	No
Spherical Gradation	No	No	No	Yes	Yes	No	No
Flashing	Yes	No	No	Yes	Yes	No	No
Border light	Yes	No	No	Yes	Yes	No	No
Picture-based							
Picture	Yes	Yes	No	Yes	Yes	Yes	Yes
Text-based							
Text	Yes	No	No	Yes	Yes	Yes	Yes
Guide-based							
Guide	Yes	Yes	No	Yes	Yes	Yes	Yes

Table 7.7: Parameters available for visual Idioms supporting navigation tasks for off-screen Points of Interest (POIs) (Continued).



(a) Ariadne's thread
(Narzt et al. [2005])

(b) Attention Funnel
(Biocca et al. [2006])

(c) Bird's eye
(Bork et al. [2019])

(d) 2D map
(Carmo et al. [2020])

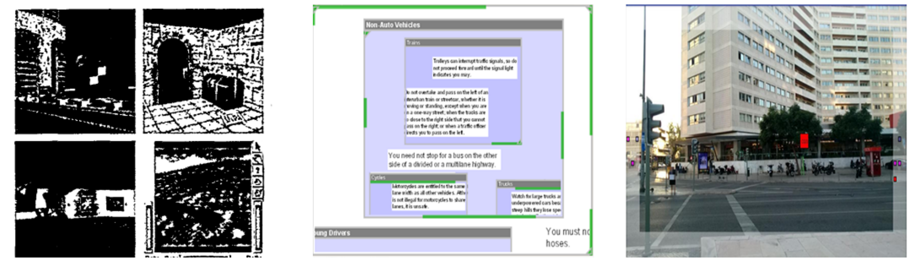
(e) Dynamic Insets
(Ghani et al. [2011])

(f) LineDrive
(Agrawala and Stolte [2001])

(g) 2D radar
(Bork et al. [2018])

(h) 3D radar
(Bork et al. [2018])

(i) World-in-Miniature
(Stoakley et al. [1995])

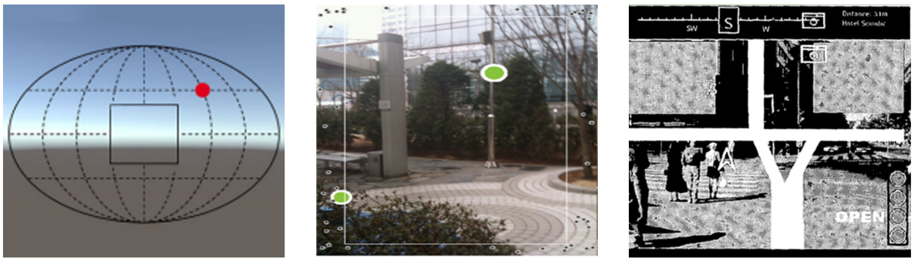


(j) Wordlets
(Elvins et al. [1997])

(k) City Light
(Zellweger et al. [2003])

(l) Edge Radar
(Carmo et al. [2020])

Figure 7.3: Illustrations of visual idioms: (a-b) path-based, (c-j) map-based, and (k-l) border-based. The source is given for each picture.



(a) EyeSee360 (Gruenefeld et al. [2019a])

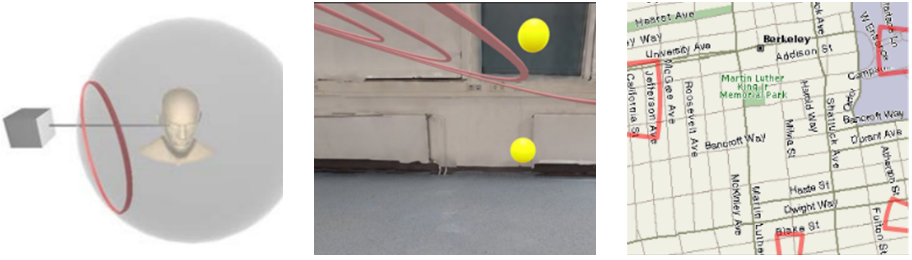
(b) Aroundplot (Jo et al. [2011])

(c) Horizontal compass (Suomela and Lehikoinen [2000])

(d) Circle (Hu et al. [2021])

(e) Halo (Baudisch and Rosenholtz [2003])

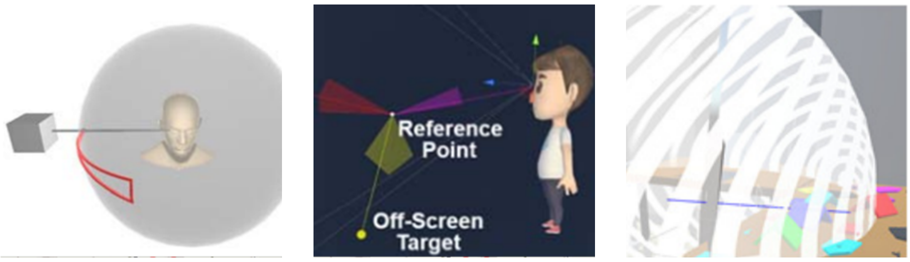
(f) HaloDot (Gonçalves et al. [2011b])



(g) HaloMR (Gruenefeld et al. [2018a])

(h) 3D Halo (Wieland et al. [2022])

(i) Wedge (Baudisch and Rosenholtz [2003])

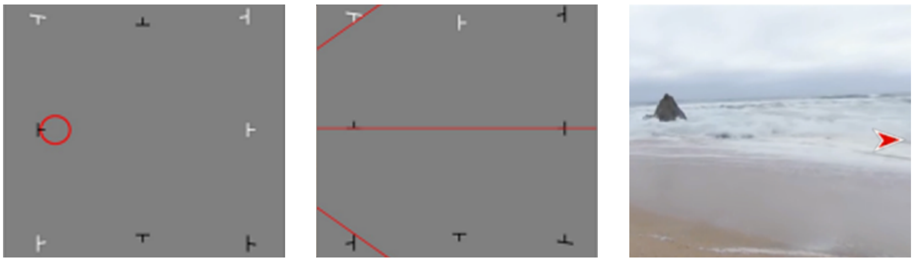


(j) 3D Wedge (Yu et al. [2020])

(k) Wedge MR (Gruenefeld et al. [2018a])

(l) SWAVE (Renner and Pfeiffer [2017a])

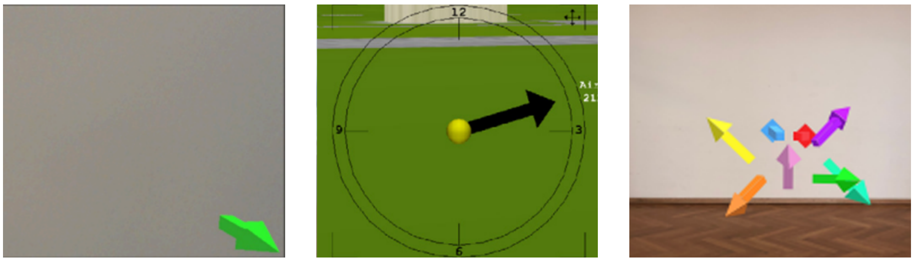
Figure 7.4: Illustrations of visual idioms (Continued): (a-d) border-based, (e-l) geometric-based. The source is given for each picture.



(a) Moving window
(Harada and Ohyama [2021])

(b) Radiation
(Harada and Ohyama [2021])

(c) 2D arrows
(Lin et al. [2017])



(d) 3D arrows
(Gruenefeld et al. [2018b])

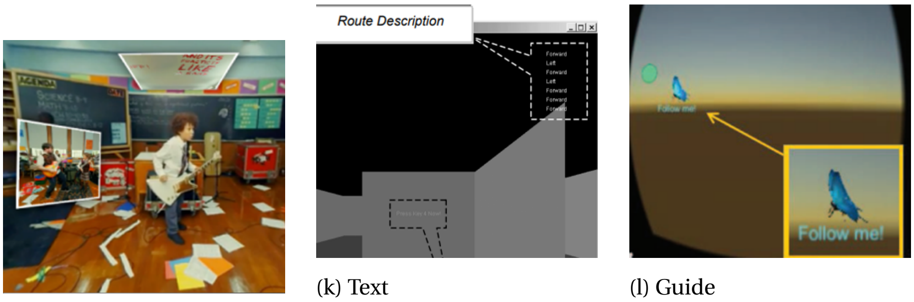
(e) 2D compass
(Burigat and Chittaro [2007])

(f) 3D compass
(Bork et al. [2018])

(g) SearchLight
(Khan et al. [2005])

(h) Spherical Gradation
(Harada and Ohyama [2021])

(i) Border light
(Hu et al. [2021])



(j) Picture
(Lin et al. [2017])

(k) Text
(Chewar and McCrickard [2002])

(l) Guide
(Wallgrun et al. [2020])

Figure 7.5: Illustrations of visual idioms (Continued): (a-b) geometric-based, (c-f) arrow-based, (g-i) light-based, (j) picture-based, (k) text-based, and (l) guide-based. The source is given for each picture.

passed, with arrows drawn on the line to indicate the direction. The lines can be placed side by side to avoid overlapping.

Papers presenting or evaluating this idiom: Narzt et al. [2005], Schwerdtfeger et al. [2006]; Schwerdtfeger and Klinker [2008]; Schwerdtfeger et al. [2011], Gay-Bellile et al. [2010], Renner et al. [2018]; Renner and Pfeiffer [2017b, 2020], Arntz et al. [2020], Tang and Zhou [2020], and Wieland et al. [2022].

Attention Funnel

Biocca et al. [2006] introduced the **Attention Funnel** (Figure 7.3b) that repeats a geometric shape along the thread with a short distance between each. The funnel is placed at eye level and takes a large part of the user's field of view to force turning the gaze. So, in contrast to Ariadne's thread, it cannot address multiple POIs simultaneously, and it is not designed to indicate an icon or a text.

Papers presenting or evaluating this idiom: Biocca et al. [2007, 2006], Schwerdtfeger and Klinker [2008]; Schwerdtfeger et al. [2011], Renner and Pfeiffer [2017a], and Hein et al. [2020].

7.4.2 Map-based Idioms

This category contains visual idioms based on a top-down view of the user's environment. The representation of POIs consists of annotations on another element giving the context. Therefore, designers can play freely with the size, shape, color, and opacity, as well as add icons and text.

Bird's eye

Bork et al. [2019] introduced **Bird's eye** (Figure 7.3c), a real-time video from a bird's-eye view of the user. This point of view can be captured by an overhead camera or by reconstructing the environment using various sensors. Although they named their approach Bird's eye, Tonnis and Klinker [2006]; Tonnis et al. [2005] proposed something more in accordance with the 2D radar approach detailed below.

Papers presenting or evaluating this idiom: Bork et al. [2019].

2D Map

A **2D map** (Figure 7.3d) represents a static picture of the user's surroundings. It is one of the best-known idioms, as it is widely used in map and GPS applications. The background with the environment details can be either an image taken from top or a drawing representing the main elements. Levinew et al. [1984] and Darken and Cevik [1999] recommended aligning the map orientation with the environment and showing where the user is.

Ghani et al. [2011], rather than displaying a single map, displays a series of minimaps on the edges of the user's field of view. These are positioned on the edge closest to the POIs they are concerned with and on which they are centered. They named their approach **Dynamic Insets** (Figure 7.3e).

Agrawala and Stolte [2001] suggested a variant (Figure 7.3f), namely **LineDrive**, where details are removed to show only annotated paths. in this version, designers can also play on the path's border to differentiate POIs.

Papers presenting or evaluating 2D map: Chewar and McCrickard [2002], Schwerdtfeger et al. [2006], Oulasvirta et al. [2008], Burigat and Chittaro [2011b]; Burigat et al. [2012], Dünser et al. [2012], Gonçalves et al. [2013], Funk et al. [2014], Watanabe et al. [2015], Carmo et al. [2016, 2020], Renner and Pfeiffer [2020], Osmers and Prilla [2020], Tang and Zhou [2020], Rovira et al. [2020], and Kraus et al. [2020].
Papers presenting or evaluating Dynamic Insets: Ghani et al. [2011].
Papers presenting or evaluating LineDrive: Agrawala and Stolte [2001].

2D Radar

A **2D radar** (Figure 7.3g) indicates only POIs around the user, unlike a 2D map that details the environment. Its most common form is nested circles centered on the user, but it can be a grid as with [Schinke et al., 2010]. These circles help to understand the distances between POIs. Moreover, Gruenefeld et al. [2019a] proposed moving the symbol representing a POI vertically to indicate its height and then drawing a vertical line to the circle. This prevents the vertical overlapping generated from the top-down view. For the same reasons, Bork et al. [2018] added a vertical arrow starting from the symbol representing a POI to the adequate height.

Papers presenting or evaluating this idiom: Chittaro and Burigat [2004], Tonnis and Klinker [2006]; Tonnis et al. [2005], Burigat and Chittaro [2007], Schinke et al. [2010], Jo et al. [2011], Dünser et al. [2012], Carmo et al. [2016], Bork et al. [2018], Petford et al. [2019], Gruenefeld et al. [2019a], Carmo et al. [2020], Yu et al. [2020], Wallgrun et al. [2020], and Chung et al. [2021].

3D Radar

Bork et al. [2018] introduced the **3D radar** (Figure 7.3h) based on spheres rather than circles like in 2D radar. Yu et al. [2020] suggested using nested spheres with more and more transparency further they are from the center to represent distance better.

Papers presenting or evaluating this idiom: Bork et al. [2018], Yu et al. [2020], and Harada and Ohyama [2021].

3D map

Stoakley et al. [1995] presented the **WiM** (Figure 7.3i), a miniature 3D version of the environment. Elements of interest or elements that contribute to the mental representation of the site are represented. Nevertheless, this does not prevent the addition of annotations to direct the user's gaze [Bell et al., 2002]. Danyluk et al. [2021] proposed a taxonomy of WiM based on height dimensions: size, scope, abstraction, geometry, reference frame, links, multiples, virtuality. Later, Elvins et al. [1997] proposed the **Wordlets** (Figure 7.3j), 3D thumbnails of each POI. Where the WiM is in one element, wordlets are a set of WiM reduced to the environment close to each POI.

Papers presenting or evaluating WiM: Stoakley et al. [1995], Bell et al. [2002], Chittaro et al. [2005], and Oulasvirta et al. [2008].
Papers presenting or evaluating Wordlets: Elvins et al. [1997].

7.4.3 Border-based Idioms

This category contains visual idioms located on the border of the user's field of view. The most significant variability between approaches concerns the space allocated on the edge and how they convert 3D coordinates in the environment to 2D coordinates on the screen. They do not support on-screen POIs except for the horizontal compass but manage multiple off-screen POIs.

City Lights

Zellweger et al. [2003] introduced **City Lights** (Figure 7.3k) that draws a thick line on the border of the user's field of view to indicate the direction where to look. Although minimalist regarding space, this approach is extremely limited in overlaps management. In addition, designers can only play on the color and opacity to convey additional information.

Papers presenting or evaluating this idiom: Zellweger et al. [2003].

Edge Radar

Edge radar (Figure 7.3l) defines a scaled space along the border where placing the representation of each POIs respecting the distance between them. Moreover, Jo et al. [2011] proposed a variant with **Aroundplot** (Figure 7.4b) using a fisheye projection. Projection is based on the position of the POI on the surface of a sphere centered on the pivot point of the user's field of view. In addition, the proposition comes with a magnification mechanism to dynamically extend the space allocated in the direction of the view movement.

Papers presenting or evaluating Edge Radar: Gustafson and Irani [2007], Bork et al. [2018], and Carmo et al. [2016, 2020].
Papers presenting or evaluating Aroundplot: Jo et al. [2011], Perea et al. [2019] (without the spherical projection), and Bork et al. [2018].

EyeSee360

Gruenefeld et al. [2017b] introduced **EyeSee360** (Figure 7.4a) inspired from Edge radar. An ellipse and a rectangle are drawn on the screen. The inner rectangle denotes the field of view and is designed not to block the user's focus. The outer ellipse extends the field of view to 360°. 45° helplines provide information on the rotation needed to access each POIs.

Papers presenting or evaluating this idiom: Gruenefeld et al. [2017b, 2019a, 2018b, 2019b], and Bork et al. [2018].

Horizontal compass

Suomela and Lehikoinen [2000] presented the **horizontal compass** (Figure 7.4c) that indicates the elements that need a left or right rotation of the user's head to be observed. It is placed at the top of the user's field of view. In addition, the cardinal directions are indicated, but it was sometimes given up, like in the version of Buchmann et al. [2008].

Papers presenting or evaluating this idiom: Suomela and Lehikoinen [2000], Buchmann et al. [2008], and Evangelista et al. [2021].

Circle

Petford et al. [2019] and Hu et al. [2021] placed a circle on the edge of the user view. We name this approach **Circle** (Figure 7.4d), although it is possible to use other representation elements.

Papers presenting or evaluating this idiom: Petford et al. [2019], and Hu et al. [2021].

7.4.4 Geometry-based Idioms

This category contains visual idioms based on geometric shapes drawn on the user's field of view or placed in the world. Since the technique depends on the shape used and aims to make it appear in the user's vision field, only the border, color, and opacity are free to design. Moreover, although some approaches add a number to manage overlap, they are not intended to associate an icon or text with it.

Halo-like

Baudisch and Rosenholtz [2003] introduced **Halo** (Figure 7.4e) that draws a circle centered on the POI, with an arc visible in the user's field of view. Several strategies have been proposed to solve the overlap issue. First, Baudisch and Rosenholtz [2003] suggested aggregating circles and rendering 2-3 thinner concentric arcs centered at their average location. Second, Perea et al. [2019] proposed to aggregate the circles and annotate with a number the circle centered at the barycentre. Third, Perea et al. [2019] suggested joining the arcs to form a region along the edge of the user's field of view. Moreover, a first variant exists with **Halodot** (Figure 7.4f) proposed by Gonçalves et al. [2011a,b]. It is almost the same thing, except that a circle or arrow pointing to the border of the user's field of view is added to the arc of the circle to reinforce the POI's direction. Concerning the overlapping issue, they adopt the second strategy described above. A second variant is **HaloMR** (Figure 7.4g) presented by Gruenefeld et al. [2018a]. It draws a circle on the inner surface of a sphere centered on the projection of the target POI; thus, we consider it a 2D approach. A last variant is **3D Halo** (Figure 7.4h) introduced by Trapp et al. [2009] and Wieland et al. [2022]. 3D circles are placed directly in the environment and enlarged to fit an arc in the user's field of view. Trapp et al. [2011] proposed to combine several circles on different planes to reinforce direction.

Papers presenting or evaluating Halo: Baudisch and Rosenholtz [2003], Burigat et al. [2006], Gustafson et al. [2008]; Gustafson and Irani [2007], Trapp et al. [2009, 2011], Henze and Boll [2010]; Henze et al. [2010], Gonçalves et al. [2011a], Gruenefeld et al. [2017a,b], Perea et al. [2019], and Biswas et al. [2022].
Papers presenting or evaluating HaloDot: Trapp et al. [2009], and Gonçalves et al. [2011a,b]; Gonçalves et al. [2013].
Papers presenting or evaluating HaloMR: Gruenefeld et al. [2018a].
Papers presenting or evaluating 3D Halo: Trapp et al. [2009, 2011] and Wieland et al. [2022].

Wedge-like

Gustafson et al. [2008] presented **Wedge** (Figure 7.4i), a Halo-like idiom based on triangles rather than circles. Designers can modify each triangle's rotation, intrusion, and aperture to prevent overlapping. Yu et al. [2020] proposed a variant, namely **3D Wedge** (Figure 7.4j), with 3D pyramids starting from a point in front of the user, thus forming a cluster. For the pyramids directed in the same direction and of the same length, they suggest modifying the width of the base to stack them. Gruenefeld et al. [2018a] proposed another variant, namely **WedgeMR** (Figure 7.4k), with the initial version of Wedge drawn on the surface of a sphere centered on the user; thus, we consider it a 2D approach. However, they must remove the ability to make space because *"The jumping of wedges would happen because of the original approach of making space for other wedges because here, they are applied onto the inner surface of a sphere, which is a double-curved geometry. Making space for other wedges combined with head-movement requires wedges to jump."*.

Papers presenting or evaluating Wedge: Gustafson et al. [2008], Burigat and Chittaro [2011b]; Burigat et al. [2012], Miau and Feiner [2016a,b], Gruenefeld et al. [2017a,b], and Petford et al. [2019].
Papers presenting or evaluating 3D Wedge: Yu et al. [2020].
Papers presenting or evaluating WedgeMR: Gruenefeld et al. [2018a].

Spherical Wave-Based Guidance

Renner and Pfeiffer [2017a] introduced **Spherical Wave-Based Guidance (SWAVE)** (Figure 7.4l) based on a sphere with often an animation of waves on the surface. The sphere is centered on the user's head and has a radius equivalent to the distance between the user's head and the target POI.

Papers presenting or evaluating this idiom: Trapp et al. [2011], Renner et al. [2018]; Renner and Pfeiffer [2017a,c], and Hein et al. [2020].

Moving Window

Harada and Ohyama [2021] presented **Moving window** (Figure 7.5a) based on a circle moving from the central position to the target direction. It is unclear if the circle can leave the user's field of view or stop at the border.

Papers presenting or evaluating this idiom: Ward et al. [2016], Harada and Ohyama [2021].

Radiation

Harada and Ohyama [2021] proposed **Radiation** (Figure 7.5b) composed of lines focused on the target direction. The approach is, therefore, located in the world, but it is necessary to create enough lines to ensure that the user always has at least one in his field of view.

Papers presenting or evaluating this idiom: Harada and Ohyama [2021].

7.4.5 Arrow-based Idioms

This category contains visual idioms based on arrows. Each technique works for on-screen POIs. The size, shape, border, color, and opacity are free. **2D arrows** is printed on the user's field of vision, usually placed on the edge closest to the POI concerned (Figure 7.5c). Two versions, in particular, are highlighted: one that plays on the size of the arrow and one that plays on the length of the arrow to indicate distance. However, when the arrows are placed in a cluster starting from the same point, the name used is **2D compass** (Figure 7.5e). Each has a 3D version called **3D arrows** (Figure 7.5d) and **3D compass** (Figure 7.5f) respectively.

Papers presenting or evaluating 2D arrows: Chewar and McCrickard [2002], Baudisch and Rosenholtz [2003], Kolbe [2004], Burigat et al. [2006], Buchmann et al. [2008], Henze and Boll [2010]; Henze et al. [2010], Burigat and Chittaro [2011b], Gonçalves et al. [2013], Funk et al. [2014], Lin et al. [2017], Gruenefeld et al. [2017a,b], Renner and Pfeiffer [2017a,b,c], Perea et al. [2019], Hein et al. [2020], Wallgrun et al. [2020], Rovira et al. [2020], and Biswas et al. [2022].
Papers presenting or evaluating 3D arrows: Gruenefeld et al. [2018b], Arntz et al. [2020], Hu et al. [2021], Chung et al. [2021], and Biswas et al. [2022].
Papers presenting or evaluating 2D compass: Chittaro and Burigat [2004], Burigat and Chittaro [2007], Buchmann et al. [2008], Miau and Feiner [2016a,b], and Osmers and Prilla [2020].
Papers presenting or evaluating 3D compass: Chittaro and Burigat [2004], Tonnis and Klinker [2006]; Tonnis et al. [2005], Burigat and Chittaro [2007], Schinke et al. [2010], Jo et al. [2011], Bork et al. [2018], Yu et al. [2020], Harada and Ohyama [2021], and Wieland et al. [2022].

7.4.6 Light-based Idioms

This category contains visual idioms based on light effects. Therefore, designers are free to play only on color and opacity to convey additional information. Khan et al. [2005] introduced **Searchlight** (Figure 7.5g) that simulates a spot effect on POIs starting from the user's field of view. Harada and Ohyama [2021] proposed **Spherical gradation** (Figure 7.5h) that generates light in all directions from the POI, fading with distance. From their side, Petford et al. [2019] presented **Flashing** that consists of a flashing light on the border of the user's field of view. Renner and Pfeiffer [2017a] adopted this approach over the entire screen to complement other visual idioms (e.g., 2D arrows). Hu et al. [2021] proposed the same approach but without flashing (Figure 7.5i). We name this latter **Border light**.

Papers presenting or evaluating SearchLight: Khan et al. [2005].

Papers presenting or evaluating Spherical Gradation: Harada and Ohyama [2021].
Papers presenting or evaluating Flashing: Renner and Pfeiffer [2017a], Petford et al. [2019].
Papers presenting or evaluating Border Light: Hu et al. [2021].

7.4.7 Picture-based Idioms

This category contains visual idioms based on pictures placed in the world but kept in the user's field of view. Funk et al. [2014] combined the last captured picture of a target object with arrows. Watanabe et al. [2015] combined pictures taken by the user with a 2D map. Lin et al. [2017] proposed an approach where pictures are placed in perspective without obstructing the center of this field (Figure 7.5j). Teodosio and Mills [1993] suggested using a panoramic picture, which are known in MR for being displayed on spheres. However, Chung et al. [2021] displayed it as a rectangle in the top of the field of view.

Papers presenting or evaluating Picture: Teodosio and Mills [1993], Funk et al. [2014], Watanabe et al. [2015], Lin et al. [2017], Rovira et al. [2020], Chung et al. [2021].

7.4.8 Text-based Idioms

This category consists of presenting textual information to the user. An example is given by Chewar and McCrickard [2002] (Figure 7.5k).

Papers presenting or evaluating Text: Chewar and McCrickard [2002], and Schwerdtfeger et al. [2006].

7.4.9 Guide-based Idioms

This category proposes a 3D model guide to the off-screen POIs. An example is given by Wallgrun et al. [2020] (Figure 7.5l).

Papers presenting or evaluating Text: Chittaro et al. [2003], Miyashita et al. [2008], Wallgrun et al. [2020], Tang and Zhou [2020], and Ye et al. [2021].

7.5 Comparison of Visual Idioms on Conventional 2D screen

This section compares the idioms presented in the previous section based on their 2D screen performance in literature evaluations. We included MR experience conducted on a smartphone or tablet. We have identified fifteen types of tasks and will discuss each separately. We aim to determine the best idiom by presenting significant study results. Note that the authors used the NASA TLX form [Human Performance Research Group, 1980] to calculate the workload. This form is composed of six subscales (mental demand, physical demand, temporal demand, performance, effort, and frustration) that allow the calculation of the participant's perceived workload during the evaluation.

7.5.1 Search Task

In this task, the participant must reach a target in the environment. Depending on the nature of the environment, the constraints are different. For a 3D virtual environment (or AR environment where participants can move), participants had to move a first-person character until they reached one or more points of interest. For a 2D (or AR environment where participants cannot move), participants had to pan on the interface or turn on themselves with the device, respectively, until they had the point(s) of interest in the center of the field of view.

Regarding 3D virtual environments, Elvins et al. [1997] conducted a study with Wordlets, Picture, and Text. They concluded that Wordlets significantly outperformed the other two except for the time spent consulting the guidebook. Chewar and McCrickard [2002] compared 2D arrows, text instruction (only one and a list), and 2D maps (partial and full). They observed significant results for 2D arrows in terms of time, accuracy, and participant preference. A second efficient idiom is the textual list of instructions regarding accuracy and distraction (i.e., time to look at the idiom rather than in the center of the field of view). On the contrary, the two 2D maps showed poor time results, worse than the condition without indicators. For their part, Chittaro and Burigat [2004] evaluated 3D compass, 2D radar, and 2D compass. They concluded that the former is at least as fast as the two others and can even be significantly faster if combined with automatic alignment of the view to the direction of the arrow when clicked. Furthermore, Burigat and Chittaro [2007] studied the same idioms as Chittaro and Burigat [2004] with inexperienced and experienced users in virtual environments. Concerning the time to complete tasks, they obtained that the 3D compass is significantly faster than the other two for inexperienced users in an air base and for both user types in an empty sphere with wireframe texture drawn on the inside face. In addition, inexperienced users were significantly faster with a 2D compass than with a 2D radar in the second environment. The same results occurred for user preference, respecting user types and environments. Next, Oulasvirta et al. [2008] determined that 2D map significantly outperformed WiM. They concluded by *“2D maps direct users into using reliable and ubiquitous cues like street names and street topology, and they better afford the use of pre-knowledge and bodily action to reduce cognitive workload. The findings point out two prominent problems of 3D in this task: the uninformativeness of the street-level perspective and the ambiguity of photorealistic cues on a small mobile display. Directionality of the viewport complicates ego-centric alignment in the 3D view, too.”*. Dünser et al. [2012] compared an AR view with a 2D radar against a 2D map on smartphone. The results showed no significance in performance but a preference of participants for the 2D map or a combination of both. Moreover, Tang and Zhou [2020] evaluated Ariadne’s thread combined with no adding visual idioms, a guide, a 2D map, and a guide with a 2D map. They observed that participants had higher navigation errors with the guide than without it, but this result is not significant. The 2D map also increased the error rate but was significant only for older adults. Last, Rovira et al. [2020] compared a 2D map with and without 2D arrows and Picture. They concluded that Picture contributed significantly to better results.

Concerning 2D environments, Schinke et al. [2010] evaluated 2D radar and 3D compass but did not obtain significant results. Henze and Boll [2010] (resp.

Gonçalves et al. [2011a]) studied Halo and 2D arrows (resp. Halodot). Their results showed that scaled and stretched arrows (resp. Halodot) were significantly faster than halo. Jo et al. [2011] compared Aroundplot, 3D compass, and 2D radar. In the first trial, they make the target appear by highlighting their representation a few seconds before resetting them to a normal state when starting to search. The results demonstrated that Aroundplot is significantly more accurate, faster, and less workload-intensive than 2D radar. However, it is faster only when there are a small number of POIs. In addition, Aroundplot is significantly more accurate than 3D compass when there are many POIs. In the second trial, they used a specific shape to indicate the target. Aroundplot and 3D compass were significantly faster than 2D radar. Aroundplot was also significantly less workload-intensive than 2D radar. For their part, Gonçalves et al. [2013] evaluated 2D map, Halodot, and 2D arrows. Their results demonstrated that a 2D map is significantly faster than Halodot, especially with numerous targets. Their participants preferred a 2D map and scaled arrows rather than Halodot. Last, Perea et al. [2019] studied Halo, 2D arrows, and Aroundplot. Halo and 2D arrows were significantly faster than Aroundplot in their first trial, with one target at a time. In addition, the former two were perceived more usable than the latter. In the second trial with multiple targets at once, they did not find significant results in time. Still, they found that their two new versions of Halo, Halo* and Halo**, were significantly less workload-intensive than Aroundplot and 2D arrows, respectively. Halo* considers that small overlaps between circles do not decrease the understanding of the environment. Halo** aggregates circles to create a region.

In conclusion of this task, we deduct the following recommendations. For 3D virtual environments, there is a preference for arrows, particularly the 3D compass, rather than a 2D map or 2D radar. Wordlets is also better than Picture and Text. For 2D environments, suggestions point more towards Halo or 2D arrows, although it is impossible to confirm this observation perfectly.

7.5.2 Locate Task

This task consists of indicating the expected location of the off-screen POI in the off-screen space. Perea et al. [2019] did not find significant results between Aroundplot, Halo, and 2D arrows. Baudisch and Rosenholtz [2003] evaluated the Halo and 2D arrows idioms. While Halo is significantly faster than 2D arrows, the latter is significantly more accurate than the former. As for the participants, they prefer Halo because it requires less mental mathematics. Burigat et al. [2006] observed that Halo is significantly more accurate than scaled and stretched arrows. Their participants also prefer Halo. Next, Gustafson et al. [2008] compared Halo and Wedge but did not find significant results. Nevertheless, Wedge tends to be better than Halo, thanks to better POI management in the corners. Oulasvirta et al. [2008] observed that 2D map outperformed WiM, but their results were discussed in search task. Next, Burigat and Chittaro [2011b] determined that Wedge and 2D maps were significantly more accurate than scaled arrows, but only for height targets, not five. Miau and Feiner [2016a] conducted many trials where they observed significant results in favor of their 2D compass than Wedge for completion time, accuracy, and preference. They concluded that Wedge is effective for close POIs, but not for far POIs, as the two sides of the triangle become too parallel. On the contrary, their compass indicates distance by the line length, which makes it easier to determine the direction and distance. Last, Wieland et al. [2022] evaluated a 3D compass, Ariadne's thread, and 3D Halo. Their results indicated that 3D halo is significantly less accurate than the other two. In addition, 3D compass is significantly less workload-intensive than 3D Halo, particularly for the effort.

In conclusion of this task, the 2D and 3D compass seem to be the better approach to adopt. Two parameters are essential for this task: direction and distance. Arrows indicate direction well but require mental effort to calculate distance. On the other hand, Halo, Wedge, and their variants are handy for knowing distance and direction, but only when the POI is close. Nevertheless, there will be a need to confirm these observations.

7.5.3 Closest Task

In this task, participants must select the closest POI of a determined position. Burigat and Chittaro [2011b] studied 2D map, Wedge, and 2D arrows, but obtained no significant results. Baudisch and Rosenholtz [2003] compared Halo to 2D arrows. Regarding time completion, Halo is significantly faster than 2D arrows for the same mathematical reason as exposed in the Locate Task. Halo is also preferred over 2D arrows. On that, Burigat et al. [2006] observed that Halo and scaled arrows are preferred over stretched arrows. Next, Gustafson et al. [2008] determined that Halo is preferred over Wedge. Last, Carmo et al. [2016] evaluated Edge radar alone and combined with a 2D map and with a 2D map and a 2D Radar on this map. They concluded that the combined three idioms imply significantly less rotation and are significantly preferred by the participants.

In conclusion of this task, Halo seems to be the best choice, but there is a need for further exploration, including other idioms.

7.5.4 Traverse Task

This task consists of determining the shortest path starting at a determined position by selecting the POIs in the correct order. Only Baudisch has studied this task, comparing Halo and 2D arrows. The results demonstrated that Halo is marginally significantly faster than 2D arrows. In addition, Halo is preferred by the participants.

7.5.5 Avoid Task

In this task, participants must select or access a POI from a determined position by avoiding danger (e.g., traffic jams). Gustafson et al. [2008] did not find significant results between Halo and Wedge, but their participants preferred Wedge to Halo. Baudisch and Rosenholtz [2003] once again found that Halo is significantly better than 2D arrows and preferred by participants.

7.5.6 Cluster task

This task involves pointing out the pair of POIs closest to each other. Burigat et al. [2006] observed equivalent results for Halo, scaled arrows, and stretched arrows.

However, participants preferred Halo and scaled arrows. Burigat and Chittaro [2011b] evaluated a 2D map, Wedge and scaled arrows. The former is significantly faster, more accurate, and preferred than the latter two.

7.5.7 Order Task

In this task, participants must order the POIs by increasing distance to a determined position. Gonçalves et al. [2011a] found no significant results between Halo and Halodot. Burigat and Chittaro [2011b] determined that Wedge and scaled arrows were significantly better than a 2D map. Last, Burigat et al. [2006] observed that Halo is significantly slower and less accurate than scaled and stretched arrows. Participants' preference goes to scaled arrows.

7.5.8 Multiple Object Tracking Task

This task consists of tracking several moving POIs, initially shown by flashing them at the start of the trial for a brief period. Gustafson and Irani [2007] are the only ones to have studied this task, and unfortunately, they did not obtain any significant results because, in their opinion, the number of participants was too small. They did, however, observe a tendency for Edge Radar to be more accurate than Halo.

7.5.9 Memory task

This task involves memorizing the position of POIs and repositioning them without the aid of the device. Schinke et al. [2010] compared 2D radar and 3D compass. Their results demonstrated that a 3D compass is significant for identifying POIs and angular deviation. Besides not being significant, the 3D compass is also slightly faster.

7.5.10 Speed Selection Task

In this task, participants must select as many targets as possible in an allocated time. For ten targets, Henze et al. [2010] observed that Halo is significantly more efficient than stretched arrows, which is significantly better than scaled arrows. For twenty targets, they determined that scaled and stretched arrows are significantly more efficient than Halo. For thirty targets, scaled arrows are significantly more efficient than Halo and stretched arrows.

7.5.11 Estimate Task

This task consists of estimating the distance of an off-screen POI. Miau and Feiner [2016a] observed that a 2D compass is significantly more accurate than Wedge, but Wedge is significantly faster than a 2D compass.

7.5.12 Described Task

In this task, participants describe what they see while exploring a video with off-screen POIs. Lin et al. [2017] evaluated pictures and 2D arrows. Pictures were significantly better in perceiving spatial context, understanding storylines, and preferred by participants.

7.5.13 Closest and Avoid Task

This task is a combination of the closest and avoid tasks. Participants must select the closest targets by simultaneously avoiding danger. According to the study of Burigat et al. [2012], a 2D map significantly outperforms Wedge in time completion, accuracy, workload intensity, and game enjoyment.

7.5.14 Closest and Search Task

This task is a combination of closest and search tasks. Participants must click on the closest relevant target. Gonçalves et al. [2013] determined that scaled arrows are significantly faster than Halodot and a 2D map. Participants also significantly preferred scaled arrows over Halodot. Moreover, Carmo et al. [2020] observed participants' preference for a 2D map rather than 2D radar and Edge radar.

7.5.15 Order and Search Task

This task is a combination of order and search tasks. Participants must search POIs in order. Carmo et al. [2020] found that 3D arrows are significantly faster than Halo and 2D arrows. Next, Wieland et al. [2022] observed significant participants' preference for 3D compass rather than 3D Halo and Ariadne's thread. The authors state that the latter is not liked due to the curved shape. In addition, 3D compass is significantly faster than Ariadne's thread, which is significantly faster than 3D Halo. 3D compass and Ariadne's thread are significantly more efficient than 3D Halo regarding screen rotation. Finally, 3D compass is significantly less workload-intensive than 3D Halo.

7.6 Comparison of Visual Idioms in Immersive Environments

This section compares the idioms presented in Section 7.5 based on their immersive performance in literature evaluations. We have identified twelve types of tasks and will discuss each separately. We aim to determine the best idiom by presenting significant study results. Note that the authors used the NASA TLX form [Human Performance Research Group, 1980] to calculate the workload. Moreover, the threshold success for the System Usability Scale (SUS) is 68 [Brooke, 2013], indicating that the approach is in the average. This threshold is based on the study of Bangor et al. [2008, 2009] on 3,500 SUS results compared to people's ratings.

7.6.1 Search Task

This task consists of finding a target while standing on the spot or moving in the environment. Regarding the former, Buchmann et al. [2008] compared a Horizontal compass, 2D arrows, and two 2D compasses. They concluded that 2D compasses were significantly faster, had less overshooting, and were preferred. Regarding the two 2D compasses, the head-up version was better than the perspective one. Later, Gruenefeld et al. [2018a] evaluated HaloMR and WedgeMR but found no significant results. Nevertheless, the two approaches passed the acceptable usability threshold of the SUS. The same year, Gruenefeld et al. [2018b] studied EyeSee360 and 3D arrows but, again, found no significant results. This time, they observed that 3D arrows are over the SUS threshold but not EyeSee360. The participants have also shown a preference for 3D arrows. Next, Hein et al. [2020] compared Attention Funnel, Spherical Wave-Based Guidance (SWAVE) and 2D arrows. In addition, they propose to combine Attention Funnel and SWAVE with 2D arrows. These approaches displayed 2D arrows under a fixed angular distance to the target; otherwise, it was the Attention Funnel (resp. SWAVE) that was displayed to guide the user. Overall, the combination of Attention Funnel with 2D arrows performed significantly in terms of efficiency, search time, and pragmatic user experience score. Furthermore, Attention Funnel was significantly more efficient than SWAVE, which was significantly faster than 2D arrows alone. 2D arrows were also the worst in hedonic user experience scores. Last, Harada and Ohyama [2021] conducted a study with 3D radar, Moving window, Radiation, 3D compass, and Spherical Gradation. They added a timer to the task in contrast to the others. They observed that Moving window and Radiation were effective for guidance in the frontal direction, and 3D compass and 3D radar were effective for guidance in the backward direction. In addition, they suggested that Moving window was effective guidance as it was better at helping the user ignore distractors than other approaches. Furthermore, their results suggested small cognitive loads for the Moving window and radiation and high cognitive loads for 3D compass, Spherical gradation, and 3D radar.

Concerning the studies that involve moving in the environment, Funk et al. [2014] evaluated a 2D map and 2D arrows combined with the last captured picture of the target. They did not find significant results, but their participant's feedback revealed a preference for the 2D map because its global view makes it easier to select the shortest route. Wallgrun et al. [2020] studied a 2D radar, 2D arrows and a guide in a VR tour guide scenario. They concluded by confirming that all three approaches effectively guide the user but that the results did not allow them to favor one regarding performance. Nevertheless, 2D arrows were the most preferred, intuitive, easy to learn and use, aesthetically pleasing, helpful, and less annoying. Moreover, Arntz et al. [2020] compared 3D arrows with Ariadne's thread. Their results demonstrated that 3D arrows had a significantly smaller path length than Ariadne's thread. In addition, the latter was perceived as easier to use and as better support for path indications. Last, Renner and Pfeiffer [2020] conducted a study with 3 types of Ariadne's thread (i.e., for one target at chest height (AT1), for multi targets at chest height (AT2), and for one target at floor height (AT3)) and a 2D map. Their results demonstrated that AT1 and AT2 were significantly faster than AT3. AT1 was significantly faster than the 2D map. AT3 had a significantly higher number of head movements. A majority of participants stated they prefer a technique for multiple POI.

In conclusion of this task, for studies where users stand on the spot, the 2D compass, Attention Funnel, and Moving Window seem to be the better approaches in this state of the literature. Attention Funnel is also better when combined with 2D arrows, as it obstructs the user's view less when close to the target. For studies where users can move in the environment, Ariadne's thread and 2D map stand out from the other approaches evaluated.

7.6.2 Locate Task

This task involves indicating the expected location of the off-screen POI in the off-screen space. Gruenefeld et al. [2017a] conducted a study with Halo, Wedge, and 2D arrows. Their results indicated that Wedge was significantly more accurate than Halo, which was significantly more accurate than 2D arrows. Concerning usability, Wedge was slightly more usable, according to Gruenefeld et al. [2017a], as it passed the SUS literature threshold compared to the other. The same year, Gruenefeld et al. [2017b] evaluated the EyeSee360 approach with Halo, Wedge, and 2D arrows. The results demonstrated that EyeSee360 had a significantly lower direction error than the other three. One year later, Gruenefeld et al. [2018a] evaluated HaloMR and WedgeMR, for which they found no significant results between the two approaches. In Gruenefeld et al. [2018b], they compared EyeSee360 and 3D arrows, for which, again, they did not find significant results. Concerning the SUS, the results were reported in the search task; HaloMR, WedgeMR, and 3D arrows were over the threshold, while EyeSee360 was not. Furthermore, Yu et al. [2020] conducted a study on 2D radar, 3D radar, 3D Wedge, and 3D compass. Their results showed that 3D compass was significantly faster than 3D radar or 2D radar. In addition, 3D Wedge, 3D compass, and 3D radar were significantly more accurate than 2D radar. 3D Wedge was more appreciated by the participants than the three others. Based on these results and three other tasks evaluated in the paper, Yu et al. [2020] proposed the 3D Wedge* that consists of a 3D Wedge combined with a dashed line and a ball arrow to reinforce the direction and distance understanding. This version of Wedge was even more appreciated by the participants than the other approaches they proposed. In conclusion of this task, there does not seem to be one approach that stands out more than another. However, 2D arrows, 3D radar, and 2D radar can be avoided.

7.6.3 Pick-up Task

This task involves picking up objects in boxes. The number of boxes is too large to all fit in the user's field of view. Schwerdtfeger et al. [2006]; Schwerdtfeger and Klinker [2008]; Schwerdtfeger et al. [2011] conducted a series of evaluations. In the first [Schwerdtfeger et al., 2006], they compared a textual list, a 2D map, and Ariadne's thread, for which they did not demonstrate significant results due to problems with depth perception. The second evaluation studied Attention Funnel and Ariadne's thread [Schwerdtfeger and Klinker, 2008], the latter pointing to either a 3D arrow or a frame to indicate the right box. Although Ariadne's thread combined with a frame demonstrated better results in time completion and accuracy, many participants preferred the Attention Funnel over Ariadne's thread because it indicates the distance. Therefore, in their third and fourth studies [Schwerdtfeger et al., 2011], they tried to combine Attention Funnel and Frame idioms. They did not find significant results except for the rectangular opaque Attention Funnel, which was slower than the other conditions (i.e., rectangular and ring semi-transparent Attention Funnels + frame and Ariadne's thread + frame). Later, Renner and Pfeiffer [2017a] adopted a picking task in a construction scenario. They did not obtain significant results between the relevant idioms, but 2D arrows and SWAVE were those that generally did best compared to Attention Funnel and 2D arrows with flickering of the view when approaching the target. Concerning the latter, participants found it too distracting.

In Renner and Pfeiffer [2017b], they used the same environment to compare the Arianne thread with 2D arrows. Their results indicated that the Arianne thread is significantly faster and is also preferred to 2D arrows that are said to be too inaccurate. Renner et al. [2018] extended their works by comparing Ariadne's thread and SWAVE, but their results did not find significant results. In conclusion of this task, Ariadne's thread seems to be the most appropriate choice, but it is better to find a way of indicating the distance to the target.

7.6.4 Closest Task

In this task, participants must select the closest POI of a determined position. Yu et al. [2020] studied 2D radar, 3D radar, 3D Wedge, and 3D compass. 3D compass and 3D Wedge were significantly faster than 3D radar. 3D Wedge was more appreciated by the participants than the three others. Subsequently, Yu et al. [2020] proposed the 3D Wedge* described in Section 7.6.2. A new study demonstrated that 3D Wedge* and 3D Wedge were significantly faster than 3D radar.

7.6.5 Cluster task

This task involves pointing out the pair of POIs closest to each other. Yu et al. [2020] observed that 3D compass was significantly faster and more accurate than 2D radar. No significant results were found for the 3D radar and 3D Wedge. 3D Wedge was more appreciated by the participants than the three others. Subsequently, Yu et al. [2020] proposed the 3D Wedge* described in Section 7.6.2. A new study demonstrated that 3D Wedge* and 3D Wedge were significantly faster than 3D radar.

7.6.6 Order Task

In this task, participants must order the POIs by increasing distance to a determined position. Yu et al. [2020] found that, in low density (i.e., four targets) 3D Wedge and 2D radar were significantly faster than 3D radar, and 3D Wedge was also significantly more accurate than 3D compass and 2D radar. In high density (i.e., nine targets), no significant differences were found. 3D Wedge was more appreciated by the participants than the three others. Subsequently, Yu et al. [2020] proposed the 3D Wedge* described in Section 7.6.2. A new study demonstrated that 3D Wedge* and 3D Wedge were significantly faster than 3D radar. In addition, 3D Wedge* and 3D Wedge were significantly more accurate than the three other approaches in low-density scenarios. In high density, 3D Wedge* was significantly more accurate than all other approaches.

7.6.7 Driving Task

In this task, participants must drive a car in a simulator and read a number on a paper as fast as possible, symbolizing a danger on the road signaled by a visual idiom for off-screen POIs. Tonnis et al. [2005] evaluated 2D radar versus 3D compass. They did not say whether their results are significant, but 2D radar was faster, more accurate, and preferred over 3D compass. They hypothesized that the monoscopic 2D projection of the 3D compass helped to explain this result and conducted a second experiment described in Tonnis and Klinker [2006] to verify this. This time, 3D compass was significantly faster than 2D radar, but there were no significant results for the accuracy. In addition, 3D compass was preferred but again without significant effect.

7.6.8 Speed Selection Task

In this task, participants must select a predetermined amount of targets as fast as possible. Bork et al. [2018] compared 2D radar, 3D radar, Edge radar, Aroundplot, EyeSee360, and 3D compass. Their results demonstrated that Aroundplot, 3D compass, 2D radar, and EyeSee360 were significantly faster than Edge radar. In addition, 3D compass, 2D radar, and EyeSee360 were significantly faster than 3D radar. Furthermore, EyeSee360 was significantly faster than Aroundplot and 3D compass. Concerning the SUS scale, EyeSee360 and 2D radar obtained high scores, followed by Aroundplot and 3D compass. The two others were under the threshold of 68. Moreover, EyeSee360 had a significantly higher usability than 3D radar and Edge radar and a significantly higher SUS score than 3D compass. Finally, EyeSee360 had a significantly lower mental effort level than 3D compass, Aroundplot, 3D radar, and Edge radar. 2D radar also had significantly lower mental effort than 3D radar. Next, Hu et al. [2021] evaluated Circle, 3D arrows, and Border light. The latter tended to fade in the background, so their results were worse than the two others. 3D arrows and Circle presented similar results for effectiveness, comprehensibility, precision, helping, effectiveness at grabbing attention, and effectiveness at making aware of off-screen targets. Regarding the workload, 3D arrows were significantly less workload-intensive than Circle, which was significantly less intensive than Border light. Last, Chung et al. [2021] studied a panorama picture against a 2D radar and 3D arrows. They found that 2D radar was significantly slower than the other two. In addition, the panorama picture implied significantly less travel distance than the 3D arrow, which implied significantly less travel distance than the 2D radar. The panorama picture and 3D arrow were also significantly considered easier by the participants. Overall, the participants were most satisfied with the panorama picture, while they preferred the panorama picture and the 3D arrow.

7.6.9 Multiple Object Tracking Task

This task consists of tracking several moving POIs, initially shown by flashing them at the start of the trial for a brief period. Gruenefeld et al. [2019a] are the only ones to have studied this task. They found significant results concerning the accuracy in favor of the 2D radar rather than EyeSee360. In addition, the 2D radar had a SUS score beyond the threshold, but this was not the case with EyeSee360, contrary to previous studies. The authors attributed this to the fact that the targets are moving.

7.6.10 Collaborative Pick-up Task

Participants have to find targets and partners to solve a puzzle. Some pieces can be moved alone, others require two people, and still others require a partner to be nearby to pick up the piece. In these conditions, Osmers and Prilla [2020] evaluated 2D map and 2D compass. They observed that the 2D map was generally better perceived in terms of utility, usability, and presence than the 2D compass. They explained this by the fact that the latter did not indicate the position of other people. In addition, the 2D compass was significantly rated as the most demanding in terms of mental workload.

7.6.11 Search and Dual-task

Participants are asked to perform a secondary task, and an off-screen point of interest is indicated at some points. The participant must find the point of interest quickly. Petford et al. [2019] conducted a study with Wedge, Circle, Flashing, and 2D radar as approaches. Regarding the time completion, all comparisons between idioms were significant. The order from the fastest to the slowest were Wedge, Flashing, 2D radar, and Circle. The workload analysis led to similar results. Ward et al. [2016] compared a central 2D arrow and Moving window. The former was significantly faster than the latter. In addition, participants made no directional errors with the 2D arrow, but 40% did with the moving window.

7.6.12 Estimate and Search Task

This task is a combination of the estimate and search tasks. First, participants must estimate the distance to the target. Next, they must search for the target. Gruenefeld et al. [2017b] evaluated the EyeSee360 approach with Halo, Wedge, and 2D arrows. They observed no significant results for distance accuracy and, regarding the SUS, Wedge was slightly better than EyeSee360, which was slightly better than Halo.

7.7 Discussion

This section provides a detailed discussion of the three research questions. For each, we present the results obtained from our literature review and elaborate on future research avenues these results open.

7.7.1 Research Question 2.1: Visual Idioms

Overall, we have identified 36 visual idioms divided into nine main families. The foundations of each family are very different from one another. Within these families, idioms can take on numerous forms, depending on the design choices made in each study. These choices are part of a complementary approach to our own to help determine which idiom to choose. For example, Li [2016] demonstrated that color value contributes less than size in distance perception. Therefore, if the distance is a critical requirement, designers should select idioms allowing to play on the size of the POIs. Studying these parameters is a future work to cross-reference results and offer a selection guide. This guide will be of help for both scientific and practice. The former will obtain a review identifying gaps and future works. Depending on their context, the latter will receive strong recommendations on which idioms to use.

Another point is that certain combinations of idioms produced exciting results. The main idioms used in these experiments are 2D map, arrows, and Arianne's thread. Either the idioms were used simultaneously or switched depending on some conditions. Combining idioms reinforces their advantages and overcomes their disadvantages. So, we would recommend further research into which idioms should be combined. Binetti et al. [2021], Chewar and McCrickard [2002], and Tonnis and Klinker [2006] have also shown that this is not only true for visual idioms but that combining them with audio is also of great interest.

7.7.2 Research Question 2.2: Tasks

We identified fifteen tasks for evaluations conducted on a 2D screen and twelve for those in immersive MR. These tasks are again varied, but the main ones are shared between the two sets. The search task involves reaching a target in the environment involving displacement or simple rotation. As mentioned in the introduction to this chapter, this is the definition of navigation. It is hardly surprising that this is the task most frequently used in evaluations. A variant in immersive scenarios is the pick-up task that adds the step to pick up the target element. This task is inspired by several industrial scenarios where the user must retrieve various parts and tools to complete the task. Alternatively, the locate task helps to determine which idioms are best at helping the user to create a mental representation of the environment with an accurate estimation of the direction and distance of POIs. Studies have shown that direction and distance are two essential properties of navigation. Finally, the closest task tests whether the user is able to determine the nearest element. This task is more likely to be used in decision-making situations. It will be quite frequent when users explore an environment, such as tourists in a city. In any case, we recommend including the search task at least in future evaluations. Still, it would be preferable to include all three as they are the most representative (i.e., search, locate, and closest tasks).

7.7.3 Research Question 2.3: Best idioms by tasks

Although some idioms seem to be emerging, it is not possible to determine the best idiom for each task. Therefore, it is not yet possible to create decision trees as in Chapter 6. There are two main reasons for this. The first is that for most tasks, too few idioms were compared. Overall, there is no single task where all idiom families have been tested. The second reason is that there are a significant number of factors influencing the results. First, there is the way in which each idiom has been implemented and designed; this differs in each study, which the same authors do not carry out. Then there are material constraints such as the devices used [Renner and Pfeiffer, 2017c]. The context in which the experiment takes place is also essential since experimental conditions are not standardized. Finally, the user himself is the most critical factor. Studies have shown differences between users according to age [Pampel et al., 2018; Tang and Zhou, 2020] and gender [Ahmad et al., 2005; Czerwinski et al., 2002; Dünser et al., 2012; Li and Chen, 2021], for example. In addition, Chewar and McCrickard [2002] also observes a difference between people with left and right brain hemisphere dominance. For the future, we recommend conducting new experiments with the emerging idioms for each task and trying to standardize as much as possible to have experimental conditions as close as possible. Statistics on participants' characteristics must also be included. By the way, a question we also raised in the previous chapter is whether some of the results achieved on older headsets are still relevant, given the technological advances of today's headsets. Overall, Our priority for the future is a more in-depth study of the results obtained, extracting, in addition to the statistical results, the various hypotheses raised by the authors regarding their results. This will enable us to further guide future research on the techniques and points of improvement to be prioritized to contribute to this field.

7.7.4 Limitations

We have included in our survey evaluations that were carried out on 2D screens, whereas our focus is on immersive augmented reality. This choice was made assuming that results on these devices could guide a choice in immersive scenarios. However, the results do not allow us to validate or invalidate this hypothesis. We learned that the devices did impact the results [Renner and Pfeiffer, 2017c], but it is unclear what that impact is.

In this chapter, we have once again presented results based on user preferences. However, it is essential to note that several factors, including social and cultural influences, can significantly impact the outcomes. Nonetheless, we have highlighted only the significant results, although this limitation should not be disregarded. In addition, some of the idioms we have identified have physical versions in the real world. For example, 2D cards existed before the advent of computers, and are part of everyday life for many people. Therefore, while the approaches may differ, the fundamentals remain the same.

Finally, our approach to paper identification is open to discussion. By basing our search almost exclusively on snowballing, we did not allow ourselves to include certain papers that could be isolated. However, the terms used in this field are relatively generic, making it difficult to use a classic keyword approach. One solution to counterbalance this is to strengthen the coverage assessment phase with additional checks in the major conferences' proceedings, which will be done in the further exploration of this research question.

7.8 Summary

We conducted a systematic literature review to identify the visual idioms and the tasks used to address off-screen POIs. In total, we found 36 idioms distributed in nine families, 15 tasks on a 2D screen, and 12 tasks in an immersive environment. Among these tasks, four are the main ones, covering the main aspects of navigation. These are reaching an off-screen POI, creating a mental image of the environment, and decision support for selecting the nearest target. We have reported the significant results found in evaluations comparing idioms for each task. The results do not allow us to select the best idiom, but they reveal the main approaches to be interested in for the future. In addition, we have highlighted various factors impacting evaluations. Finally, we conclude with the proposition of future research directions to extend the results and promote them outside scientific publications.

CHAPTER 8

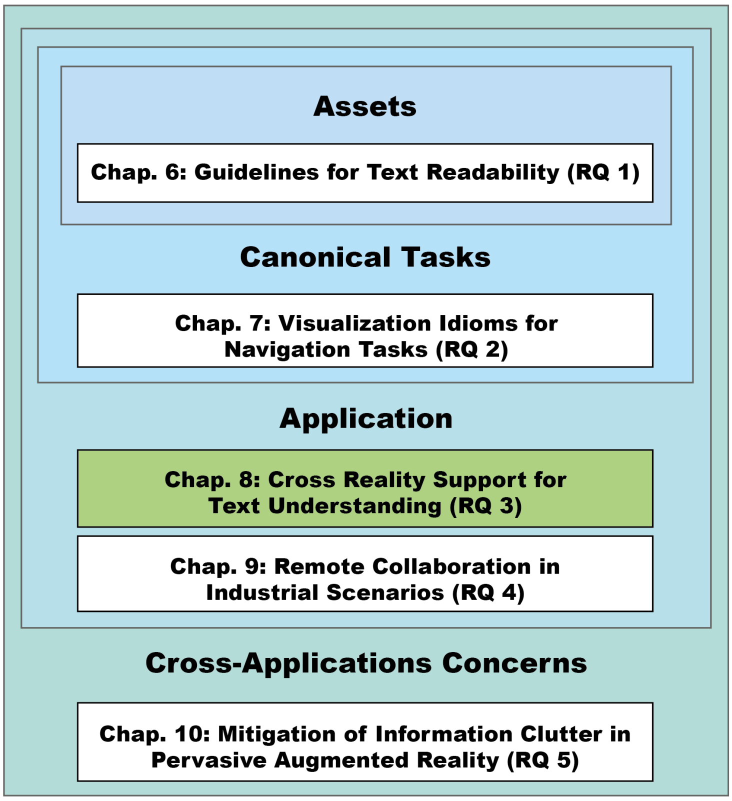
CROSS-TERMINAL SUPPORT FOR TEXT UNDERSTANDING

8.1 General Introduction

In this chapter, we focus on answering research question 3: **"How can Immersive Augmented Reality support existing conventional applications for text understanding?"** Many human sciences research protocols rely on analyzing interview transcripts, and several IT tools are already available to support this task. However, most of these tools leave the interpretation task to the analyst or involve an implicit conception of language, which is rarely questioned [Linden et al., 2020]. Various software supports text analysis, many based on visual analytics principles, as discussed in Chapter 3. In an in-depth study of their functionalities, Lejeune [2010] categorized them into five families based on two axes: compute/show and explore/analyze. In their conclusions, they note that although these functionalities are helpful, they cannot replace the researcher's ability to generate a stimulating interpretation of the text, which requires a robust theoretical and methodological framework.

The **Evoq** tool, developed in the context of the Evocative Framework For Text Analysis - Mediality Models (EFFaTA-MeM) transdisciplinary research project [Clarinval et al., 2018], takes a radically innovative approach to text analysis. It has been created as a web text analysis tool for experts and researchers in social sciences and humanities. Its design follows a text analysis method inspired by post-structuralism [Barthes, 1966; Cooper, 1989; Derrida, 1967; Piret et al., 1996] (explained in Section 8.2) to support the actual practice of target users [Linden et al., 2020]. The tool enables manual analysis supported by interactive visualizations that generate automatic behaviors, which differs from an automated analysis of text highlighted by interactive visualizations.

As discussed in Chapter 3, combining immersive approaches and visual analytics



has led to the development of **Immersive Analytics**. It refers to *"the use of engaging, embodied analysis tools to support data understanding and decision-making"* [Marriott et al., 2018]. The main idea behind Immersive Analytics is to reduce the barrier between the data and the analyst by providing visualizations that allow users to immerse themselves in the data, fully or partially. This approach provides users with a new perspective and a more omniscient point of view through a 3D space, which offers new analysis possibilities and a different perspective on the data. This, in turn, should lead to a better conceptualization of the data. As stated by Schwajda et al. [2023], the strength of combining 2D and 3D interfaces appeared with the works of Feiner and Shamash [1991] and Kijima and Ojika [1997]. While most works use this approach to extend the work area [Schwajda et al., 2023], some have also noted its value in collaboration [Büschel et al., 2021; Seraji and Stuerzlinger, 2022a,b].

This chapter looks at how three metaphorical and immersive approaches to a word graph can support an expert's analysis using the Evoq application. Combining conventional and immersive applications leads us to cross-terminal systems. In other words, systems designed on different levels of Milgram and Kishino [1994]'s continuum, introduced in Chapter 2. The contributions of this chapter are twofold. First, we explore how three immersive visualizations can be added to existing visualizations within a web application, offering a new vision of the analyzed text. Second, we conduct an exploration of three immersive metaphors for the visualization of text analysis.

Publications and supervised works

The content of this chapter is based on peer-reviewed publications in scientific conferences and student works for which I was co-supervisor:

Linden, I., Wallemacq, A., Dumas, B., Deville, G., Clarinval, A., and Cauz, M. (2020). Text as Semantic Fields: Integration of an Enriched Language Conception in the Text Analysis Tool Evoq®. In **Research Challenges in Information Science (RCIS)**, volume 385, pages 543–548, Cham

- This demo paper briefly introduces the main concepts of poststructuralism. Then, it presents how these concepts are modeled in a formal system. Finally, it shows how this approach is integrated into the **Evoq** software and how human scientist users can benefit from its functionalities.

Cauz, M., Albert, J., Wallemacq, A., Linden, I., and Dumas, B. (2021). Shock wave: a graph layout algorithm for text analyzing. In **Proceedings of the 21st ACM Symposium on Document Engineering**, Article 32, pages 1–4

- This paper describes the node-link diagram implemented in the Evoq tool. On the one hand, it focuses on the force algorithm that ensures the logic of the graph through time and manipulation, and on the other, it also introduces the new placement algorithm called Shock Wave, which places nodes initially in a position that allows visualization interpretation.

Sanfilippo, U. (2023). État de l'art: Visualization de graphes 3D en réalité augmentée. Unpublished technical report, University of Namur

- This student work aimed to carry out a state-of-the-art study of 3D graph visualization. The student thus analyzed the representation of nodes, links, and graphs through their classical representation and that of metaphors. This work provided an introduction to the literature on the subject.

Dejardin, G. (2022). CSApp, A Mixed Reality Game for Cybersecurity Awareness. Master's thesis, University of Namur

- Although not directly related to the content of this chapter, this master's thesis explores the use of daily life metaphors to facilitate understanding of cybersecurity concepts through an immersive serious game. It briefly examines the application of this concept to a different use case than the one discussed in this chapter.

Outline

The organization of this chapter adheres to the subsequent structure. Section 8.2 introduces the web tool and the analysis process it supports. Following that, the graph visualization on which the immersive solutions were inspired is described in greater detail in Section 8.3. The solution combines geometric notions and business constraints to provide real-time visualization tailored to the needs of the analysis process. Then, Section 8.4 discusses the three immersive visualizations and their integration within the web tool. Finally, in Section 8.5, we discuss the implications of our results, limitations, and future works before concluding in Section 8.6.

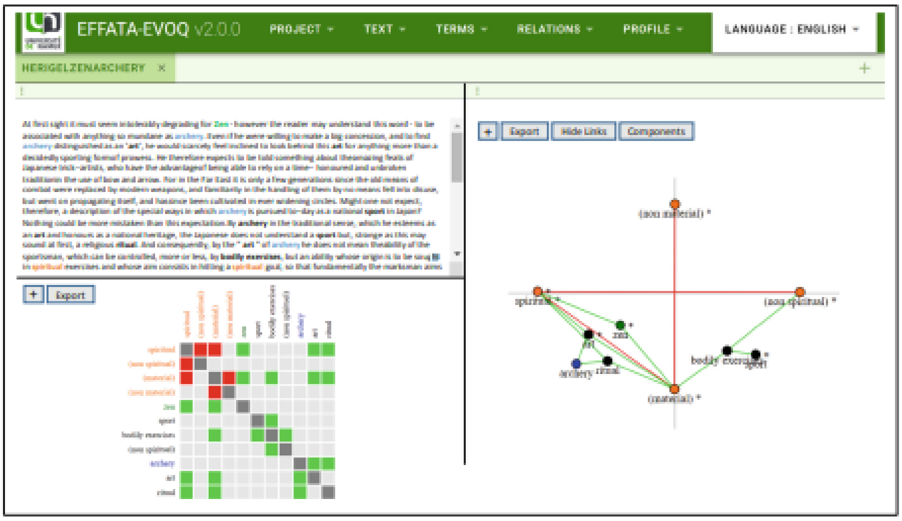


Figure 8.1: Interface of the Evoq tool: the text itself is displayed in the top-left corner, the matrix of relations in the bottom-left, and the node-link diagram on the right.

8.2 The Evoq Tool

The Evoq tool is designed to assist analysts in identifying tensions between pairs of words in a given text. Rather than viewing the text as a linear progression from start to finish, the tool treats it as a semantic landscape. The analyst pairs words based on association or opposition to use the tool. For instance [Clarinval et al., 2018; Linden et al., 2020], they may associate words such as "*white, pure, and good*" on one side and "*black, dirty, and bad*" on the other to represent two distinct semantic fields. By adding an oppositional relationship between "*white*" and "*black*", they can highlight the underlying concepts of "*purity*" and "*impurity*" and the tension between the two semantic fields.

The EVOQ tool is designed to help users identify relevant words or phrases, called **terms**, in a text and connect them through binary **relations** of **association** and **opposition**. The tool also offers a range of visualization and interaction tools to help users explore the meaning of a text [Clarinval et al., 2018]. These include the text itself, a matrix of relations, a list of terms and relations, and a node-link diagram [Tarawaneh et al., 2012] (see Figure 8.1). The node-link diagram is the primary visualization tool in EVOQ, allowing users to arrange all the text concepts regardless of reading order and structure them according to opposition relations, which are more important than association relations for the analysis method [Linden et al., 2020]. The node-link diagram is an interactive analysis space where users can manipulate and play with the visual representation. They can add or remove terms or relations and change the interpretation of the text by moving nodes.

8.3 Node-Link Diagram

The graph layout in the node-link diagram must meet some requirements arising from our target users' needs. It should respect our analysis method, allow the user to understand a text's structure more efficiently, and reduce the number of nodes displaced by the user. In particular, the opposition relations should structure the graph layout because they are the most important according to the text analysis method. The nodes of a connected component by association relations, which form a semantic field, should also be placed near one other. They should not overlap the area dedicated to another connected component (i.e., another semantic field). Finally, the general readability of the graph layout should be preserved.

As stated in the literature on graph layout algorithms [Gibson et al., 2013; Kobourov, 2013] (i.e., algorithms that compute the position of nodes on a node-link diagram [Tarawaneh et al., 2012]), the main paradigm for computing layouts is to use force-directed methods. These methods consider the graph as *"a physical system where nodes are attracted and repelled according to some force"* [Gibson et al., 2013]. To do so, they mainly use the structural properties of graphs. For instance, Eades' Spring Embedder Algorithm [Eades, 1984], which is one of the first main approaches and the basis for almost all force-directed techniques [Gibson et al., 2013], considers nodes as steel rings and links as springs. Starting from a random configuration, the nodes are positioned according to the resulting forces applied to them until the system reaches a stable state. However, force-directed methods alone can hardly deal with the vital user requirements of our case. The different types of links (i.e., associations and oppositions) and the prevalence of oppositions over associations need to emerge from the structure of the resulting layout.

Numerous recent contributions in the domain focus on computing layouts for more extensive graphs [Kobourov, 2013]. Despite addressing a different challenge from us, some techniques primarily designed to address this problem can also be relevant in our case. This is notably the case of multi-level techniques, which combine sequentially different techniques to compute layout. These techniques are mainly motivated by the will to reduce the complexity of the global computation [Gibson et al., 2013]. Hendrickson and Leland [1995] proposes an example of these who used a multi-level technique to minimize the number of crossing links [Gibson et al., 2013]. Another idea is to introduce a preprocessing step to initialize the algorithm rather than starting from a random state [Kobourov, 2013]. For example, Fowler and Kobourov [2012] compare different methods of initial positioning with different force-directed algorithms to see if there is any aesthetic improvement compared to random positioning. This last idea is particularly relevant in our case because this preprocessing step may be an occasion to inject some user requirements into the graph layout.

Inspired by this last idea, our solution to computing graph layouts [Cauz et al., 2021], described in the following sections, combines a **force-directed algorithm** and a **placement algorithm**. The force-directed algorithm follows the state-of-the-art solutions. In contrast, **Shock Wave** is a novel placement algorithm designed to answer our user requirements while preserving efficiency with more extensive graphs.

Figure 8.2: A graph layout in the Evoq tool with a random initial placement of nodes on the left and an initial placement done by the Shock Wave Algorithm on the right. As can be seen, our algorithm helps to distinguish the different semantic fields extracted by the user (i.e., groups of nodes linked by association relations association (green links)). It also highlights opposition relations (red links), a key part of interpreting these semantic fields.

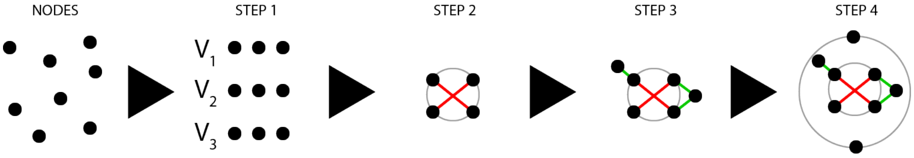


Figure 8.3: The Shock Wave algorithm is divided into four steps: 1) categorizing the nodes by their direct, indirect, or no implication in opposition relation, 2) placing the nodes of the first category (direct opposition), 3) placing the nodes of the second category (indirect opposition), and 4) placing the nodes of the third category (no implication in opposition relation).

8.3.1 Force-Directed Algorithm

The force-directed algorithm used in the Evoq tool directly follows Eades' Spring Embedder Algorithm [Eades, 1984] but with some adaptations specific to the user requirements explained above. It works by applying different types of forces to nodes. First, a repulsive force is applied to all pairs of nodes to ensure sufficient space between them. Second, an attractive force is applied to couples of nodes linked by an association relation. Third, a repulsive force is also applied to couples of nodes connected by an opposition relation. The strengths of these types of forces depend linearly on the distance between involved nodes. To ensure convergence, a limit to the number of iterations to compute the final position is also fixed (20 in our implementation).

All the repulsive/attractive forces performed are based on minimal and maximal distances for each relation type or the absence of a relation. If two nodes have no relation, they must respect a minimal comfort distance to prevent overlap of labels (i.e., a node is represented by a dot with the term as a label). The maximum distance between two associate nodes must be less than the minimum distance between two opposite nodes. Figure 8.2 is an example of the same graph in two different equilibrium states due to the different initial placement algorithms.

8.3.2 Shock Wave Algorithm

The algorithm is divided into four successive steps (see Figure 8.3). The first step categorizes nodes into three sets, V_1, V_2, and V_3, depending on their implication level in opposition relations (i.e., respectively direct, indirect, or no implication). The next three steps place the nodes from the different sets on the visualization in three waves, starting from the center towards the border. The name **Shock Wave** comes from this arrangement in successive circles. The choice of circles as the main pattern to arrange nodes is motivated by the will to preserve the readability of the layout independently of the graph configuration. In addition, the first circle divides the space in two to keep the opposition relation inside and the rest outside. In this way, a circular arc never overlaps another node than the two attached to it (independently of the number of nodes on this circle), and the user can perceive the different axes structuring the text more easily. Finally, the circular construction guides the user's gaze towards the center, where the opposition relations stand.

The nodes directly implied in an opposition relation are categorized in V_1. By positioning them, the algorithm places all the opposition relations. The nodes with a path to a node of V_1 and not in V_1 are included in V_2. The nodes in V_2 represent semantic fields that help enrich oppositions described in V_1. In the example explained above, not only *white* and *black* are in opposition, but the concepts behind (i.e., *purity* and *impurity*), that may only be understood with the other associated words (i.e., *pure* - *good* and *dirty* - *bad*). Finally, the rest of the nodes form the set V_3. This set comprises terms that may be irrelevant to the analysis or terms that might be connected later to nodes from V_1 or V_2 by the user. Due to their yet unclear utility, these are placed on the outer layer of the graph. This way, the user can clearly distinguish them and decide what to do with them.

The placement of nodes in V_1 must highlight the groups of opposition relations by crossing them inside the first circle. Indeed, an associative relation may exist between nodes of V_1. Associative paths may also exist between two nodes of V_1 with some nodes of V_2. For example, it is the case in Figure 8.4 for *Sky* and *Earth*. Thus, the two opposition relations from these terms are in the same group, while the opposition relation starting from *Observe* is in another. Thus, nodes linked by an associative path have to be closer to each other. So, the nodes of V_1 are grouped into *poles* in such a way that each pole contains all the nodes of V_1 that are connected by an association path (i.e. without an opposition relation, and by considering the transitive association through the nodes of V_2). Then, the poles are grouped into connected components by considering only the opposition relations (i.e., a connected component regroups all poles connected by transitive opposition relations). Next, in each connected component, the poles are distributed into two sets by minimizing the number of opposition relations between nodes of the same set. As the number of sets is limited to two, opposition relations can exist between two poles of the same set (e.g., three poles in opposition to each other). Finally, these

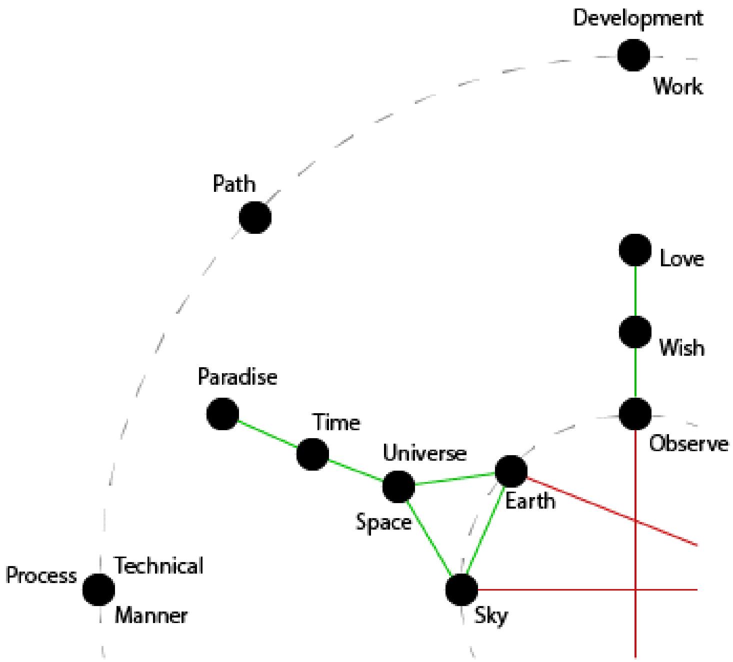


Figure 8.4: Disposition of nodes by the Shock Wave algorithm before executing the force-directed algorithm. The two circles are just displayed for the illustration.

two sets are placed in opposition on the circle to create an axis. The sets of each connected component are intertwined to create the intersection between axes. In the Figure 8.4, the relation from *Observe* crosses the ones from *Earth* and *Sky* as two structural axes. The order of the nodes in each pole and the poles themselves are given by minimizing the number of intersections between the relations. To reduce the permutations, the algorithm chooses the best order of nodes in each pole first and then the best order of poles. For example, if we take two poles with four and five nodes, the number of permutations reduces from 9! (362.880) with a brute force solution to $4! + 5!$ (144) with our heuristic. This trick respects the user's requirements since a pole represents a set of associative nodes that must stay closer to one another. Once the order is known, the nodes of V_1 are placed on the inner circle, where the radius is calculated to ensure that the minimal distance between two nodes is always the minimal distance for an opposition relation (see Section 8.3.1 on the force-directed algorithm). In other words, the nodes are the vertices of a regular polygon inscribed in the circle. Because the nodes of V_1 with an associative path between them are in the same pole, the force-directed algorithm can bring them closer without breaking the circle organization.

The third step consists of placing the nodes of V_2 depending on the nodes of V_1. They are placed outside the circle near their reference nodes in V_1 (a.k.a. nodes of V_1 linked by an associative path). The force-directed algorithm is responsible for computing the exact positions. Because the nodes of V_1 are placed according to the connected components, the reference nodes are close together on the circle. So, the algorithm selects the one or two most centered to choose the position of the nodes of V_2. To preserve user requirements, the nodes are placed on a line depending on the distance to the reference nodes (i.e., the size of the shorter path between them and the reference nodes).

Finally, during the last step, the nodes of V_3 are placed on an outer circle around the nodes of V_1 and V_2. The nodes are grouped by connected components and separated by the force-directed algorithm like the nodes of V_2. Figure 8.4 illustrates the position of a few nodes of each category before executing the force-directed algorithm.

8.3.3 Preliminary Feedback

The initial evaluations of the Shock Wave Algorithm were done as part of the evaluation of the Evoq tool, with mainly a qualitative approach. The evaluation of the Evoq tool is a continuous process that is an integral part of its development, and the validations we present below are part of this evaluation cycle. In this way, we can discover the practices and needs of users through observation, which facilitates their understanding and, ultimately, their integration into the tool.

Once the second prototype was mature enough, some sessions of evaluation, based on the quasi-empirical evaluation approach Hartson and Pyla [2012b], were organized with three users who were not involved in the earlier development. These three users were all social scientists with extensive expertise in manual and semi-automatic text analysis techniques. One of the three users had good knowledge of analysis techniques based on post-structuralism.

The first lessons learned with this preliminary evaluation are encouraging. Users underlined the stimulating character of the Evoq tool and its ability to enrich the analysis process. They also enjoy that Evoq supports their actual analysis practice instead of replacing it like some other available tools do. In that regard, the different visualizations offered, particularly the Shock Wave Algorithm, were considered central tools supporting the analysis. Indeed, an essential lesson from the evaluations is that researchers and analysts in social sciences and humanities can use some automation in their tools as long as this automation supports their detailed, word-per-word analysis instead of trying to replace it. This evaluation phase is still in progress, and we intend to expand the panel of users.

In the end, we determined two main challenges. First, we noticed that with the increasing size of the graph, it became difficult for users to get an overall idea of it due to screen limitations and the space occupied by the graph. This is a well-known issue that arises with Big Data, which we have already discussed in the introduction of this thesis. To address this problem, we have implemented a mini-map in the Evoq tool, which provides an overview. However, the mini-map also occupies part of the screen and does not display the words associated with the nodes. As a result, the user needs to memorize the general structure of the graph. A solution will be moving the graph outside the screen to take advantage of the space offered by the user's environment.

Second, the representation of oppositions and associations through links in red and green, respectively, is well perceived. However, we wondered if this symbology could be further strengthened by incorporating common concepts from human

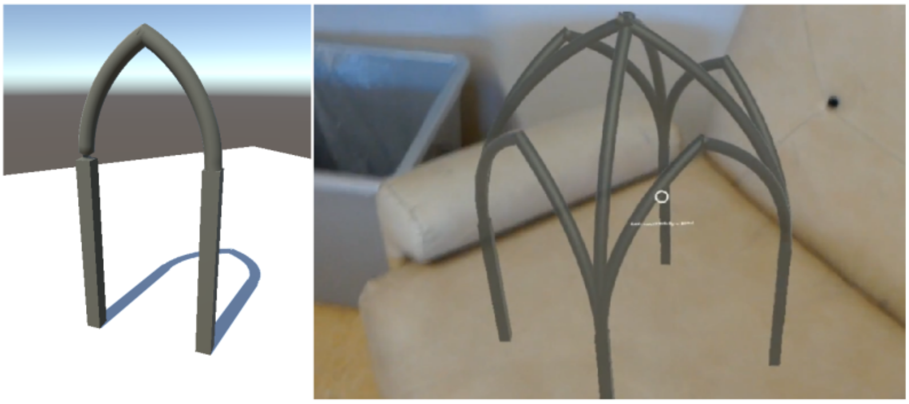


Figure 8.5: Exploration of a text through the construction of a cathedral. The pillars represent the terms, and the arches represent the associations between these terms. (Source: the EFFaTA-MeM project)

.

social and historical backgrounds, such as war and trade. The symbology behind this principle implies that trade between people represents a certain level of friendship, while wars oppose them. In our opinion, immersive technologies can help us to better perceive metaphors by making them more realistic and, in a certain sense, more tangible.

8.4 Immersive Visualizations

To overcome the challenges exposed in the previous section, we relied on metaphors to guide the design of three solutions: the **Atoms** graph, the **Cathedral** graph, and the **Island** graph. The Atom addresses the first challenge, the Island graph addresses the second challenge, and the Cathedral graph seeks to address both. These metaphors resulted from ideation sessions with project members composed of UX and human science experts, practicing the methodology supported by the tool for the latter. All three solutions were considered as supplements to the web tool. Users can access them through a QR code in the application that links the immersive application to the server that manages project information.

The decision to use metaphors was based on a review of the field [Sanfilippo, 2023] and an exploratory study that was conducted on building a cathedral in immersive Augmented Reality (AR) (as shown in Figure 8.5). The cathedral was made up of arches supported by pillars. Each pillar represented a term, while the arches represented oppositions between these terms. The structure was meant to showcase the tensions between these terms, as removing one could cause the entire structure to collapse. However, this approach had limitations in representing association relations in this context.

In their recent study, Wijayawardena et al. [2023] proposed a taxonomy of 3D metaphoric information visualization. The taxonomy is divided into several categories, starting with composite metaphors that make up 65% of the metaphors identified. Composite metaphors are elementary metaphors combined to create a more complex representation. The source of the metaphor can be drawn from the spatial, human-defined, or geographical world. Alternatively, it can be a geometric or a compound object. The most common world metaphor is the city [Wettel and Lanza, 2008] (see Figure 8.6a), which is often used to represent software systems [Wettel and Lanza, 2007] or databases [Meurice and Cleve, 2016]. Another example of composite metaphors is the use of a sphere for the geometric metaphor [Schwajda et al., 2023] (see Figure 8.6c) and an aquarium with plants, fish, and fish animation for the compound object metaphor [Latvala et al., 2017] (see Figure 8.6b). The following categories are the glyphs and structural metaphors such as text visualization, matrix, and dictionaries in the Evoq tool. Finally, natural phenomena, biological, botanical, and other categories are listed. For instance, Barreiros et al. [2018] proposes to correlate the water level, temperatures, and elapsed time before maintenance of a coffee maker to the thickness of the tree's greenery, the leaf color, and the presence of flowers of a virtual tree.

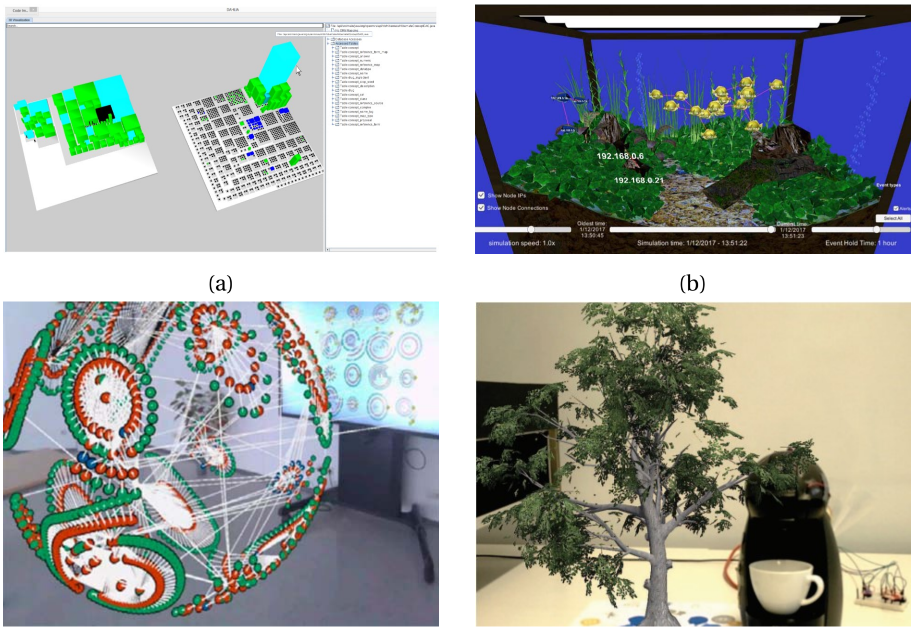
However, comparing different metaphors may be challenging, as their use depends on each use case. For instance, Dejardin [2022] applied them to the cybersecurity field in contrast to the one selected in this chapter. Therefore, Wijayawardena et al. [2023] recommended paying attention to the mapping of elements and their interactions with the properties of each metaphor. Additionally, they determined three metaphoric designs to minimize cognitive load: (1) the use of real-world metaphoric concepts, (2) the use of aesthetics, and (3) animations.

8.4.1 Atom Graph

The Atom graph uses a force-directed algorithm within a 3D spatial context. This means that atoms, represented as spheres, correspond to terms, while molecules, depicted as cubes, symbolize semantic fields established by associative relations (see Figure 8.7). This allows for a dual-level exploration of the graph. The first level shows semantic fields to emphasize oppositions. The second level involves a more detailed examination of these molecules by unfolding the atoms (nodes), allowing for a more precise text analysis. The Atom graph has the distinct advantage of existing in a 3D spatial space, which has been shown to outperform 2D graphs in mitigating issues related to edge crossings [Schwajda et al., 2023]. All nodes within the Atom graph are manipulable through manual gestures, allowing for intuitive exploration and navigation.

8.4.2 Cathedral Graph

The Cathedral graph is based on the cathedral metaphor, as the first exploration described above, but focuses on the tension between the building's pillars rather than the manual construction aspect. In this representation, an opposition relationship forms a dome over the user, which consists of nodes of V_1, upheld by pillars constituted by association relationships, which consist of nodes of V_2. These pillars manifest as stacks of circles, each consisting of nodes having the same distance to their referent node in V_1. The closest nodes are towards the top of the pillar, while the furthest nodes tend to be closer to the ground. Figure 8.8 illustrates the graph with the sphere representing the terms, the lines representing the relations, and the semi-transparent gray cylinders reinforcing the pillar aspect. At this exploratory stage,



(a) (b)

(c) (d)

Figure 8.6: Illustration of 4 metaphors: (a) the city metaphor (Source: Meurice and Cleve [2016]), (b) the aquarium metaphor (Source: Latvala et al. [2017]), (c) the sphere metaphor (Source: Schwajda et al. [2023]), and (d) the tree metaphor (Source: Barreiros et al. [2018]).

V_3 nodes have been omitted; however, these could be positioned on the ground or in the air surrounding the dome structure. Additionally, all nodes can be moved through manual gestures.

This approach aims to change the analyst's perspective on data. Unlike the web application, which allows observation from an omniscient standpoint, the 3D version of Shock Wave requires viewing the graph from within, precluding the simultaneous observation of the entire graph and emphasizing the exploration aspect. The primary focus of this version is to evaluate the concept of exploration, and a secondary objective is to assess the 3D version of Shock Wave. Nevertheless, our hypothesis posits that the Cathedral graph may not be favorably received due to the need to look upwards to fully engage with the metaphorical construct.

8.4.3 Island Graph

The Island graph visually represents the conceptual notions of opposition and association by mixing the geographical metaphor of islands and the war and trade metaphor (see Figure 8.9). An island represents each term within an expansive sea. The relationship between these terms is symbolized by animated naval trading routes and exchanges of cannon fire, effectively illustrating associations and oppositions. The visualization is confined within a defined boundary, and users can

Figure 8.7: Illustration of the Atom graph. The spheres and cubes represent, respectively, the terms and the semantic fields.

Figure 8.8: Illustration of the Cathedral graph. Opposition relations form a dome above the user, while semantic fields are arranged as pillars. A gray tube reinforces the appearance of the latter.

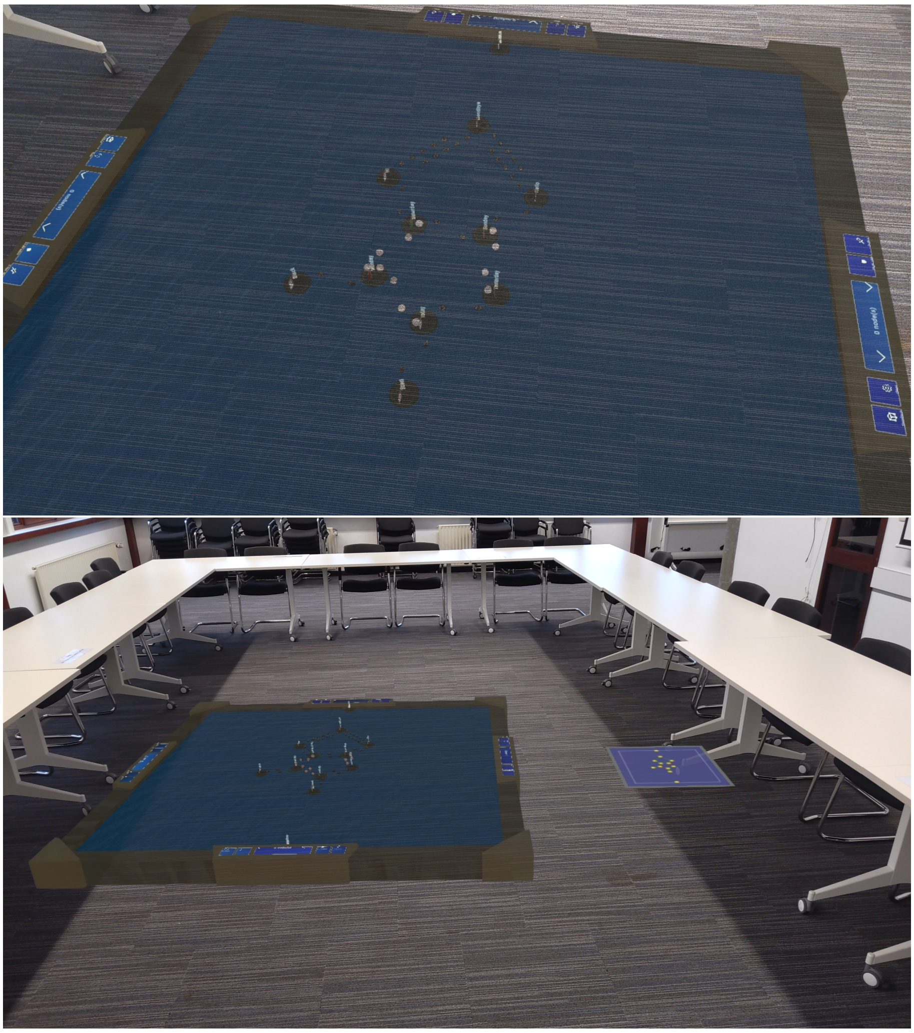


Figure 8.9: Illustration of the Island graph. Naval trades and cannon fires link islands, representing the terms (nodes). The size of the board limits the sea containing the islands. A mini-map is available to visualize the general shape of the graph.

Figure 8.10: Link to a video illustrating the three graph metaphors: https://www.youtube.com/watch?v=qc7VzwzhSHw

manipulate the board's position, rotation, and size using hand manipulations. They can also change the view by moving the sea to move all islands, moving only one island, or zooming in the sea. Buttons are placed on each side of the board to reset the islands' positions to their original positions, and the number of islands hidden beyond the board's limits is indicated for each direction. Moreover, a separated 2D window displays a top-down radar view of all islands and a frame delineating the visible area, which can be manipulated independently of the boundary.

The Island graph is only based on the Shock Wave algorithm, as moving the islands based on the distance between them will break the metaphor. Although the islands are represented in three dimensions, the underlying spatial logic remains in two dimensions due to the board's aspect of a strategic map. This visualization is designed to assist users when working on the web application or collaborating with others. However, it may be challenging to read due to its heavy content, and the notion of war and trade risks may distract from the analyzed text's focus.

8.4.4 Internal Evaluation

We presented the three metaphors to the team's two methodological experts (of the post-structuralism approach). This first evaluation taught us valuable lessons and highlighted several areas for improvement besides the interaction difficulties they encountered due to their low experience with Mixed Reality (MR) technologies. They were asked to explore and appropriate each metaphor through a pre-established analysis. No time limit was set. They gave their live feedback, which two interviewers noted.

The Atom graph was not particularly well received. Experts questioned its usefulness when compared to the web version. This was primarily due to the low number of nodes and relations used to illustrate the graph. The idea of aggregating nodes into a semantic field also failed to prove its usefulness for the same reasons. In the future, more extensive examples will be necessary to evaluate the usefulness of the

Atom graph in contrast to the web graph. Anyway, according to the experts, this approach is not the one that would bring the most benefit to text analysis.

The Cathedral graph was a little bit more appreciated. While experts find some tensions between the terms, they are not sufficiently represented. The experts pointed out that the tensions must lie between the arches rather than in the pillars sinking into the ground. They also asked for the force-directed algorithm to be reintroduced, even if it meant breaking the metaphor by automatically moving the nodes. Experts also noted the need for space around us to take full advantage of the graph without the nodes ending outside the room. Retrospectively, a proposition to solve this problem is a miniature version of the cathedral, similar to the first experiments of the cathedral metaphor, resolving the constraint of having to look up. It is assumed that experts did not raise this latter constraint due to the short time allowed with each graph.

The Island graph was less appreciated than the other two graphs. The island metaphor failed to immerse the user in the graph, making it less interesting than the web version. The other two graphs were better in this regard. In addition, the war and trade metaphor represented in the graph was too complex and required a lot of thought processes about opposition and association; the cathedral metaphor seemed to be a much better option. However, the experts did show interest in the animation of ships and cannonballs in the visualization. Adding animation to the cathedral or another metaphor closer to the desired meaning of the relationships may be a potential improvement.

As the work presented here is in an initial state, it is necessary to conduct a more thorough evaluation in the longer term. Our next objective will be to refine the cathedral metaphor before presenting it to the team's experts. Once we have identified a satisfactory version that meets their needs, we will present it to external users in the same way as we did for the web application.

8.5 General Discussion

Our preliminary results show that metaphors provide a fresh perspective and immersion compared to the visualizations offered on the web application. The cathedral metaphor, in particular, seemed to reinforce the notion of tension compared to the other two graphs. This means that metaphors seem to be an effective approach to address the second challenge discussed in Section 8.4. However, there is still a lot of work to be done to refine this aspect. Regarding the first challenge discussed in Section 8.4, the problem of size was not the main preoccupation of the experts, who seem to have come to terms with the current web solution. However, a more thorough evaluation is needed with an expensive graph to ensure these results.

In the context of an application aimed at a specific group of experts, immersive visualizations should focus more on highlighting underlying analysis aspects rather than trying to reproduce what already exists in 2D. The experts are interested in the tension and how to make the user feel it, rather than the position itself. The positioning of the nodes allows them to modify the behavior of the tensions, which must be automatic to assist the analyst. At this level, the experts are willing to add automatic behaviors that may not be part of the metaphor's foundation, such as the force-directed algorithm. This feedback goes against Wijayawardena et al. [2023]'s suggestion to use a metaphor rooted in the real world. This was done to emphasize the reconfiguration of the structure when a node is moved; moving an element modifies the structure itself.

Additionally, it appears that simply associating a structure with the graph is not enough. More visual support in the form of a 3D model that gives a tangible form to the structure is needed, as well as animation to support analysis. For instance, this can be done in the style of the first exploration of the cathedral metaphor. In a way, these feedbacks aligns with the last two notions mentioned by Wijayawardena et al. [2023]: aesthetic and animation.

Finally, what perhaps stood out most in this initial assessment was a notion underlying Weiser [1991]'s vision of ubiquitous computing, discussed in Chapter 2. The vision involves computing taking place in the real world through various terminals, such as smartphones, tablets, and computers. Each terminal has its own advantages and disadvantages, making them complementary in their use. This combination of terminals also reinforces the need to define what an AR application is. Indeed, in addition to the difference in definition between Pervasive Augmented Reality (PAR) and Conventional Augmented Reality (CAR), there is the question of whether or not the application is common to all terminals. It would, therefore, seem that the notion of AR application evolves along a continuum depending on how AR is used. In the particular context of this chapter, experts suggest that immersive technologies should not replicate what web applications already do well on computers, but rather complement them with new uses. For EFFaTA-MeM, this new use would be a collaborative space to explore the tensions in the text, where the web application will support the creation of this space.

8.6 Summary

This chapter discusses the advantages of using metaphors to define immersive visualization for text analysis in an existing web application. The web application primarily relied on graph-based text visualization, but the use of metaphors aimed to address two issues: first, the limited screen space on the web application, and second, the need to reinforce the concept of opposition and association that is central to the graph. Three different solutions based on different metaphors were explored, but they were not well-received by the experts on the project team. However, this initial evaluation revealed some interesting insights for developing cross-terminal applications and using metaphors in text analysis. First, mixing different types of displays has its advantages, as each type brings unique benefits and uses. Second, in this context, immersive AR is more helpful in providing a collaborative space for analysis and exploration by highlighting ways of thinking through adequate metaphors rather than reproducing visualizations that are already effective on a 2D screen. Of course, these preliminary results require further study in future iterations. Moreover, as they are based on only a few participants, they deserve to be reworked for presentation to a broader range of participants.

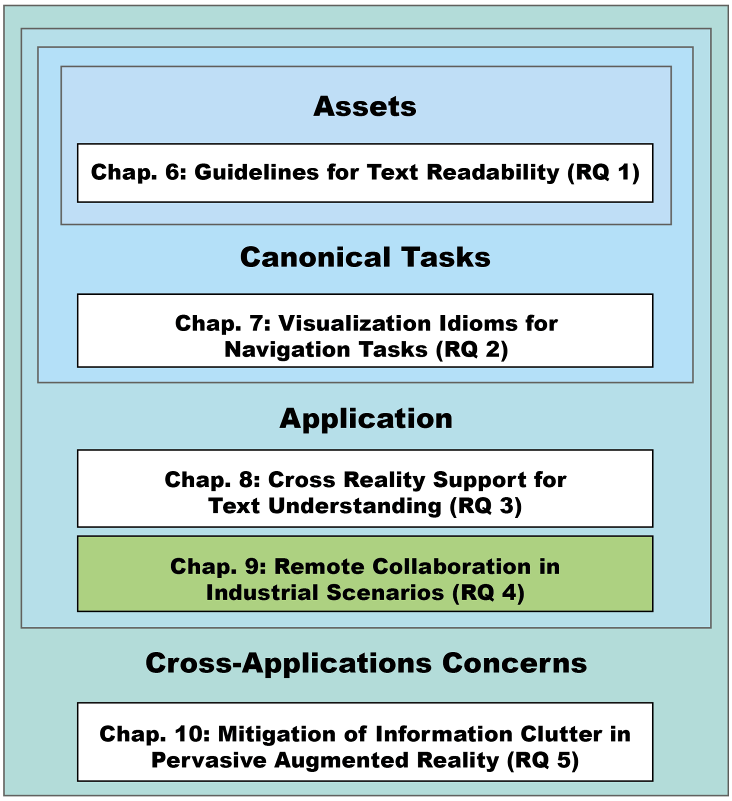
CHAPTER 9

REMOTE COLLABORATION IN INDUSTRIAL SCENARIOS

9.1 General Introduction

This chapter is dedicated to research question 4: **Which approaches best enable remote collaboration in industrial scenarios?**. Computer-Supported Collaborative Work (CSCW) is a field that explores the integration of technology in facilitating collaborative activities among individuals and groups. This domain fills, among other things, the geographical and temporal gaps faced by traditional face-to-face collaboration works. In the maintenance field, for example, this approach has been shown to reduce facilities and infrastructure costs and improve service quality [Sanchez et al., 2011]. Furthermore, the CSCW field combines with the Mixed Reality (MR) field (MR-CSCW) opens up new perspectives in terms of interactivity, presence, engagement and awareness [Sereno et al., 2022]. Wang et al. [2021] even observes that combining these research fields is becoming indispensable for remote collaboration on physical tasks.

Sereno et al. [2022] distinguished two research sub-branches, with studies on co-located and remote collaboration. This chapter focuses on the latter, with the basic scenario of one or more remote experts helping on-site workers solve a local task, such as a maintenance-like problem. The system must enable remote users to take ownership of the problem and the state of the workers' local environment [Rasmussen and Gronbak, 2019]. In addition, it must propose efficient communication channels to guide the on-site workers in the resolution process [Rasmussen and Gronbak, 2019]. This chapter's contribution is identifying current system limitations in industrial maintenance through an expert interview and analyzing cues to address these limitations.



Publications and supervised works

The content of this chapter is based on peer-reviewed publications in scientific conferences, my master's thesis, and one student work for which I was co-supervisor:

Cauz, M. and Cleve, A. (2019). Interacting with overlaid information in augmented reality systems for maintenance: A preliminary review. In **2019 13th International Conference on Research Challenges in Information Science (RCIS)**, pages 1–2

- This paper conducted a preliminary study on how augmented reality systems for industrial maintenance present and let users interact with virtual information. Ten papers are analyzed through three comparison tables presenting the type of information displayed, how the information is displayed concerning the real world, and the interaction techniques available to users.

Cauz, M. (2019). Augmented Reality visualization and interaction for maintenance. Master's thesis, University of Namur

- My thesis focuses on developing a design framework for immersive augmented reality applications. The framework is built as a succession of steps, to which designers must pay attention to correctly design their applications. The steps go from choosing the devices to addressing challenges at the application and cross-application levels.

André, M., Bayet, A., Jetzen, T., Luycx, P., Cauz, M., and Dumas, B. (2023). Engineering User Interfaces with Beat Gestures. In **Companion Proceedings of the 2023 ACM**

SIGCHI Symposium on Engineering Interactive Computing Systems, pages 76–78
- This paper explores using virtual hands to guide the user through precise gestures. The experiment occurs in a game where the user must follow the music rhythm. The results demonstrated that virtual hands are effectively an efficient way to guide users in specific manipulation. The work was first realized in a student project before being converted into a publication.

Thielemans, D. (2023). Virtualisation d'un manuel de maintenance industrielle en réalité augmentée. Master's thesis, University of Namur
- This master's thesis explores the conception of an immersive Augmented Reality (AR) application to guide a user in changing an expensive maintenance part. The primary focus of the work was the design of the application. However, results have highlighted the desire of some technicians to see this kind of application equipped with a remote collaboration function to reinforce the assistance provided during the procedure.

Outline

The organization of this chapter adheres to the subsequent structure. Section 9.2 discusses the results of an interview with a remote expert. Next, Section 9.3 presents relevant related works about the remote collaboration field. Then, an application prototype is designed in Section 9.4. The solution combines an AR headset for the on-site worker and a Virtual Reality (VR) headset for the remote user, including various cues to address the challenge exposed in Section 9.2. Finally, Section 9.5 concludes the chapter, as no prototype evaluation was carried out when writing this thesis.

9.2 Industrial Needs Elicitation

As part of the FLARACC project, we interviewed a remote expert to gain insights into the challenges they face in industrial maintenance. The interviewee always works remotely and never on-site, but his experience in this latter position is equally valuable due to his previous experience. He was a former fieldman before becoming the head of a remote help center for a defense department. Its job is to help military personnel solve any problems they encounter with their equipment. Communication can, therefore, take place from anywhere in the world, depending on the conflict in which soldiers are deployed. We aimed to supplement the existing literature with first-hand accounts of the difficulties encountered by remote experts. During the interview, we allowed the interviewee to introduce their business and discuss the solutions they had already implemented in real-life scenarios. Throughout the meeting, we asked questions to understand the issues at hand better.

The context in which the on-site worker is located is frequently a restricted area (a seat) with a weak network and where the situation's urgency can be critical to the worker's well-being. In addition to possibly being on the other side of the world, the worker on-site may not know the machine well enough to repair it on his own, hence the need for remote intervention. Intervention generally involves dismantling and reassembling various parts of the mechanism to resolve a fault. As a result, the on-site worker needs to keep his hands free while filming his work to continue to be guided by the remote expert. Moreover, the remote expert told us that the language barrier can also be strong even if English is used during the exchange due to poor reception and accents. The remote expert must also share content, such as 3D models, to illustrate his talk. On these 3D models, he can hide or display parts to show the inside of the machinery. Last, the remote expert must draw up a report at the end of the intervention as a record and documentation for future interventions.

In terms of solutions, he tested Vuzix smart glasses, which add a screen in front of the wearer's eyes. In this way, the remote expert can see the situation of the on-site worker and relay what he sees on the Vuzix's screen, with additional annotations such as drawings to frame a particular room. However, the problem is twofold. He cannot see anything on the remote expert's side because the worker constantly moves his head. On the site worker's side, there is a difference in focus between his two eyes that can quickly become distracting. In addition, the small screen size can make it difficult to perceive small details. The remote expert also added that ergonomics were not a success, notably because interactions are exclusively on the side of the glasses, with buttons and a trackpad that are not easy to manage on the first try. A solution via the smartphone was also tested but gave worse results, in addition to the constraint of putting the phone down to free the hands. Due to the restricted area, a solution with Spatial Augmented Reality (SAR) or Stationary AR is impossible.

At the end of the interview, we tested a solution based on the Hololens 2 immersive headset with the remote expert. This headset is an all-in-one Optical See-Through (OST) display, i.e., not connected to a computer. We made a call between the headset (on-site worker) and Microsoft's Teams application on the computer's desktop of the remote expert. The interviewee was very enthusiastic about this approach, mainly due to the high video quality compared with other approaches and the ease of use for both him and the on-site worker. It was also the first approach that allowed him to show his screen to share 3D models easily. All the content shared by the remote expert was displayed in 2D windows in the workers' environment. The annotations made by the remote expert on the video, on the other hand, were displayed directly in the worker's environment.

This first approach has enabled us to define a list of requirements. First, the remote expert can send documents but has no control over windows and 3D models. This means that it is not possible for him to assist the on-site worker in managing the workspace or placing the models in the right places. Second, the application remains relatively limited in terms of cues. Users can talk, exchange files, draw, and that is about it. There are no elements to support comprehension with subtitles, generate automatic reports, etc. Thirdly, the remote expert remains locked into the view of the on-site worker, with the disadvantage of the latter's head shaking.

9.3 Background

This section addresses the literature specific to the remote MR-CSCW field. The three psychological aspects that impact the design are presented in first [Sereno et al., 2022]. This is followed by discussing the various interfaces available to local and remote users. Then, the two existing strategies to immerse the remote expert in the local environment are presented. Finally, the different cues used to support the communication are described.

9.3.1 Psychological Aspects

Presence

Witmer and Singer [1998]'s definition of presence entails the psychological sense of being situated in an environment, even when one is not physically present, and requires both involvement and immersion. They refer to the involvement as "*a psychological state experienced as a consequence of focusing one's energy and attention on a coherent set of stimuli or meaningfully related activities and events*". On the other hand, immersion is described as "*a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences*". According to them, presence has a small positive impact on remote users' task performance. The sense of presence will depend on the degree and intuitiveness of the controls offered to the user, the number of sensory modalities, the consistency of the information between the sensory modalities, the isolation of the user in the virtual environment (VE), and the level of realism of the VE. However, excessive sense of presence in VEs or sudden and abrupt returning to reality can induce a dissociation from actual reality, resulting in behaviors that pose risks to individuals, others, and nearby objects. The proliferation of internet videos showcasing individuals inadvertently colliding with walls or damaging their televisions upon using VR headsets illustrates this phenomenon. This highlights the potential danger of losing touch with one's immediate surroundings due to an intense immersion in the virtual world.

Engagement

Attfield et al. [2011] define engagement as "*the emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource*". The authors identified several characteristics of engagement, including task-focused attention, positive affect (i.e., emotions experienced), aesthetics of the interface, endurability (i.e., enjoyable experience that a user wants to repeat), novelty, task richness and control, reputation, trust, expectation, and user context (e.g., motivation and benefits). The necessity to concentrate on a task can be related to the concept of presence, which is influenced by the user's sense of isolation. MR technologies have been shown to enhance many of these engagement characteristics, as discussed by Sereno et al. [2022]. These systems provide higher perceived user control and novelty levels due to enhanced sensory feedback and ubiquitous interfaces. Additionally, MR systems seamlessly blend virtual and real-world objects/tools, facilitating smooth focus transitions. Furthermore, AR systems stand out regarding social presence as they closely adhere to real-world rules and preserve social protocols, distinguishing them from regular workstations and VEs. This higher social presence may increase engagement in AR experiences.

Awareness

According to Gutwin and Greenberg [2002], previous researchers had defined awareness as "*knowledge created through interaction between an agent and its environment - in simple terms, knowing what is going on*". This knowledge is defined in time and space, necessitating updates. They defined a Placement×Presentation matrix to characterize the awareness. The first dimension discerns whether the information is displayed in the workspace (situated) or externally (separate). The second dimension determines whether the information is displayed in its raw manifestation (literal) or a coded configuration (symbolic). The authors advocate for adopting the situated-literal approach, deeming it the most natural and effective. This approach entails presenting information to users consistent with how it is presented to other users within their specific contexts. Nonetheless, the authors acknowledge the potential challenge of users scanning the entire environment to access information. To resolve this challenge, they propose incorporating techniques such as those discussed in Chapter 7 on the navigation to offer users a comprehensive overview of the environment. In addition, they mentioned the critical elements of the embodiment, which is the representation of the users in the environment.

9.3.2 Interfaces

In their recent survey dedicated to remote collaboration on physical tasks, Wang et al. [2021] identified and categorized four main types of devices that remote or local users use. The first category, referred to as Windows-Icon-Menu-Pointer (WIMP), encompasses traditional 2D computing displayed on a screen with interactive functionalities through a mouse and keyboard. It has the advantage of being easy to set up and has well-established interaction paradigms, which have been extensively studied throughout the modern computing era. However, the authors noted a significant concern when used by on-site workers: it implies users divide their attention between the screen and the physical task. The second category, the SAR interfaces, allows instructions to be projected directly into the local user's environment but is limited by a fixed projection location. As a result, SAR is less suitable for medium and large environments and remote users. The third category, the Hand-Held Display (HHD), enables users to maintain an augmented view of the environment while breaking away from fixed information positions. Nonetheless, using HHDs comes at the cost of restricting the user's hands. Lastly, the Head-Mounted Display (HMD) category blurs the boundary between the real and virtual environments while freeing the user's hands. However, the current inconvenience associated with HMDs includes their size, weight, and social acceptance.

Regarding local users, the prevailing trend is to equip them with AR tools that offer a heightened level of immersion compared to conventional screens displaying augmented environments. The objective is to present information promptly and, whenever possible, in the relevant spatial context. However, insights from consultations with an expert on a hotline from the FLARACC project indicate that HHDs should be avoided, first to free up the workers' hands, and second to address the issue of excessively jerky video streams, which hinder effective assimilation of the augmented scene and could potentially cause discomfort for the remote expert. HMDs share similar concerns, as the video feed aligns with the local user's viewpoint.

Nonetheless, according to the interviewed expert, these issues are considerably less pronounced with HMDs compared to HHDs. Especially as the effect will be more or less intense depending on the headset and image stabilization, according to the arguments in favor or not of the different categories of interface, it would seem that SAR interfaces should be favored when space is limited and can be covered by a projection system (e.g., a beamer). Otherwise, when the system is intended for short-term use, HMDs may be the preferred choice. For extended usage periods, it is advisable to favor HHDs or WIMP interfaces. In all cases, the rapid advancement of HMD technology in the coming years is anticipated to lead to HMDs becoming the favored interface across all contexts.

As for remote users, the conventional method typically involves using WIMP interfaces. Nonetheless, a few research studies have explored SAR, HHD or HMD approaches for remote users [Wang et al., 2021]. However, rather than pushing AR to the side of remote users, recent studies have focused instead on pushing the use of VR, which offers a critical advantage to AR within such a context: the presence, the engagement and the awareness of the local space. The VR headset can be an efficient solution as it maximizes immersion, involvement, and engagement by isolating the user in the local environment. Additionally, suppose the information received on the environment allows it. In that case, the user can locate himself and move in 3D space, increasing their awareness either freely within a reconstruction of the environment or at key positions corresponding to cameras. This freedom can also enable greater control over the virtual elements they share with the local operator. As Munzner [2015] and Fuchs et al. [2014] point out, it's easier to experience 3D visualizations in 3D rather than in a 2D space such as a screen. Yet, 3D models have become a common element and are even the most recurrent interface element, according to Gattullo et al. [2022]. Consequently, the future of interfaces for remote users appears to be oriented towards VR headsets once the problems of size, weight, motion sickness, etc., will be resolved. If VR is not possible or not desired, then WIMP interfaces are preferred, or even HHDs if the user needs to be reachable anywhere [Rasmussen and Gronbak, 2019].

9.3.3 Access to the Local Environment

One of the underlying principles within the domain of MR-CSCW involves facilitating the cognitive awareness of distant users, encompassing not solely the task itself but also the contextual environment within which the task is situated. This objective is achieved by exploring two distinct yet not mutually exclusive categories of methodologies [Fuchs et al., 2014]: the utilization of real-time video streams and the 3D reconstruction of the spatial environment.

Real-Time Video Stream

The utilization of video streaming, which is the most time-honored approach, constitutes a fundamental element in numerous applications involving remote communication. The archetypal strategy entails transmitting the visual perspective of the local user, which offers the advantage of shared focus, thereby facilitating a deeper comprehension of the local user's intentions. Nonetheless, this approach curtails the degree of autonomy for the remote user to navigate through the environment, effectively relegating them to the role of a spectator rather than a collaborative participant. Tait and Billinghurst [2015] determined that augmenting the level of independence in the remote user's viewpoint resulted in expedited task completion, heightened confidence, and reduced verbal exchanges. Subsequently, Lee et al. [2017b] examined the selection between dependent and independent views in an application that used AR for the on-site worker and VR for the remote user. They employed a 360-degree video stream, allowing the remote user to explore the environment unrestrictedly. The findings strongly favored the independent mode, particularly considering the presence of visual cues, such as arrows and frames, that indicate the users' focus. Piumsomboon et al. [2019b] named the approach of an independent view for the remote user the "Giant-Miniature Collaboration" metaphor. This analogy casts the on-site worker as a giant and the remote user as a miniature entity whose size correlates with the camera's positioning. For instance, if the camera is at ground level, the remote user assumes the scale of the camera. In conclusion of their research, they advocate granting the on-site worker the capability to summon the remote user into their field of view as needed. Another proposition by Rasmussen et al. [2022] extended to the remote user the option to switch between different camera perspectives to attain diverse viewpoints of the scene. Veas et al. [2012], as for them, recommend providing access to the video stream to the local user. In a divergent approach, Stafford et al. [2006] introduced a top-down view using a metaphor of a god-like perspective. In this depiction, the local user perceives the "hand of god" (i.e., the remote user) extending from the heavens to indicate elements within the environment. Finally, regarding camera positioning, it can be either at a location in the environment, which may be moved by the local user or positioned on the local user's body. Piumsomboon et al. [2019b] recommend a protruded shoulder-worn position rather than a backpack, head, or handheld position. This position allows seeing the environment and the collaborator's face at the same level as the latter's eyes. It is also unaffected by hand and head movements.

3D Reconstruction of Task Environment

The predominant approach to reconstructing the task environment often involves utilizing the output from one or more cameras to reassemble the environment's spatial layout [Gao et al., 2017; Gauglitz et al., 2014b; Orts-Escolano et al., 2016]. It is also possible to use a pre-created environment model as a basis or directly, but at the risk of inconsistency with reality. Nevertheless, since the assorted methodologies for such reconstitution are beyond the scope of this thesis, an in-depth examination of these techniques is precluded. Notwithstanding, empirical findings by Kleiber et al. [2012] indicated superior outcomes when utilizing a VR system instead of a conventional video setup. Furthermore, an investigation by Teo et al. [2019] juxtaposed the two methodologies (i.e., video and reconstruction) and concluded that proposing both might constitute an optimal strategy. A salient advantage of employing 3D reconstitution instead of video lies in its capacity to afford users unrestricted movement within the environment. However, its effectiveness depends on the textures' realism and accentuation of the cues' difficulty in ascertaining the users' positioning and focus.

9.3.4 Exchange of Information

The process of exchanging information involves the transmission of messages among multiple individuals, which can be deconstructed into distinct components: 55% is attributed to non-verbal cues, 38% to variations in intonation, and 7% to explicit verbal content, as outlined by Bioy and Bourgeois [2003]. Consequently, a mere 45% of the conveyed message can be comprehended in scenarios where only audio is available. Conversely, incorporating video allows for observing immediate surroundings, as discussed in the previous section, and facilitates the complete perception of the message. Non-verbal communication encompasses facial expressions, eye contact, and gestures, constituting a set of variables that can be manipulated to enhance collaborative interactions. However, a notable challenge in capturing user video is the inherent system focus on recording the operational context rather than the users themselves. This is particularly evident when a local operator employs a HHD or a HMD to provide the remote user's exclusive visual perspective. Another scenario involves a remote user wearing a VR HMD that obscures the facial region. Additionally, enabling the remote user to move freely in the environment or adopt various camera viewpoints, as mentioned earlier, introduces the issue of constantly tracking the user's position. This crucial information helps the user better understand the partner's focus and message. Moreover, certain gestures lose their significance when reproduced on a 2D screen due to the challenge of mapping between the 2D screen and the 3D environment. For instance, pointing at an object becomes less effective. Researchers have explored five distinct methodologies documented in the existing literature: spatialized voice, avatar representation, frustum-based visualization, gaze tracking, and virtual hands. They complement annotation techniques such as element pointing and drawing.

Spatialized Voice

The study by Yang et al. [2020] evaluated the effects of using spatialized voice on collaboration effectiveness and spatial awareness. While spatialized voice enhanced spatial awareness, discernible improvements were not observed in collaboration effectiveness when vocal localization was absent. Interestingly, spatialized verbal cues could even lead to distraction due to audio volume fluctuations. Therefore, researchers suggested that incorporating spatial audio with minimal volume could supplement visual indicators like frustum and manual gestures, enhancing the sense of social presence. Consistent with these findings, Rasmussen et al. [2022] obtained similar results, affirming the effectiveness of visual cues over audio cues. Günther et al. [2018], as for them, recommend using tactile and audio cues as auxiliary to the visual cues.

Frustum-Based Visualization and Avatar

Concerning frustum-based visualization, their utility has been substantiated by previous research [Chénéchal et al., 2015; Piumsomboon et al., 2019b; Yang et al., 2020]. Users' frustum is the idiom used to indicate the part of the environment they can see. The most accurate approach involves delineating a frame in the user's field of view within the environment. Unlike gaze, the boundary of the field of view is visible. Additionally, an avatar can accompany the frustum at the user's position. If the user's face is trackable, a real-time video window of the user can be placed there [Billinghurst et al., 1998]; otherwise, a virtual avatar represents the user. The complexity of the avatar varies based on the study, ranging from a basic head model [Teo et al., 2019] to a highly realistic representation [Joachimczak et al., 2017; Lincoln et al., 2009; Maimone et al., 2013]. To indicate the position of other users' avatars/frustums when they are out of a user's view, arrows [Lee et al., 2017b] or gaze color [Rasmussen et al., 2022] (i.e., the gaze takes the color associated with the camera) can be used. However, the gaze color method has limitations in environments with numerous positions and free user movement. Another approach involves using a mini-avatar to show gaze direction and the position of life-size avatars [Piumsomboon et al., 2018]. Their findings demonstrated that besides offering good performance, the mini-avatar reduces the need to look at the life-size avatar.

Gaze Tracking and Pointing Technique

Utilizing WIMP interfaces enables users to employ interactive actions, such as clicking on the visual data stream to designate specific elements within a 3D spatial context. Notably, Gupta et al. [2016] observed that amalgamating this methodology with gaze tracking noticeably reduces the time required to accomplish the designated task. Conversely, Piumsomboon et al. [2019a] identified a preference for employing head gaze rather than eye gaze. Concurrent investigations [Gupta et al., 2016; Higuch et al., 2016; Lee et al., 2017a; Piumsomboon et al., 2019a] further substantiated the efficacy of utilizing gaze indication in synergy with the frustum technique. An alternative approach involves employing the index finger as a pointing tool, which may also serve as a viable solution, circumventing the need to immobilize head movement [Teo et al., 2018].

Virtual Hands and Drawing

As discussed above, gestures are significant in communication, complementing verbal content, particularly in intricate explications. Consequently, the apprehension of these gestures through a two-dimensional screen constitutes a viable resolution [André et al., 2023; Teo et al., 2018]. Nonetheless, empirical evidence indicates that their efficacy is further heightened when conveyed through AR or VR mediums [Chenechal et al., 2016; Chénéchal et al., 2015]. This augmentation allows contextualizing gestures within their spatial context. For example, simulate the rotation of a bolt by replicating the motion of twisting one's finger in front of it.

In addition, Teo et al. [2018], Huang et al. [2018], and Kim et al. [2019b] have demonstrated a significant interest in enabling users to engage in drawing activities within VE, particularly when employing virtual hands. However, challenges arise under specific circumstances. When the user's perspective is affixed to another user's viewpoint, capturing a static image (screenshot) becomes essential to facilitate accurate drawing. Another complication arises when a user's view is contingent upon a real-time video stream; in such cases, incorporating a drawn element into the environment necessitates accounting for depth considerations. Gauglitz et al. [2014a] investigated this predicament and proposed three distinct approaches. The first approach involves spray points: independently determining the depth of each point by projecting it onto the surface located behind the point. The second approach entails projecting all points onto a plane orthogonal to the viewing direction, situated at a depth defined by the application's designer. The third approach involves identifying the dominant plane within the 3D space formed by the 2D points and its associated depth, subsequently projecting the drawn shape onto this plane.

Annotations and Virtual Replica

The last cues investigated within the literature pertain to annotations and virtual replicas. Annotations can take many forms, but in an industrial maintenance context, we will mainly find icons [Cauz, 2019; Cauz and Cleve, 2019; Thielemans, 2023]. At the same time, an explanatory text will be provided on a 2D window or via the presence of a collaborator [Thielemans, 2023]. While both Günther et al. [2018] and Kim et al. [2014] observed interest in the former cue, a pointing technique somewhat overshadowed its popularity, likely attributed to its relatively faster nature. On the other hand, the latter stimulus, a virtual replica, underwent assessment and recommendation by Oda et al. [2015] in a co-located collaboration context. However, the insights garnered from their study hold potential applicability in the context of a remote collaboration.

9.4 Prototype

We designed an application prototype to address the challenge discussed in Section 9.2. First, the remote expert needs access to the on-site worker's video without blocking the latter's hands. Second, the remote expert must be able to share information such as images and 3D models and manipulate them. Third, the solution must support verbal language, which can sometimes be complicated by accents or poor language knowledge. Last, a report of the intervention must be drawn up automatically to facilitate the work of the remote expert after the intervention. All these challenges must be addressed, considering the on-site worker's space is restricted to a seat.

In this prototype, we aim to analyze two important aspects. First, we want to explore how we can integrate the various cues presented in the previous section to achieve a certain consistency within the application. We also want to determine whether these cues duplicate each other or if users prefer one cue over another. Second, we want to investigate how to remove the spatial barrier between users by allowing them to control each other's environment directly. Our hypothesis is that this last element can significantly improve communication between them.

Unfortunately, evaluation of the prototype has not yet taken place at the time of writing. Therefore, this section presents the solution that will be evaluated in future work.

9.4.1 Devices and Views

Our application combines the advantages of OST HMD on the on-site worker side and Video See-Through (VST) HMD on the remote expert side. The latter had access

(a) (b)

Figure 9.1: Illustrations of the prototype: (a) the hand menu and (b) a remote expert pointing to a phone.

to four views: the worker's view, the expert's real view, a 360° view thanks to a movable camera on the worker's side, and a virtual view focused on the repair machine (i.e., where there is only the remote expert and the repair machine). They can switch between the views by clicking on virtual buttons that appear when you look at the palm of your hand. We propose all these views to give the expert the freedom to navigate the environment. An indication is always displayed to the worker on-site to let him know which view the expert is on. In addition, in the 360° view, an indication is displayed at the camera level to show which direction the expert is looking. Finally, the on-site worker can ask the expert to switch back to their view by clicking on a button when looking at the palm of their hand. At this point, a dialog box will appear on the remote expert's side, giving the choice of accepting.

9.4.2 Hands, Gaze and Draws

To help users indicate certain elements more precisely in the environment and to support verbal language, they will be able to see each other's hands. Via the latter, they can also point with their index finger (a line will run from the index finger to the object pointed at) (Figure 9.1b) or draw in the air. For the remote expert, drawing will be done on a frozen frame to make the task easier. Pointing with the index finger is done by closing the hand except for the index finger. On the other hand, drawing is done by clicking on the option in the menu accessible from the palm of the hand.

9.4.3 Windows and 3D models

To enhance user collaboration, we decided to make all windows and 3D models visible on both sides, except for the menu associated with the palm of the hand (Figure 9.1a). In other words, the virtual space is shared; if the on-site worker opens a window, it also appears on the remote expert side, and vice versa. However, for the sake of privacy, the window's content can be shared via a button at the top of the window (Figure 9.2). Thus, if the content is not shared, the correspondent will only see a blue window, which can be moved if required. 3D models, on the other hand, cannot be hidden from the correspondent. There are two approaches to accessing

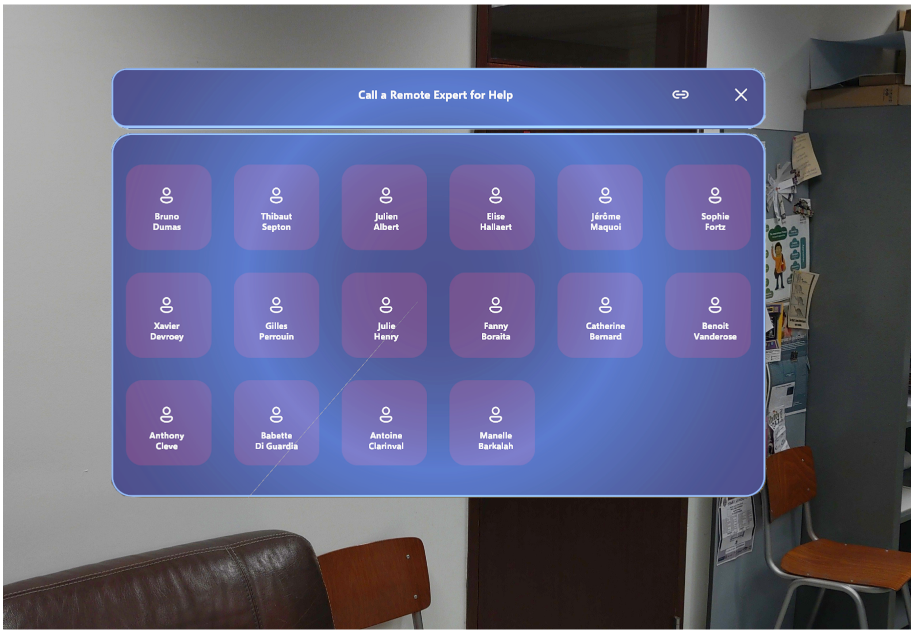


Figure 9.2: Example of a window in the prototype. The button with the link symbol next to the cross allows showing/hiding the window's content to the collaborator.

3D models. The first is to select the desired part in a window and make it appear in the air in front of the users. The second is for the remote expert to go into the virtual view, where they can make the part appear directly superimposed on the real part. For the on-site worker, simply click directly on the real part to select the virtual part to place.

9.4.4 Subtitles and Report

We have set up an automatic subtitling system for discussions. These subtitles are positioned at the bottom of the user view, with each user seeing the other's subtitles. These subtitles are also saved in a file with the time codes. In addition, all notable actions carried out during the intervention are recorded.

9.5 Summary

The interview with a remote expert and a literature review showed us that there is still work to be done in immersive MR collaboration. The literature on the subject has focused on improving non-verbal communication, presence, engagement, and awareness. In a solution that has unfortunately not yet been tested, we propose to add a fully shared virtual space and automatic report generation based on verbal communication and key actions performed by both users. The next step in this research is, of course, to evaluate this prototype.

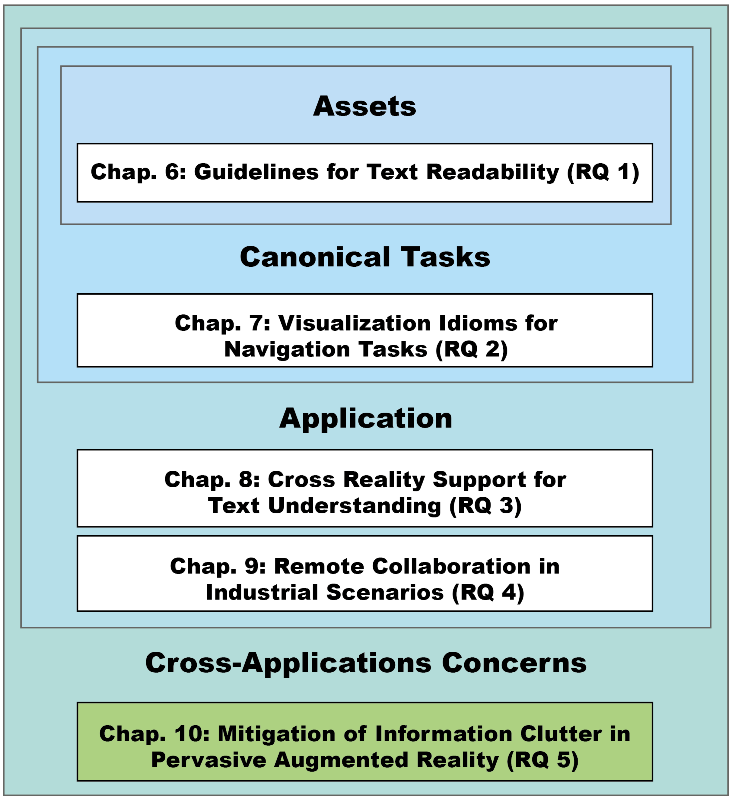
CHAPTER 10

MITIGATION OF INFORMATION CLUTTER IN PERVASIVE AUGMENTED REALITY

10.1 General Introduction

This chapter is dedicated to research question 5: **How can the issue of information clutter be mitigated through active control from the user in Pervasive Augmented Reality scenarios?**. Information clutter is a persistent issue in computer applications, worsened by the amount of information displayed, screen size limitations, and visual obstructions. In Augmented Reality (AR), the challenge is even more significant due to the uncontrollable real-world environment unless the application is used in a confined setting. The problem of information clutter can quickly arise when developing complex AR applications, especially when combining multiple applications like in Pervasive Augmented Reality (PAR). This chapter specifically addresses the issue of information clutter in this latter context.

As elucidated in Chapter 2, Grubert et al. [2017] define PAR as an AR usage paradigm where various applications coexist simultaneously and continuously, never closing. In such dynamic environments, reliance on context-aware solutions becomes imperative, complementing sporadic manual interactions to prevent users from expending excessive effort managing the multitude of open applications. In science fiction, various media productions have explored futuristic scenarios, both dystopian and utopian. A particularly thought-provoking work is Matsuda [2016]'s "Hyper-Reality" video, which portrays a world where AR permeates every facet of daily life, including advertising, entertainment, and tourism (See Figure 10.1). Matsuda's work raises important questions, including one relating to the novel concept of situated information. Situated information denotes data visualizations closely associated with their physical counterparts [Prouzeau et al., 2020]. In a world resembling Matsuda's vision, individuals could generate situated information to



be shared and perceived by diverse users, encompassing professionals, artists, and public and private entities. However, the continuous influx of such information from various sources can lead to visualizations obscuring critical real-world elements, resulting in obtrusiveness and overwhelming experiences, as noted by Lu [2021].

Consequently, this chapter contributes to AR knowledge in several ways. First, it proposes, assesses, and refines various strategies to empower users with control over situated information flow. Second, it utilizes the selected use case to conduct experiments, shedding light on user behavior and cognitive load within a PAR environment characterized by multiple information producers. These insights are poised to pave the path for the broader and safer adoption of pervasive augmented reality throughout society.

Publications

The content of this chapter is based on the following peer-reviewed publications in scientific conferences:

Cauz, M., Septon, T., and Dumas, B. (2023). Interaction Techniques to Control Information Clutter in a Pervasive Augmented Reality Scenario. In **Proceedings of the Internation Conference on Human-Computer Interaction (HCI) - Late Breaking Papers: Virtual, Augmented and Mixed Reality**, pages 3–21

Figure 10.1: Illustration of an augmented street by Matsuda [2016] in his "Hyper Reality" video.

- This paper presents the pilot study and the first experiment conducted in this chapter. In particular, it presents and evaluates the initial four strategies for returning control of the information flow to the user.

Septon, T., Cauz, M., and Dumas, B. (2023). Extending the World-In-Miniature Metaphor to Access Situated Information in a Pervasive Augmented Reality Environment. In **IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 543–548

- This paper presents the second experiment conducted in this chapter and designed following the lessons learned from the first experiment. It consists of presenting and evaluating a new strategy, combining World-in-Miniature (WiM) and a filtering mechanism to access information without moving to their location.

Outline

The organization of this chapter adheres to the subsequent structure. Section 10.2 presents relevant works evolving in a context of PAR. Next, Section 10.3 presents interviews to highlight the problems behind the augmented street illustrated by Matsuda [2016]. Then, Section 10.5 explores the effectiveness of four user-centered strategies attempting to solve the identified problems. Section 10.6 explores an exciting approach to accessing situated information without moving, highlighting the feedback of the first study. Then, Section 10.7, in the process of iteration, shows the evolution of strategies and the emergence of new ones. Unfortunately, these have not yet been tested at the time of writing. Finally, in Section 10.8, we discuss the implications of our results, limitations, and future works before concluding in Section 10.9.

10.2 Background

The emergence of PAR inside our society raises ethical, social, legal, political, and technical issues. Gugenheimer et al. [2020] organized a workshop on ethical, social, and political impacts in Mixed Reality (MR) research. Unfortunately, the workshop was canceled due to the pandemic. Regenbrecht et al. [2022] opened the discussion on the privacy, safety, belief, and rights applied to the PAR. What stands out is a careful modulation of what the user sees, proposing a control for managing the number of virtual elements displayed and defining social and legal norms around the subject. For their part, O'Hagan et al. [2023] listed a series of ethical issues linked to the PAR. There is a risk to the user's identity, which could be subject to a desire to conform to ideals, be stolen (e.g., via deep fakes), or be ridiculed by other users. It will also take time for this technology to be democratized for everyone, which will also cause inequalities among people. Another risk is the alteration of the perception of reality, which can lead to an unproven ideological representation by hiding inconvenient elements (poverty, people with divergent political opinions, etc.) or even lead to persuasion and manipulation by third parties. Regarding the actual state of ethical and legal scope on these subjects, the authors say *"[...] at present, it would seem that existing digital human rights do not sufficiently address the exposed societal vulnerabilities of everyday AR."*. To this end, they posed a series of questions to help establish this ethical and legal framework: *"What elements in reality are permissible to alter, augment, extend or diminish?"*, *"Who can mediate your perception of reality?"*, and *"How this capacity for mediated perception will be exploited to enact and amplify abuse?"*.

Besides ethical work done in the context of PAR, other works focus more on view management techniques. Lu and Bowman [2021] explore how AR applications can be presented and interacted with while being unobtrusive and with minimal cognitive effort. Their solution is called glanceable applications (i.e., applications such as Email, Calendar, or Fitness minimized inside users' peripheral vision) with which users can use gaze to expand their content. Their goal was, on the one hand, to compare applications with their equivalent in conventional devices (e.g., smartphones or computers) and, on the other hand, to study user behaviors in this pervasive environment. Their results demonstrate interest in the use of pervasive applications. Furthermore, the experiment participants desired more interactions than just information access. Also, they wanted to be able to turn on and off the AR to be relieved from the pressure of the digital world. Lindlbauer et al. [2019] explore the view management problem of having multiple applications opened as windows for continuous use of AR. They submit an optimization-based approach to determining where, when, and how applications should be displayed using interfaces with multiple levels of detail according to the user context and task. Others have also used context awareness as a way to determine what and how information should be displayed, such as Lu and Xu [2022], Caggianese et al. [2015], Julier et al. [2000], Lee et al. [2008], Gebhardt et al. [2019], and Orlosky et al. [2015]. However, reducing information clutter in the virtual environment may help identify near-user information but won't help find results for hidden or distant information. Besides, automatic techniques are based on heuristics or learning models. Although they try to determine the user's wants, they are never an exact model of their thoughts and reflections. Manual techniques include highlighting the desired content or hiding the not desired content, as discussed in Chapter 7. Another PAR related work was realized by Marques et al. [2019]. In it, the authors test research and manipulation tasks in a PAR environment. They help the user in their task thanks to three navigation aids: (1) increasing/decreasing the element's size over time, (2) a 3D frame around the element, and (3) arrows pointing to the element. Results showed good participant appreciation and interest in diverse daily life activities. The navigation aids were discussed in greater detail in Chapter 7. If we disregard how users enter their search, they help guide them to the various points of interest. All techniques have advantages and drawbacks, but the precise location is generally not indicated until the user has moved to the target.

MacIntyre et al. [2011] defined multiple requirements that must be met for a PAR ecosystem to exist. These are the following: "*multiple sources (or channels of interactive information) must be able to be simultaneously displayed and interacted with, channels must be isolated from each other (for security and stability), channel authors must have the flexibility to design the content and interactivity of their channel, and the application must fluidly integrate with the ever-growing cloud of systems and services that define our digital lives.*". To our knowledge, no other work addresses issues arising from multiple content producers in a PAR environment. Most works focus on applications designed as we know them today on smartphones and do not fully exploit the potential of AR. Except for MacIntyre et al. [2011], none of the works above explicitly address the problems that emerge from a world where multiple sources of information coexist. Most view management techniques solve information clutter using context-aware approaches, mainly for smartphone-like applications. In this chapter, we would like to proceed otherwise by first establishing interaction strategies that can be applied to let the user keep control over a PAR environment.

As we will see, most works focus on applications designed as we know them today on our smartphones or studied view management techniques, solving information clutter using context-aware approaches. Few works explicitly address the problems that emerge from a world where multiple sources of information coexist. In this chapter, we would like to proceed otherwise by first trying to establish user-centered interaction strategies that can be applied to let the user keep control over a PAR environment constructed by considering the interests of different content producers. It is essential to reduce the frustration and fears caused by unexpected changes by ensuring user control. PAR environments need to stay transparent, with the user's agreement and acknowledgment, by mixing automated and interactive approaches.

10.3 Pilot Study

As a first step, we conducted semi-structured interviews to understand better potential problems conveyed by the illustration of the street in the "Hyper-Reality" video of Matsuda [2016]. This section was part of a scientific publication [Cauz et al., 2023]. We asked the participants to imagine themselves as pedestrians in an unknown augmented city (i.e., all streets are augmented like in Figure 10.1). They were told to be visiting the city or simply wanting to reach a specific place. Participants knew they wore augmented glasses or contact lenses displaying virtual information.

We hypothesized that the high quantity of information, animations, and the weak coherence of the environment would make information irrelevant and the world oppressive, thus increasing cognitive load. To verify that, we formulated multiple research questions: (1) "How do users feel in this type of environment?", (2) "Is the separation between what is real and what is virtual clear?" and (3) "Is the virtual information understood and relevant?".

10.3.1 Participants

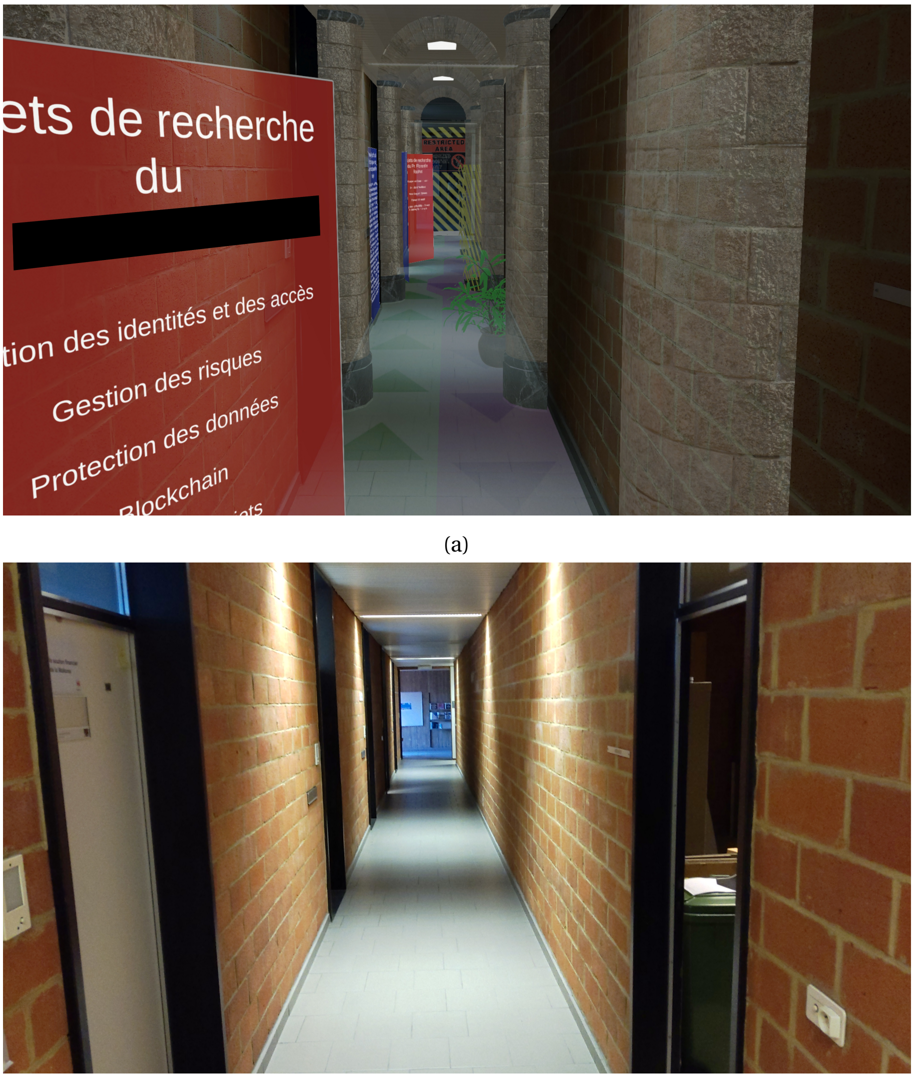
A total of five men and four women took part in the interviews. They were all volunteers; no remuneration was given. Five participants were between 25 and 30 years old, two were between 35 and 45, and two were 60. All participants had a master's degree, including four with a PhD in computer science. However, no participants came from our laboratory team, and none worked in the AR field. Apart from that, their background was heterogeneous.

10.3.2 Protocol

The meeting was not audio or video recorded; our team members took notes. When two interviewers were present, they were both allowed to ask questions, but one had the lead. The interviewees began by giving their agreements for the interview and the use of their answers. After that, their personal information described in the previous section was collected, and the context was exposed. Next, the photo in Figure 10.1 was presented to the participants. They were requested to immerse themselves in the depicted environment. The first question stated above was formulated. The interviewees were free to speak and never interrupted. The interviewers asked questions only for more details on a remark or to redirect the discussion through one of the research questions of the interview. When the participants said everything they had in mind and the interviewers had no more questions, the corresponding video part of the street was shown; the question asked was: "Does the video change anything to your feelings?". The same procedure as the one for the photo was applied until the end of the interview. Beginning with a photo was a deliberate choice to help the participants analyze the details of the scene.

10.3.3 Outcomes

Results indicate that valuable information is lost in the middle of the advertising and could become useless due to information clutter. On the other hand, some participants mentioned the similarities of such a street with cities like Singapore or New York, noting that the difference with today's big cities is the possibility offered by AR to modify our perception of reality. According to six participants, the perception of what is real is deteriorated, and three participants feared for their physical safety, again due to information clutter. Other observations not related to information clutter are the following. While some participants described a feeling of sickness, at least all agreed that this environment was tiring. Two participants working in the medical field also noted that the environment, as presented, could greatly affect the population's health either by "*driving people crazy*" or due to medical syndromes. To conclude, we want to highlight a remark from one of the participants: "*This is about*



(a)

(b)

Figure 10.2: Chosen floor for the experiment, with (a) and without (b) the virtual content. Note that the virtual elements appear more transparent in the capture than they appear to users.

the future of cities, not the countryside.". This raises the question of "what will be the limit of the intrusive advertising in front of these free spaces?".

10.4 Street-Like Pervasive Augmented Environment

This section, extracted from Cauz et al. [2023], presents the environment used for the following experimentation. We realized the experiments on a building floor to ensure a controlled environment. We used its corridor as the principal place for the experiments, mapping it to the vision of Matsuda [2016] of an augmented street. We matched the corridor to the street, the doors to shop windows, etc. Realizing the experiment in such an environment allowed us to keep control over weather conditions, passers-by, privacy, and social acceptance. It also enabled us to fully exploit the Microsoft Hololens 2's potential. Such headsets, being Optical See-Through (OST) devices, are vulnerable to light variation; therefore, controlling the light exposition of the environment solves the issue. In addition, OST devices have the advantage of reducing motion sickness and are less intrusive than other types of AR headsets. However, it should be noted that the Hololens 2 suffers from a restrained field of view; therefore, users do not perceive virtual content inside their peripheral vision.

To augment the floor, we designed five categories of virtual elements: 1) decorations, 2) social norms, 3) navigation indications, 4) room information, and 5) advertising. The work of Matsuda [2016] inspired all these categories as illustrated in Figure 10.1. Figure 10.3 shows the placement of the virtual elements on top of the physical layout of the floor. However, virtual pillars and roof, as seen in Figure 10.2, have not been represented for clarity reasons. We added pillars and a false roof for the decoration category to break the corridor's square red brick look. While the virtual roof is higher than the real one, it is still visible and hides the physical roof. In addition to these architectural modifications, we added plants at multiple locations and flying paper planes in the corridor. Finally, we added a chirping bird sound in the background. For social norms, we consider any information considered to be cultural, behavioral guidelines, and expectations. In these experiments, this is expressed with animated arrows on the ground indicating which side to walk on depending on the direction you want to go. These are not to be confused with navigation information. Those are either wayfinding signs indicating classroom names or other visual elements, such as the usual toilet sign on the dedicated doors, a bucket, and a broom indicating the storage facility. We also placed a panel on the cafeteria entry to indicate that the access is for staff members only. Two other panels denying access were placed to prevent the user from entering a room always opened and accessing the end of the corridor due to a lack of luminosity that makes the headset behave unexpectedly. For room indications, we defined content depending on the use of the room. For professors' offices, their biographies, agendas, last published papers, or research domains were added as separate panels. We also added audio content where the professors share their research area to attract passers-by. We added panels with agendas and background noise for classrooms to give the impression that the rooms were occupied. On the student open space doors, we placed panels with information on the usage of the room, rules of use, and, if the room was named after a person, the biography of the corresponding personality. For advertising, we added four bulletin boards (see Figure 10.4f) with advertisements coming from students, faculty, or university activities. Each panel is a square that takes the wall height. It rotates slightly towards the user to be more readable. The panels' content updates every ten seconds, and seven ads, on average, are presented simultaneously. We also added a teddy bear on one end of the corridor, moving up and down. When looked at for more than 0.8 seconds, an advertisement appears progressively towards the participant. Finally, in addition to these elements, we

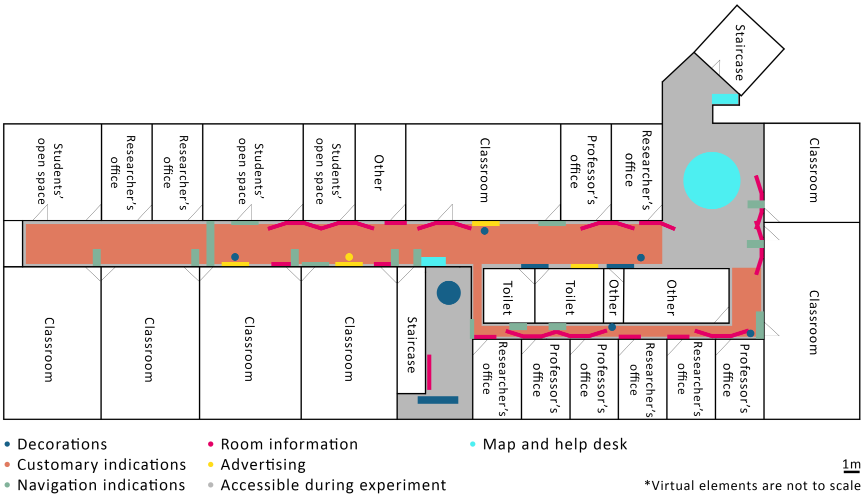


Figure 10.3: Layout of the chosen floor.

added two maps on opposite sides of the floor and a help desk at the floor entry near where the experiment would begin. The help desk holds a penguin in its center, and the penguin asks the participants if they need any help when they get close to it.

10.5 First Experiment: Initial Strategies

Through this first experiment, we study the impact on the cognitive load of four strategies to solve the problem of information clutter highlighted in the pilot study (see Section 10.3). This section is the main part of our publication [Cauz et al., 2023]. By applying these four strategies, we study their impact on the user's cognitive load while evaluating their implementation to determine how to improve them, should they be effective.

10.5.1 Strategies

Based on the results of the Pilot study presented in Section 10.3, we designed four strategies to help users control their environment. Grubert et al. [2017] argue that PAR environments need context-aware automation to reduce manual user interactions. Although we agree, context-aware approaches must support and not replace user interactions. Therefore, users can keep control over the system. For this reason, we propose three strategies that are said to be active, meaning they make use of interactivity and solicit user actions, and one passive. These strategies are foundations on which context-aware automation could be applied. In an attempt to make the techniques accessible to the general user, we developed a User Interface (UI) that simulates a smartphone (see Figure 10.4d). That virtual smartphone gives the user access to the third and fourth strategies. While the latter resembles an actual smartphone app, the former directly influences the surrounding world; therefore, the virtual smartphone is an intrinsic part of the environment.

(a) (b) (c) (d) (e) (f)

Figure 10.4: Illustration of the four strategies. (a-c) represent the succession of steps for the situated information on demand by gaze interaction. (d) represents the smartphone with the slider at the top. (e) represents the mini-map available as a situated application. (f) represents the bulletin board with transparency due to the distance to the user.

Strategy 1: Situated Information on Demand with Gaze Interactions

As described in Section 10.4, room indications convey building-related information. Since multiple types of information are available in a restricted area, we hide them inside a gaze-based interactive menu. First, an eye icon (see Figure 10.4a) is visible to the user. If the user wants to check available information, they can open the menu (see Figure 10.4b) by focusing their gaze on the icon and then selecting the desired information. This technique has the advantage of keeping information situated while still being visible from far away and reducing information clutter. As explained by Plopski et al. [2022], using gaze as a means of interaction is problematic. It is usually used as a means of observation, and it is, therefore, difficult to distinguish a user-intended interaction. This problem is more commonly known as the Midas-touch problem. According to Duchowski [2018], multiple solutions exist to get around the Midas-touch problem, such as eye gestures, multi-modality, boundary crossing, etc. However, the most widely accepted solution is a dwell time. Our choice went to the latter with a time of 0.8s.

Strategy 2: Transparency Management for Non-readable Content

Advertising and university bulletin boards need further reflection. This kind of information is typically paid to be displayed. Thus, there exists a conflict of interest between the user's need to reduce information clutter and the needs of the entity paying for its advertisement to be displayed and visible. The solution is to play on the opacity of the information according to the user's distance from the ad (see Figure 10.4f). That way, one can be aware that information is available at that place if their distance from it is considered close enough that the information is readable.

Strategy 3: Reducing or Increasing the Amount of Virtual Elements in the User's Environment

The previous section explains that virtual content can be divided into several categories. Each has a different impact on the user, depending on the context in which they are included. For example, the decorations could be appreciated by a new visitor but too distracting for an employee. Thus, we designed a slider at the top of a virtual smartphone to decrease or increase the number of categories displayed (see Figure 10.4d). Each level includes the elements of the lower levels. Five levels were available: 1) Absolutely no virtual elements except the virtual smartphone, 2) only the eye tracking icons for the situated information and bulletin boards, 3) the navigation indications, 4) the social norms, and 5) the decorations with the ambient sound. Regenbrecht et al. [2022] also proposed this strategy and allows the on/off option recommended by Lu and Bowman [2021] as discussed in Section 10.2.

Strategy 4: Situated Applications

Some types of information offered by the building may not be situated. That is the case for the help desk answering visitors' questions or the maps placed in specific places described in Section 10.4. However, they are related to the building as a whole. Therefore, we decided to remove information from the virtual environment while still letting them accessible from the smartphone as applications, making them accessible as the user wishes. Here, we describe the concept of situated applications. These are available only when accessing the place they are associated with and would be removed and replaced by others as the user moves by. We consider these applications and their installation secure.

10.5.2 Tasks

We asked the participants to find successively four pieces of hidden information inside the experiment corridor at each trial. This approach lets us, on the one hand, explore users' behavior in a street-like environment by exploring the virtual environment and, on the other hand, force the participants to test the different interactions proposed by the strategies. An example of requested information was, for instance, the course title given at a specific time in a specific classroom, the name of a person, or a date for a specific event. The tasks were designed to force participants to explore and interact across the virtual environment. The information to be found differed from one trial to the next but corresponded to the same difficulties (i.e., similar information).

10.5.3 Variables

Our first objective is to study the evolution of users' mental load between an environment using no strategies and an environment using the strategies described in Section 10.5.1. Hence, we chose the NASA TLX form Human Performance Research Group [1980] with six scales to compute the overall workload score of a participant. This form is one of the most widely used in computer science for calculating cognitive load. The standard way of filling out the NASA TLX form asks participants to complete the scales and give a weight for each. However, as we wanted to evaluate the impact of the entire environment on the cognitive load through a daily life task, the weight of each scale was predefined and communicated to the participants. Users had to walk around and observe the environment. We put the following order of importance on the scales, from most important to least important: Mental Demand, Frustration, Effort, Performance, Temporal Demand, and Physical Demand. Finally, we performed a semi-structured interview to retrieve the participants' feedback. This allowed us to understand the TLX reports correctly and focus on our second goal, the ease of interaction and implementation of the four strategies. The questions that structured the interviews were: 1) "How do strategies impact user feelings and their perception of reality?", 2) "Is each strategy useful and properly implemented?", 3) "Are there other useful strategies?" and 4) "What virtual elements were the most frustrating?".

10.5.4 Pretests

We conducted pretests to refine the procedure described in Section 10.5.6 and to check the concordance with the pilot study described in Section 10.3. Five members of the Computer Science department tested the experiment. Four of them were participants in the pilot study. They concluded that the virtual environment was oppressive and overwhelming if they had to experiment it all day long. Regarding the strategies, users considered them as of great interest. Note that they knew the corridor where the experiment took place; thus, they knew where to find some answers during the experimentation.

10.5.5 Participants

A total of eight men and eight women took part in the experimentation. None of them knew the floors of the building, and none had previously participated in the Pilot Study or the Pretests. They were all volunteers; no remuneration was given. The age range was between 21 and 35, with an average of 27.31 years old. Three participants had a high school degree, four had a bachelor's degree, and nine had a master's degree. Their background was heterogeneous. Ten participants had never tested AR or Virtual Reality (VR). Four played occasionally with VR, one played

Figure 10.5: The picture of Times Square was used to determine the participants' comfort level in big cities. Photo taken by Tagger Yancey IV.

occasionally with AR and VR, and one played regularly with VR. Finally, based on results from Section 10.3, we further asked each participant how they felt if they were in Times Square based on a photo (See Figure 10.5). One answered they felt nauseous, one was uncomfortable, six were neutral, four were calm, and four were comfortable.

10.5.6 Procedure

We received each participant individually on the first floor of the building and accompanied them to a meeting room. To begin with, we explained how the experiment would be conducted. First, the participants would go through two experiments, each followed by a NASA TLX form. This part would last approximately 40 minutes. Then, participants would answer questions during a semi-structured interview, which would last about 20 minutes. Before the experiments started, the participants were asked to sign a data processing agreement and complete a form with personal information (i.e., the information described within Section 10.5.5). Then, they performed the ocular calibration on the headset, allowing them to discover how to interact with virtual content as the device displays a floating window with a button to click on to start the calibration.

After this setup phase, the first trial, with the strategies disabled, could start. Explanations were given to the participants as follows: They play the role of a building staff member, and they can walk anywhere on the floor except for some rooms. The participants had time to get used to the Hololens 2 device and the environment. Four successive questions were then asked to the participant without further

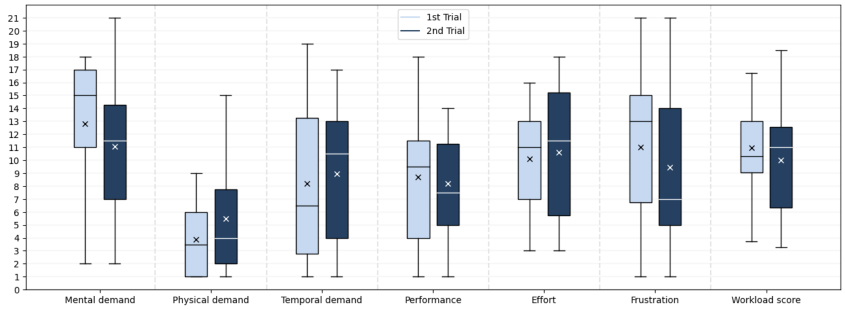


Figure 10.6: Box plots of each subscale and the workload scores of the TLX. The first trial's results are on the left each time, while the second trial's results are on the right.

indication. There was no time limitation, but the investigators would give some clues if a participant felt completely lost. The investigators would also ensure the participant's security while taking notes. Once the four questions were answered, the participant returned to the meeting room to complete the NASA TLX form. Next, the second trial with the four strategies enabled began following the same protocol. The investigators gave explanations of the three active strategies. During the adaptation time, the participant could explore the virtual smartphone capabilities. At the end of this second trial, the participant completed a second NASA TLX form, and a semi-structured interview was conducted. The interviews were audio recorded. Not balancing the trials was a deliberate choice, as we did not want participants to have to apprehend the virtual environment simultaneously with the interactions required for the second trial.

10.5.7 Results

From the NASA TLX forms, we begin by computing the workload score of each participant. This is followed by a Wilcoxon signed-rank test to determine significant improvements in the second trial. Then, we test the dependence between the forms and the participants' familiarity with MR technologies and their comfort in big cities with a chi-square independence test. Finally, we finish with the results of the semi-structured interviews presented in Section 10.5.3. Results were obtained with a thematic analysis of the audio recordings.

TLX Workload Score and Wilcoxon Signed-rank Test

Figure 10.6 presents the distribution for each scale and the overall workload score. Regarding the mean and median, we observe an improvement in Mental Demand, Performance, and Frustration. Conversely, we observe a deterioration of Temporal Demand. To verify the significance of these effects, we performed a Wilcoxon signed-rank test on all scales and the workload score. No significant results were found.

Chi-square Independence Test

We performed chi-square tests to determine the impact of MR technology awareness and comfort level in big cities on each scale and the workload score. For technology awareness, we observe no significant results. However, we can observe a trend in the performance (X^2 = 38.67, p-value = 0.07, DoF = 27) of the first trial and on the temporal demand (X^2 = 42.4, p-value = 0.03, DoF = 27) and frustration (X^2 = 41.47, p-value = 0.08, DoF = 30) of the second trial. These results seem logical as better control of these novel technologies reduces the learning time and, thus, the frustration. Similarly, no significant effects were found for comfort level except a trend on the Mental Demand (X^2 = 53.33, p-value = 0.08, DoF = 40) of the first trial. Similarly, it might be acceptable to assume that if the participant feels good at overwhelming cities, they felt good in our experiment's environment.

Impact of the Strategies on Participants' Feelings and Perception of Reality

Finally, we describe the results of the semi-structured interviews. Relative to the first trial, participants felt mainly calm. Nevertheless, everyone said being overwhelmed by the amount of information, while one specified it as intrusive. The participants pointed out the sound within the environment as a major contributor to their feelings. Therefore, 12 participants considered the experience as unusable in their daily lives. 12 of the participants also considered their perception of reality to be deteriorated. In addition, 8 participants found themselves isolated from reality. Eventually, all participants considered the information to be well-placed. For the second trial, 14 participants found their perception of reality enhanced compared with the first trial. Additionally, 8 participants considered the different strategies improved their experience.

Strategy 1: Situated Information on Demand with Gaze Interactions

As we expected, 15 participants said that this strategy reduced information clutter. 4 said that it is most useful when knowing where to find the information. This last statement is supported by 8 participants saying it negatively hides the information when searching for it. 9 said it structured the information, allowing them to navigate and isolate the wanted information easily. Finally, 5 participants said the gaze interaction icon was placed too high, and two participants would have preferred another way of triggering the interaction than the dwell time.

Strategy 2: Transparency Management for Non-readable Content

10 participants stated that they found bulletin board transparency useful, while 3 were against this strategy, and 3 had no opinion. 12 participants explicitly said that being unable to see information availability while searching for it from far away may trouble them. One participant stated that universities' bulletin boards are not equivalent to advertising panels, and 7 said the technique could benefit from an icon indicating available information when too far away from the user. Finally, 12 participants said that the moving bear drawing attention was manageable. However, how the advertisement was displayed was not.

Strategy 3: Reducing or Increasing the Amount of Virtual Elements in the User's Environment

9 participants mentioned that they had difficulties interacting with the slider. We observed that most participants selected the third or fourth level of information during the experiment because they did not see any difference between the two while experimenting. However, they noticed that these levels mute the audio. 5 expressly wished for a separate way of muting the audio, while 2 participants stated that they kept a high level of information because they feared they would lose some information otherwise. Also, 7 explicitly said they would have preferred an alternative way of interacting where they could turn on and off the chosen information level as they pleased. All participants liked the concept and the control it gave them over the environment.

Strategy 4: Situated Applications

13 of all participants found the idea behind situated applications useful. However, 5 were concerned about security issues and their privacy, even though we focused on security during the question phrasing. 7 stated that it was helpful since it created a pattern to access and find available information. Finally, 2 participants said that the applications should stay installed and only their content should be situated; thus, users would know how to use the applications and what information would be available.

Ideas and Frustrations

Before finishing the interviews, we requested each participant to classify the three most frustrating things they experienced during the experiment. Moreover, we asked them if they had any ideas for improving such an environment. Recurring answers for the former were 1) mute the ambient sound they considered noise, 2) remove advertisements occluding the view, and 3) improve the interaction for the different active techniques. For the latter, some expressed the need for more interaction with the situated information, such as grasping it, resizing it, or making it scroll. Others would have wanted the ability to access the desired information without having to walk or save it for later inside the map application. Finally, a participant said that daytime could be used to filter out unneeded information.

10.5.8 Discussion

Participants' feedback shows that an unmanaged and unconstrained augmented world is difficult to withstand; thus, elaborating strategies to handle augmented information clutter is vital. However, it is worth noting that the Hololens headset's limited field of view impacted the results obtained. This led to a reduction in visual distractions, which may have contributed to lower results than those that would have been obtained with a larger field of view. However, since both trials were affected equally, the observations made in comparing the two remain valid and unaffected.

The strategies proposed in this section were appreciated. Participants' mental demand and perception of reality tend to be improved as expected. Decreasing the amount of virtual situated information frees the user's environment, while the slider allows them to specify their preferences depending on their needs. For instance, a participant said that, in a museum, they would typically set on the last level to fully experiment with the virtual content, but in their daily life, they set it on a lower level to prevent distractions. The slider allows users to specify their preferences rather than share a common level across all users. In addition, situated applications offer users quick access to information specific to their location and an easy way to grasp its spatial organization. As for bulletin boards, as they are advertisements that are paid to be displayed, participants understood that the same technique as the gaze input could not be applied. Despite the majority of participants finding the transparency strategy useful, they were only partially convinced by it. First, it hides information; last, bulletin boards were not the most obtrusive elements.

Nevertheless, it is difficult to determine the strategies' efficiency, as no statistically significant effects were found. First, a separate evaluation of the four strategies was not carried out, partly due to technical limitations (headset battery, evaluation duration, etc.) and partly because our aim in this evaluation is, above all, to determine the user interest in the four strategies. Depending on the results obtained, a process of refinement of each strategy found to be of interest will be carried out in future work. Second, we observed frustration and mental fatigue in the participants who had difficulty interacting. A small training session outside the experiment could prevent this in future works. Third, the short duration of each trial did not allow the participants to properly overcome the novelty effect of trying an AR headset. For instance, a participant said they selected the fourth level on the slider to stop the sound while keeping the maximum amount of virtual elements to get the most out of this innovative experience. As far as we know, the minimum adaptation time to correctly study users' behaviors in pervasive environments is unknown. We plan to conduct a more extended experiment once the strategies are refined to study the evolution of the participants' cognitive load. Last, while using physically available information to guide or answer questions was not prohibited, participants prevented themselves from doing so. A participant explained that they needed to find out if the virtual information was associated with the real location.

We now discuss the implementation of the different strategies and their ease of interaction. For strategy 1, some participants recommended directly displaying the different categories of information (as can be seen in Figure 10.4b). The current state increases the necessary interaction time significantly. This suggestion raises the question of the right balance between information clutter and the number of interaction steps needed to access available information. For frequent use, two successive dwell times of 0.8 seconds seem too burdensome. An alternative would be to display a reduced level of information and then let users keep and expand only what is desired. For strategy 3, participants wanted more freedom in selecting what is displayed and what is not. Isolating visual and audio controls is also crucial. Thus, a better solution would be a toggle for each category of audio and visual content rather than a slider. Therefore, users would be able to select only the ones they want. A toggle to turn augmented reality content on and off without changing all the specified preferences would also be required. This result is consistent with Lu et al. Lu and Bowman [2021] stating that participants sometimes want to be relieved. For strategy 4, some participants wish to standardize visuals and interactions, thus pushing for situated content rather than a situated application. The help application should be implemented as a chatbot (i.e., Siri, Alexia, ChatGPT, etc.). Depending on the location, the application should have access to a dataset of information specific to the place to help it answer users' questions. Answers presentation must be explored in future works but should not be limited to oral answers. The map application must indicate the information's position relative to the world and let the user access it without moving. One participant mentioned that they felt like they were regressing to a time before smartphones existed when you always had to go to different places for information. Eventually, the map must guide users to locations only on demand. Regarding strategy 2, participants only asked to add an icon when the boards disappeared. As said above, participants considered this strategy as not mandatory, even if they understood the interest in freeing the view when unreadable. Some participants said they are accustomed to advertising panels and no longer pay attention.

10.6 Second Experiment: Extending the WiM Metaphor

In the first experiment, participants perceived the need to move around to search and obtain virtual information as a step backward. Indeed, current technologies allow anyone with internet access to access any information from anywhere in the world. However, in a scenario such as the one described above, a large part of the meaning conveyed by the information displayed depends on where the information is located. Different local authorities (e.g., government or shopkeepers) may add information in the virtual environment (VE), and this information, when situated in its context, will take its whole meaning when situated in its context and environment Marriott et al. [2018]. Therefore, decoupling visualizations from a VE from the world in which they are situated can only be done by losing part of their semantics. Based upon these three identified requirements (i.e., 1. search and filter information of use wherever it might be, 2. access potentially distant and occluded information without any required physical movement, and 3. lose the minimum amount of information when viewing distant situated information), this experiment addresses the following question: "how can we provide access to distant and potentially occluded situated information from any location without losing the semantic aspect of its placement?". This second experimentation has made the content of a publication [Septon et al., 2023].

10.6.1 Existing Techniques

This section reviews the current literature on techniques that can address the research question mentioned previously. Three main families of techniques can be identified: a) view management techniques, b) the egocentric techniques, and c) the exocentric techniques. For the last two, we follow the definition of Poupyrev and Ichikawa [1999] applied to selection techniques. The egocentric category regroups any interaction centered around the user, where they interact directly from inside the world. The latter presents techniques allowing the user to interact as if from outside the virtual environment with a God's eye point of view.

All these techniques were discussed in Section 10.2 and Chapter 7. However, none of these techniques can meet all the constraints set out: 1) search and filter information of use wherever it might be, 2) access distant and obstructed information without any required physical movement, and 3) keep the context of distant and situated information. Indeed, while view management techniques partially solve information clutter, they are not usable for distant information as they force the user to move around the environment to find the desired information. Egocentric techniques present the same issue, even if the user is guided to the information. Moreover, if more than one information corresponds to what the user is searching for, the user will have to go from one place to another. Finally, the automatic scaling technique is limited to in-view content so that obstructed information remains unreachable and suffers from information clutter. Only the World-in-Miniature (WiM) metaphor allows for distant and obstructed information to be visualized without imposing too much physical constraint over its user. Furthermore, presenting a miniature version of the VE allows for information to remain situated. However, the metaphor suffers greatly from information clutter. Therefore, we propose to combine the WiM metaphor with a view management technique to allow for information clutter to be attenuated inside the WiM as inside the VE.

10.6.2 View Management Integration

We turn to a text-input-based approach to find content in the VE. Doing so allows the user to specify the desired information directly through keywords and filter out unwanted information inside the WiM and the VE. It is implemented using a virtual keyboard, which requires freehand typing. This approach quickly reduces information clutter and redirects user attention to matching information. Moreover, since matching information appears inside the WiM metaphor, this gives the user an overview of all information around them, even if it is obstructed or out of sight. The text-based input solution is somewhat similar to what is done in AR browsers, such as in the work from Langlotz et al. [2014], for instance. Indeed, based upon a search specified by the user, they display results directly inside the VE. However, in our use case, results contain information already inside the environment, and keywords serve as filters instead of search terms. Therefore, we will refer to this technique as the **filtering mechanism** for the rest of the chapter.

10.6.3 WiM Design Choices

The built WiM (illustrated by Figure 10.7) was designed as follows. We chose to model a corridor on a floor of our building as the environment for the first experiment. The logical structure of the corridor (see Figure 10.3) was mapped, as did Trueba et al. [2009], to a 3D representation. We included the floor and the walls of the corridor, as both serve as visual aids to information placement, indicating the corridor's shape and the location of various visual cues, such as doors. However, we excluded the roof as it adds no value. We then chose to represent precisely the VE in the WiM as it is. We made the WiM accessible anywhere, using a button accessible to the user when looking at their palm (see Figure 10.8a). When pressed, the WiM appears world-fixed in front of the user. It can be closed using the same button (see Figure 10.8b).

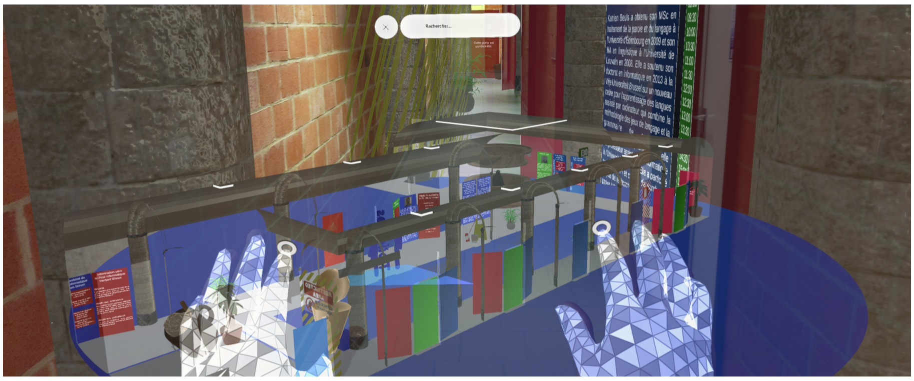
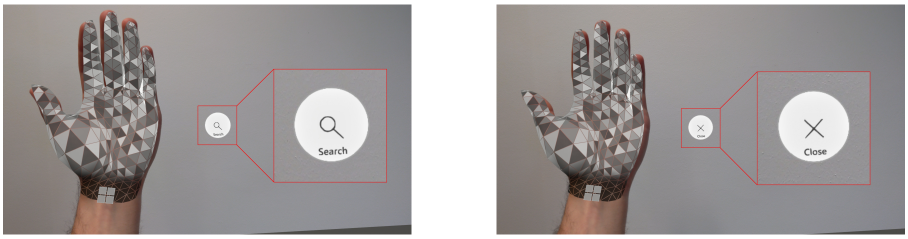


Figure 10.7: Screenshot of the WiM prototype running inside the experimental environment.

The implemented filtering mechanism UI is visible in Figure 10.8d on top of the WiM. It was designed to be the equivalent of current smartphone internet search bars. Indeed, most users are familiar with this kind of tool and how it operates. Therefore, we used an equivalent approach for its visual placement: it is placed above the metaphor since it is used as an input for keywords, and the WiM serves as the visualization method for the results. When entering keywords, it automatically updates the WiM and the VE content, leaving only matching items. Finally, when the WiM has been modified or if any keyword was written, the displayed content and the map placement can be reset through a button available next to the filtering mechanism.

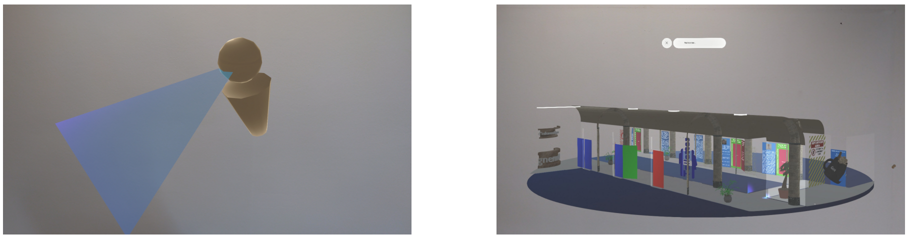
To allow the user to refine the information, we allowed three types of manipulation: rotating the map, zooming in and out, and moving the map. All manipulations required both hands to be realized and were performed by grabbing the map inside the prototype. Once each hand is grabbing, moving the map is allowed by moving both hands in the same direction; rotation can be performed by moving one hand around the other, and finally, zooming in and out is done by moving each hand away or closer respectively. While prototyping, we realized that letting the user move the map along the Y axis inside the WiM restriction zone (see the blue plane in Figure 10.8d) made the designed filtering mechanism position to be somewhat unpredictable since placed above the map. Therefore, we blocked the Y-axis movement manipulation but allowed the user to move the map freely over the X and Z axes. Note that we use a left-handed convention for the axes, with the Y pointing to the top and the Z pointing forward. We first considered blocking all axes for rotation since it helps users transcribe the WiM content to the world as explained by Wingrave et al. [2006]. However, after some testing and reflections, we realized that blocking the Y-axis implied too many physical constraints to gain access to all information. We, therefore, unlocked that specific axis. To overcome the potential induced orientation problem, we represented the user in the form of an avatar (see Figure 10.8c) moving inside the WiM according to their position and orientation inside the environment.

To enable the WiM to be an autonomous tool and be used without further user movement, we allowed continuous zooming inside the logical structure without



(a) Button to open the WiM.

(b) Button to close the WiM.



(c) User representation used inside WiM. The blue halo representing user's gaze.

(d) A blue place delimits the zone of interaction of the technique.

Figure 10.8: Overview of different functionalities of the WiM prototype.

Figure 10.9: Link to a video of our WiM implementation: https://www.youtube.com/watch?v=Mt3kq6Ct4gQ

restrictions, as did Wingrave et al. [2006]. Thus, any information inside the VE can be zoomed in and visualized without the user having to walk to it. However, the WiM can be scaled without limit, creating information clutter. Therefore, we limited the WiM interaction area to a specific zone (see the blue plane from Figure 10.8d) and removed elements from the WiM that would be outside that delimited zone. As it did not serve any purpose in our use case, we restrained zooming out when exceeding the default value (set to 1:67) so the map can stay visible even after a faulty handling. The WiM is illustrated by a video in Figure 10.9.

10.6.4 Participants

A total of 26 people participated in the experiment, with a distribution of 22 men and 4 women. Except for one participant, none were laboratory colleagues. However, that person did not work on Augmented reality-linked topics and had no previous knowledge of our work. Six had a high school diploma, one had a bachelor's degree, seventeen had a master's degree, and two had a PhD. Their background was heterogeneous. When asked about their familiarity with AR, ten said they had no previous experience, fifteen said they had little experience, and one mentioned having some acquaintance with the technology. The participants ranged in age from 19 to 38 years old, with an average age of 27.

10.6.5 Protocol

We received each participant separately inside our office, where two experimenters were systematically present. We began the experiment by asking the participants to fill in a data protection consent form. Then, we let them fill in a short survey asking for the information detailed in Section 10.6.4. We would then describe how the experiment would take place before going to the experiment floor, where the participant would be given a Hololens 2 headset running the experiment and were asked to adjust it to feel comfortable.

The experimenters then gave the participant time to adjust to the environment. After they felt ready, they would be asked to find information inside the environment. We deliberately did not plan any real baseline for the experiment, as our primary objective was to check that the tool was usable and met the needs expressed in the first study. It is the only solution we have found to meet these needs, and so comparing its effectiveness with that of not having a tool was of no interest to us at this stage, since it was a request from the participants. Once the information was found, the participant would be presented with the developed prototype as explained in Sections 10.6.2 and 10.6.3. They were told that they could open and close the WiM (presented as a tool) by looking at one of their hands and interacting with the virtual button, as can be seen inside Figures 10.8a and 10.8b. WiM manipulations were also presented to them by one of the experimenters and, finally, the filtering mechanism would be presented as a filter working with keywords and having an impact on the VE and the WiM. To let the participants take hold of the tool, they were then asked to use it to find the same information they found earlier without it. Once the participant was comfortable enough and without any other indications from the experimenters, the participant had to use the WiM to find three different pieces of information in the environment. The participant is free to use the tool as he wishes. Experimenters took note of how the WiM was used by the participant. Experimenters and participants returned to the office when all questions were answered. The participant was then asked to fill in the User Experience Questionnaire (UEQ) about the tool usage. This form is specifically designed to measure the User Experience of interactive products. Moreover, it is designed in thirty languages, including French, the language the participants speak, which avoids language bias when they fill in the form. Finally, a few open questions were asked to confirm observations taken during the experiment.

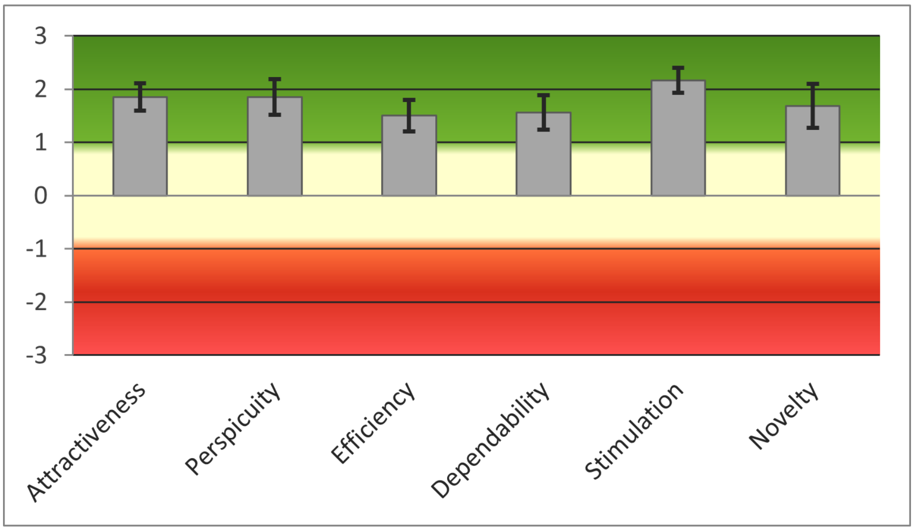


Figure 10.10: Results from the UEQ.

10.6.6 Technical Notes and Limitations

The described prototype was implemented through Unity version 2021.3.6f1. It uses the Mixed Reality Toolkit (MRTK) version 2.8 and was designed to be run on the Microsoft Hololens 2. The VE was registered upon the corridor using the MRTK World Locking feature. The filtering mechanism uses the native keyboard from the device and, as such, includes the possibility of using the dedicated speech-to-text feature. Also, it uses a Levenshtein distance-based algorithm to filter the content from the scene (number of characters to be modified to move from one string to another).

10.6.7 Results

The results are illustrated in Figures 10.10 and 10.11. From the initial 26 questionnaires, 3 had to be removed because of inconsistencies, as advised inside the UEQ Handbook Schrepp [2023]. UEQ results show strongly positive user experiences, with a score of 1.855 on the Attractiveness scale, showing that participants liked the overall experience of the WiM prototype. Regarding the prototype's hedonic quality (score of 1.92), it shows participants liked the non-goal-oriented aspect of the WiM prototype. Indeed, with a score of 2.163 for the Stimulation and 1.685 for the Novelty, their respective benchmarks are in the "excellent" category. Concerning the pragmatic quality (score of 1.64) and its different components, scores range from 1.5 to 1.85. The Perspicuity aspect reflects how easy it was for them to apprehend the prototype and its interactions. While we told participants how to interact with the prototype, they had no further instructions. The map and the filtering mechanism made understanding their use easy, as confirmed by the score of 1.848 for Perspicuity and its benchmark qualified as *"good"*. The Efficiency aspect reflects the difficulty of using the prototype to solve a task. A score of 1.5 confirms that the

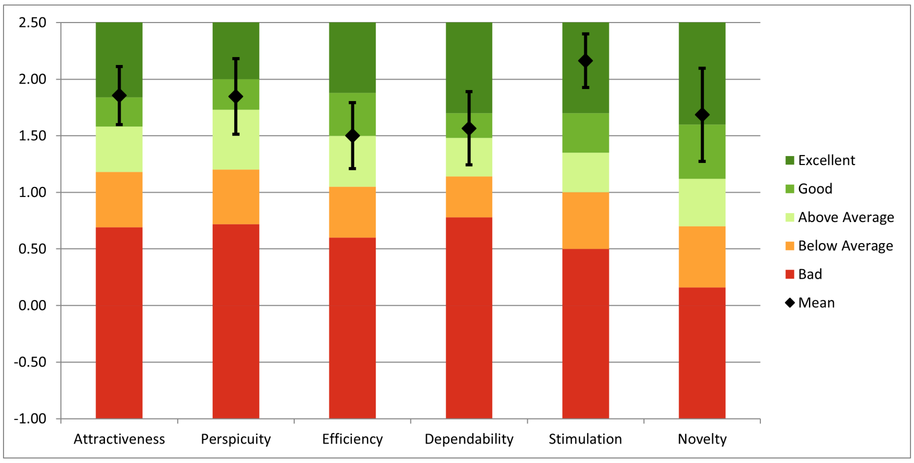


Figure 10.11: Benchmarks from the UEQ.

WiM metaphor along the filtering mechanism developed helped participants find information quickly across the developed VE. Finally, for the Dependability aspect, our prototype shows good results, too. A score of 1.565 and a benchmark marked as *"good"* indicates that users felt controlling the tool.

Furthermore, notes taken by the experimenters identified different behaviors among the participants. First, while using the keyboard to enter keywords inside the filtering mechanism, 7 participants tried to use integrated voice input, but only one of them succeeded. 2 others said that it would be a nice feature to have but did not find the dedicated button. However, almost all participants agreed that the keyboard was not easy to use at first and was time-consuming. Another interesting user behavior observed is the way participants obtained the information. Once they were presented with the WiM, all but one user found it natural to filter information in the VE. 19 participants used the zoom feature of the WiM to reach for the information, while 4 identified the place before going there physically. 3 others used both options. Subsequent discussions with the four participants who walked revealed that 2 of them did not understand that the zoom feature allowed them to get the information but would have probably used it. Four participants would have liked more visual links between the filtering mechanism and the WiM itself. Additionally, 8 participants have tried or would have wanted to click on the content in the WiM to enlarge it. Finally, all participants except one saw the interest and use of such a tool.

10.6.8 Discussion

The results presented in the previous section prove that the WiM as presented is effective and relatively well adopted. However, the excitement of the novelty brought by AR must have played a part in achieving such good results for the hedonic qualities despite the acclimatization periods given to the participants. Also, it should be noted that some participants expressed difficulties while using the device keyboard. While speech-to-text was available to the participants through a button inside the keyboard, the lack of haptic feedback made freehand typing on the keyboard hard for some participants. Exploring other text entry methods for the filtering mechanism using multimodal interaction, such as the work from Lystbæk et al. [2022a], could enhance user experience and improve Efficiency. Additionally, the short period of use cannot let the participants entirely experiment with the prototype in a case of PAR where the continuous use of the AR without interruption in the day must play an important role. Material constraints such as weight and battery life mean we cannot assess this point, but future evaluations will have to be carried out when technological developments allow.

The fact that some of the participants had taken part in the first experiment may be a risk of bias in the results, as they were accustomed to the proposed environment. Nevertheless, the information to be searched was different from the last time, and participants were asked to at least use the tool to locate the information they were looking for. Since our main objective was to check that the tool was easy to use, and to ensure that it addressed the problem highlighted in the first experiment, it is conceivable that the participation of these people was more of a gain than a bias. In fact, it allowed us to verify the second part of our objective.

As expected, most participants used zoom to consult information instead of moving to the actual position in the environment. This stayed true unless the participants considered the information close enough to directly walk to it instead of relying on the zoom feature of the WiM. That is when it takes longer to zoom than to move. This observation suggests that zooming alone is probably not the most effective technique for consulting WiM information. This hypothesis was supported by the 8 participants who would have liked to have clicked on the element instead. Combining the two techniques should help reach a larger group of users.

Another observation of this study is the user's preference on the feedback after a search. Despite our explanation that the WiM is a mirror of the environment, some participants did not perceive the change after typing their request. Reinforce visual feedback by guiding the user's gaze towards the WiM to help a novel user understand where the response appears. Additionally, we choose to hide non-related elements to the request. However, we remarked that it was impossible to determine whether certain elements had matched quickly. The user must scan all the WiM to determine all locations where information is still displayed. Presenting results inside the WiM ordered according to a search-based notion of importance would probably improve the effectiveness of this tool.

Finally, our prototype is a working solution for situated information but cannot evaluate embedded information. Embedded information [Marriott et al., 2018] goes a step beyond situated information, as it is closely embedded with its surrounding environment (as much contextually as semantically) (see Figure 10.12). As presented, the WiM allows the visualization of all virtual information and schematized real-life information, such as the floor and wall. Therefore, future works may be interested in whether and, if so, how embedded information should be integrated into WiM. In the same way, for both situated and embedded information, it is important to study the amount of real information that needs to find its way onto the WiM. In all cases, this problem reinforces the need to associate guiding techniques with our solution, such as those discussed in Chapter 7 on the navigation.

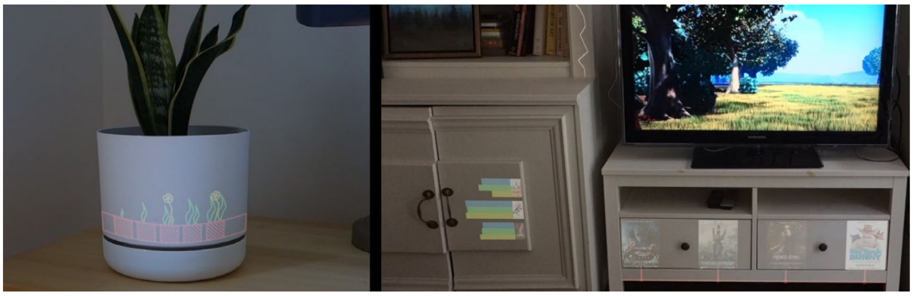


Figure 10.12: Illustration of embedded information (Source: Joshi et al. [2022]).

10.7 Enhanced Strategies for Future Work

The aim of the third experimental iteration resides in formulating novel strategies designed from the outcomes of the initial two experimental phases. These strategies encompass refinements of previously assessed strategies and novel approaches extrapolated from the input provided by participants. Unfortunately, these strategies have not yet been the subject of a user study at the time of writing. Therefore, this section will describe them without discussing their usefulness, effectiveness, implementation, and cognitive effort.

Out of the strategies investigated in the initial experiment, the approach involving the modulation of transparency in advertising based on user proximity is excluded due to user perceptions of its non-essential nature and limited usefulness. The remaining strategies, including the approach explored in the second experiment, underwent revision to yield the ensuing set of strategies described below.

In addition to the strategies, we also intend to strengthen our methodological protocol during the study. To this end, we intend to rely more on Rauschnabel et al. [2024]'s work to determine more precisely the impact of the hedonic properties through the 4C framework (consumer, content, context, and computing device) on the results, and on Lagasse et al. [2023]'s work to help users to recall more easily what they have experienced through the *REMIND* method.

Strategy 1: Virtual Smartphone

In the first experiment, the virtual smartphone was designed as a virtualization of the physical smartphone, supporting strategies without being one itself and keeping the conventional presentation of the application in a grid form. In contrast, the virtual smartphone now evolved in a strategy of its own by extending the combination of the filtering mechanism and WiM explored in the second experiment. In addition to the capability of text-based content filtering and the reset button, this virtual smartphone iteration incorporates two buttons and three tabs. The first button facilitates access to smartphone applications, presented in a grid layout akin to physical smartphones or the earlier virtual smartphone model (refer to Figure 10.4d). The second button provides access to the headset and user experience settings, enabling users to personalize their interaction with the various strategies proposed in this study. Insights from the previous experiments underscored individual preferences in tool operation, highlighting the significance of user-customizable experiences. These settings mirror those encountered when acquiring new devices or the initial launch of applications, aimed at initial adjustments rather than frequent adjustments. Next, the three tabs serve to display filtered keyword results from three distinct information sources: (1) outcomes derived from applications (i.e., applications that may respond to the user need), (2) outcomes sourced from the Internet similar to a browser's functionality, and (3) outcomes originating from the user's environment, using a WiM such as in the second experiment. The tab-based design mitigates information overload by categorizing results, offering a user-friendly structure for targeted information retrieval. Furthermore, this virtual smartphone iteration is anchored to the user's body to trail their movements. Notably, it permits the user to pivot independently without imposing synchronous smartphone orientation adjustments, as would be the case with a fixed screen arrangement. Finally, the user can decide in the settings menu to click on an element of the WiM either to zoom automatically on it, to display a copy in a readable size over the map, or to do nothing, as is the case in the second experiment. This choice was made in response to the behavior of some participants who tried to click on the WiM elements to read its content.

Strategy 2: Enabling or Disabling Element Categories

In the initial experimental phase, the approach employed was a slider mechanism, facilitating the incremental or decremental adjustment of categories attributed to virtual elements. The term "sequential" herein denotes a condition where each distinct category corresponds to a distinct position along the slider, thereby enforcing the deactivation of all positions above the targeted one. Nonetheless, participants highlighted the necessity of disassociating visual and auditory stimuli. They also expressed the need to turn off a singular category without imposing the deactivation of others, as mandated by the progressive nature of the slider-based method. To address these concerns, we abandon the slider mechanism in favor of an array of checkboxes housed within a panel reminiscent of the settings menu in the virtual smartphone interfaces. This panel will offer alternatives for either the deactivation of diverse categories of constituent elements within the environments (as discussed in Section 10.4), with a separation of visual and auditory components, or the deactivation based on the creators of said elements. In the context of our experimental setup, these creators might encompass entities like the university administration, faculty administration, professors, or students. Recognizing that access to this menu will likely occur more frequently than engagements with the virtual smartphone's conventional settings menu, we advocate including a physical button on the device's periphery and voice command to facilitate access to the panel. This button can also quickly switch between displaying nothing, displaying the selected configuration, and displaying everything without altering the selected configuration. The difference between opening the menu and changing the level of details can be made according to the type of click on the button (i.e., fast click and long click).

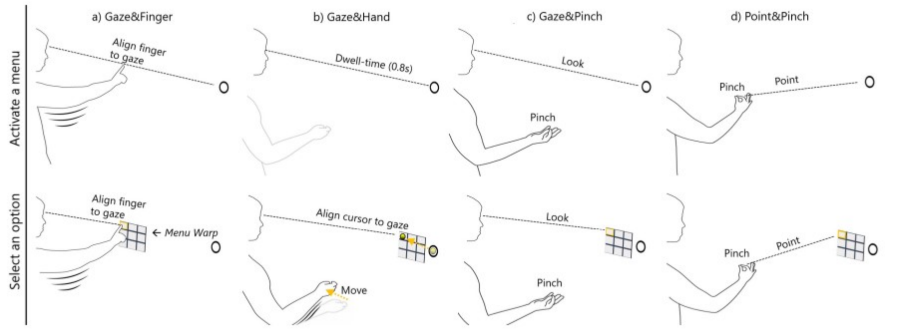


Figure 10.13: Illustration of the four techniques explored by Lystbæk et al. [2022b]. The illustration comes from their paper.

Strategy 3: Situated Information on Demand with Gaze Interactions

This is the strategy that will have evolved the least from its implementation in the first experiment. On the advice of some of the participants, it is proposed that the transition between the eye icon (see Figure 10.4a) and the thumbnails representing the various elements (see Figure 10.4b) is automatic when the user gets closer. In addition, an indication that the thumbnails can be activated with the eye gaze must be added. The thumbnails are to be moved to the center of the doors so that people of different heights can access them. Once selected, the content of each thumbnail can also be displayed in the form of a resonant-sized window in front of the user so that the user doesn't have to stand right in front of the door to read the content. Content is primarily text or images, but audio can be activated at the user's request. In addition, the method to resolve the Midas problem encountered when using eye gaze, as discussed in Section 10.5.1, can be selected in a set of propositions in the settings menu of the virtual smartphone. Once again, these choices are made in an attempt to adapt to the entire population.

Strategy 4: Selecting Information from Far Away

The last proposed strategy involves the deliberate distant selection of visible information. This strategy's rationale mirrors the previously discussed use of the WiM, primarily aimed at minimizing user physical movements. However, in this instance, the objective is to grant users access to items that are visually perceptible yet physically out of arm's reach without the process of entering keywords in the filtering mechanism. When a distant object is selected, it is virtually replicated in front of the user. Subsequently, the user can interact with this virtual duplicate without physically approaching the object. Nonetheless, it's essential to recognize that the diverse scenarios in which this technique may be employed impose significant constraints.

As elucidated by Bowman et al. [1999], the selection of elements can be deconstructed into three distinct subtasks: the **indication of object**, the **indication of selection** (confirmation), and **feedback**. The separation of the first two subtasks allows for the utilization of distinct input methods to address them, while feedback is essential to ensure the user's successful execution of the operation. First, to maximize the utility of this approach, users must be afforded as much freedom of movement as possible. For instance, the user's hand should remain unencumbered to perform other actions, whether in the physical or virtual worlds. Second, the selection process should be both fast and intuitive. Ideally, users should be able to execute the action without focusing on the process. Any cognitive burden or effort required for this process can lead to user fatigue and disrupt immersion. Third, as emphasized by Argelaguet and Andujar [2013], Mendes et al. [2017], and Weise et al. [2019] in their respective taxonomies and classifications, the spatial boundaries within which users can select objects must be clearly defined. There is no compelling reason for this space to extend beyond the user's visual field, i.e., the space containing perceptible elements. In the first scenario, a selection space that is too large would necessitate physical movement, potentially causing user frustration, an outcome we seek to avoid. In the second scenario, if the user cannot visually perceive the elements, they will also be unable to discern any feedback. So, selecting information from this concealed space without physically adjusting the user's viewpoint should not be considered. This reasoning leads us to exclude the virtual hand technique family (techniques based on the user's hand or on virtual representation of the hands), which, unlike the virtual pointing technique family, does not enable control throughout the entire visual space [Argelaguet and Andujar, 2013]. Lastly, Argelaguet and Andujar [2013], Mendes et al. [2017], Weise et al. [2019], and Leusmann [2021] have also explored the question of selecting one or multiple elements simultaneously. We propose allowing users to specify a reasonable selection radius to accommodate user preferences. If the user is close to the target, they can precisely select the item of interest. Conversely, if they are too distant, all selected items are presented as thumbnails, permitting the user to make a final choice later. The specific selection radius can be tailored to user preferences and adjustable in real-time during the selection process. The selection radius is unmistakably reminiscent of the notion of nimbus (i.e., space in which an element is detectable by another element) and focus (i.e., space in which an element can perceive other elements) introduced by Höllerer et al. [2001], which indicates when two elements can interact with each other.

Following Lystbæk et al. [2022b], gaze represents the user's natural and practical means of pointing at objects of interest, effectively indicating their focus of attention. Additionally, hands provide a highly expressive means of pointing and convey semantic input. Their research explored four techniques combining gaze and hand input (see Figure 10.13). These techniques include (1) aligning the finger with the gaze, (2) moving a cursor using the hand and aligning it with the gaze during a dwell time, (3) confirming actions indicated by the gaze through pinching, and (4) pointing and pinching using only the hand. Their findings demonstrated that combining gaze and hand input outperforms using hands alone. Furthermore, the first approach, aligning the finger with the gaze, yields the most favorable results. It would also enable the selection radius to be selected by moving the finger closer to or further away from the face. Furthermore, we propose including voice commands as a secondary modality for confirming elements of interest indicated by gaze. Lastly, as a third option, we consider the potential for detecting hand movements via a connected watch or bracelet to eliminate the need for users to raise their arms during interaction.

10.8 General Discussion

The feedback from the pilot study, corroborated by findings from the initial experiment, lends empirical support to our hypothesis regarding the necessity of implementing strategies in the domain of PAR to assist people in coping with an escalating oppressive and overwhelming context. Nevertheless, it is noteworthy that including advertising content did not significantly escalate users' cognitive loads despite being the only strategy exhibiting marginal redundancy. This may be attributable, in part, to the relatively non-intrusive nature of most advertisements in our experiments, akin to conventional wall posters, and the development of societal mechanisms that shield individuals from their influence. In our study, the overall amalgamation of environmental stimuli contributed to a sense of suffocation and mild cognitive fatigue. However, these effects were not statistically significant due to the limited sample size. Additionally, we observed that auditory stimuli induced saturation more rapidly than visual clutter.

A surprising observation was the need for more research addressing similar issues in this domain. Furthermore, the existing body of work primarily focuses on smartphone virtualization through context-aware mechanisms that operate autonomously, circumventing user intervention. However, our solutions emphasize conscious user choice and have demonstrated considerable promise. Users expressed appreciation for tailoring their surroundings to their preferences, advocating for synthesizing both approaches in future research endeavors. While automated methods can alleviate mental and physical burdens, they can engender confusion and unpredictability. Context-aware solutions, whether heuristic-based or reliant on machine learning algorithms, necessitate extensive training data, a challenge exacerbated by the diverse user behaviors, preferences, and environmental contexts we identified in our experiments.

The diversity of user behaviors and the myriad situations and environments they may encounter underscores the importance of adopting a multimodal approach to implementing each strategy. Additionally, we advocate for a more in-depth exploration of the boundary between user customization and imposed solutions, especially within personal contexts governed primarily by legal, ethical, and security standards.

Turning our attention to the proposed strategies, three out of our initial four strategies exhibited promise. Nevertheless, our initial approach to emulating behaviors observed in everyday life proved to be imperfect. User feedback indicated that bridging the gap between the real and virtual worlds requires not only the integration of primary sensory modalities (sight, hearing, touch, taste, and smell) but also social behaviors. In essence, augmented reality should enable real-world environments to harness the capabilities of computer technologies, such as web-based information retrieval. So, we would like to advocate for further exploration of the fusion between web browsers and real-world environments. As said, on the one hand, bringing information search techniques to search for real elements and, on the other hand, using the real world to display internet search results. MacIntyre et al. [2011] made a first step in this direction. This strategy encourages users to attend a location only when necessary physically. Moreover, we postulate that enabling users to access distant information remotely will prove advantageous, as it follows a similar principle.

However, the validity of these two hypotheses necessitates validation through the completion of the third experiment. This observation also raises novel questions, particularly concerning the role of remote collaboration within organizations that leverage such MR technologies, permitting users to visit locations virtually without physical presence.

Two other points that will require particular attention in future work are, on the one hand, the time given to each user to experiment with the prototype and, on the other, the types of tasks demanded to the user. By definition, PAR is used continuously in everyday life. However, in our experiments, due to human and technical constraints, we were only able to leave the user with the helmet on for up to 20 minutes. Moreover, it is necessary to clarify how much time is required to correctly simulate a PAR experience. Second, the diversity of the tasks to be performed is also important. In our case, we have now limited ourselves to information retrieval, but we should extend this range of tasks in the future. Ideally, of course, we would like to leave the user with the system that would assist him in his various daily tasks, but this implies having a sufficiently robust and complete system to last over time, adapt to each user, and assist him in a wide variety of tasks involving or not involving virtual content. Unfortunately, the technical constraints of current headsets (i.e., weight, size, batteries, etc.) do not yet allow this kind of experimentation.

In summation, the field of PAR is in its nascent stages and is just beginning to yield promising initial outcomes. Numerous unexplored avenues beckon, with increasing urgency due to major corporations' gradual integration of mixed technologies into everyday life. Thus, it is imperative to establish comprehensive guidelines to mitigate information clutter. Our work was a first step in this direction, but many challenges remain to be explored.

10.9 Summary

This chapter has explored the challenges of managing information overload in PAR environments. We conducted user testing of various information control techniques, including simplifying information down to an icon, hiding information based on distance or user preferences, and providing applications to assist users with navigation. Our studies revealed that hiding information based on distance was not considered efficient. In addition, in MR, the integration of real and virtual worlds must not only consider the visual and interactive aspects but also human behavior. A significant example of this is how information is accessed. In this regard, combining WiM with a text filtering system proved to be an effective approach.

Considering the environment used to conduct the experiments, a vast area of research still needs to be explored. So far, many studies have concentrated on the immersive version of smartphone applications or automatic context-aware approaches, but very few have examined the augmentation of the world itself. With the rapid evolution of MR technologies, the research conducted in this chapter serves as an entry point to the plethora of questions that may arise regarding PAR.

Part IV

Discussion, Future Works and Conclusion

CHAPTER 11

DISCUSSION

This chapter summarizes the contributions of this thesis and discusses their implications for research and practice. Section 11.1 maps the contributions and discusses answers to the research questions formulated in Chapter 4. Next, Section 11.2 explains how these contributions impact various practice domains. Finally, Section 11.3 addresses the limitations of this thesis.

11.1 Implications for Research

The research conducted in this thesis was focused on five research questions, each concerning a specific design level of User Interface (UI) (see Figure 11.1). The first level, represented by research question 1, is the **asset** level that studies assets' characteristics. The second level, represented by research question 2, is the **canonical tasks** level that defines visualization and interaction idioms adapted to these kind of tasks. The third level, represented by research questions 3 and 4, is the **applications** level that concerns the assembly of all design elements required for the application's purpose. Finally, the last level, represented by research question 5, is the **cross-applications** level that studies all issues common to various applications. Table 11.1 reminds us of the research questions.

Overall, twelve main contributions (i.e., related to the research questions) and three secondary contributions (i.e., not related to the research questions) are presented in this thesis. All main contributions concerned immersive Augmented Reality (AR) on Head-Mounted Display (HMD). Table 11.1 presents the mapping between this thesis's contributions and research questions. We discuss them in the following paragraphs.

The first research contribution is the **survey literature (C1)** conducted in Chapter 6, which gathered existing guidelines on text readability from scientific and practice sources, highlighting the research gaps. It is linked to three practical inputs: a **list**

	Chapter 6				Chapter 7				Chapter 8				Chapter 9		Chapter 10		
Research Questions	**C1**	**P1**	**P2**	**P3**	**C2**	**P4**	**P5**	**P6**	**C3**	**P7**	**P8**	**P9**	**P10**	**P11**	**C4**	**P12**	**P13**
RQ1: What text parameters, including the associated contextual constraints affecting these parameters, can designers tune to improve text readability?	**X**	**X**	**X**	**X**													
RQ2: What are the best visualization idioms that can be used to guide the user to different points of interest?				X	**X**	**X**	**X**	**X**									
RQ3: How can Immersive Augmented Reality support existing conventional applications for text understanding?				X				X	**X**	**X**					X		
RQ4: Which approaches best enable remote collaboration in industrial scenarios?				X				X	X				**X**	**X**	X	X	
RQ5: How can the issue of information clutter be mitigated through active control from the user in Pervasive Augmented Reality scenarios?				X				X							**X**	**X**	
Bonus 1: Popularization of immersive AR in society				**X**				**X**					**X**	**X**			**X**
Bonus 2: Development of a web application for text analysis								**X**			**X**	**X**					

Table 11.1: Mapping between the contributions, the practical inputs, and the research questions of this thesis. A black X indicates a direct contribution to the research question. A gray X indicates an indirect contribution to the research question.

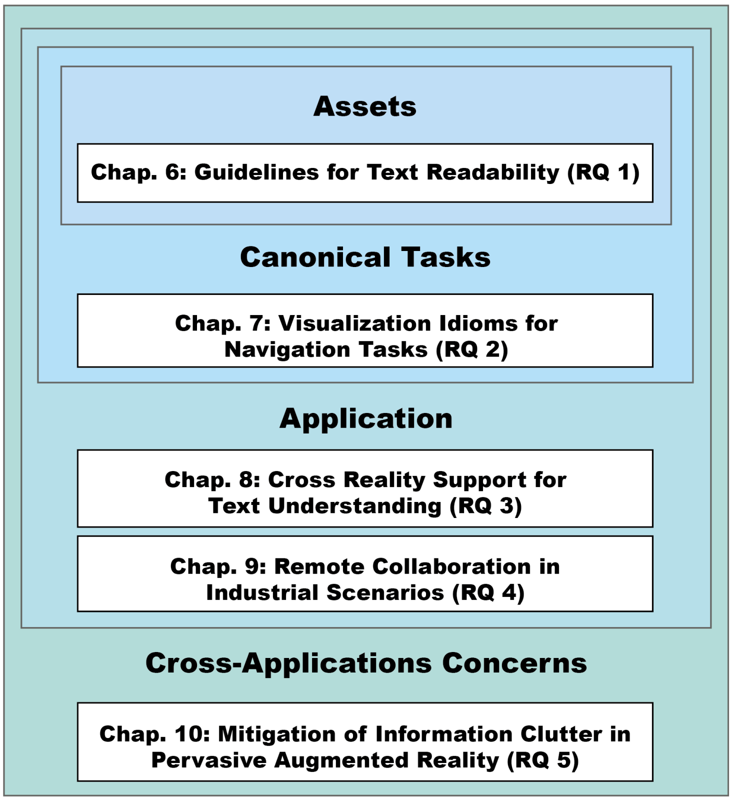


Figure 11.1: Plan of the doctoral focus with the four levels of design and their association with the research questions.

of text parameters (P1) and **contextual constraints (P2)** impacting text readability, as well as a **set of guidelines (P3)** for improving text readability. Subsequently, a subset of these guidelines took the form of two decision trees. Regarding the research questions, this contribution and these practical inputs directly answer research question 1. Additionally, the third contribution (C3) impacts the other research questions, as this level of design lays the foundations for subsequent levels. Moreover, it also contributes to popularizing immersive AR in society.

Chapter 7 discusses the following research contribution. A scientific **literature review (C2)** determined a list of visualization idioms suitable for localization and guiding to off-screen real and virtual points of interest (POIs) without modifying the user's perception of real. As for the first contribution, this review determined research gaps and future works. The review also led to a **list of idioms (P4)**, a **list of tasks (P5)** used to compare these visualization idioms, and a **comparison of visualization idioms for each task (P6)**. Concerning the research questions, this contribution and these practical inputs directly answer research question 2. For the same reasons as the third practical input, the sixth practical input (P6) impacts the other research questions and popularizes immersive AR in society. Additionally, as 2D visualization idioms were included in the review, these idioms can be applied to the web application detailed in Chapter 8.

Next, chapter 8 presents the third research contribution that consist in the exploration of a **cross-terminal application (C3)** combining a web application with an immersive AR application. This takes place within a particular use case of text analysis with experts in humanity sciences. Regarding practical input, **we evaluated four metaphors (P7)** assigned to various analysis elements. This contribution and this practical input answer research question 3, each on a different level. The seventh practical input enables us to understand better how metaphors can support experts in analyzing text with the post-structuralism approach, thus contributing at an idioms level. The third contribution (C3) allows us to understand the expected use of applications combining web and Mixed Reality (MR) for a perfect combination of both, thus contributing at a usage level. Furthermore, about this last contribution, the experts interviewed stressed the importance of collaboration within the immersive application, which ties in with research question 4. As it is an expert application, we do not consider that we are popularizing immersive AR in society. Finally, the chapter discusses the development of a **web application (P8)** and the **Shock Wave algorithm (P9)**, which, while not central to this thesis, are practical inputs of their own for the development of web applications for text analysis.

Chapter 9 does not introduce a research contribution as there is no evaluation of the prototype at this time (it will be conducted in future work as will be discussed in Chapter 12), but it introduces two practical inputs that will lead to a research contribution: the **interview of a remote expert on challenges for collaboration application (p10)** and the **design of a remote collaboration application (p11)** that respects an expert's need in industrial maintenance. The practical inputs answer directly to research question 4. In addition, as collaboration is an important characteristic of the immersive AR, the design process conducted for the proposed application design contributes to popularizing immersive AR in society.

Finally, Chapter 10 discusses the last research contribution. We evaluated four strategies to **mitigate information clutter (C4)** and, subsequently following participant's feedback, we evaluated the use of a World-in-Miniature (WiM) combining with a filtering mechanism to access distant information. Consequently, each strategy validated by the participants can be considered a practical input, except that the implementation of the first four was considered a failure by the participants. Therefore, we consider only the **WiM as a practical input (P12)**. This contribution and this practical input occur in a Pervasive Augmented Reality (PAR) environment and, thus, answer directly to research question 5. Naturally, given that this research was conducted at the cross-application level of design levels, the contribution of this research has an impact on research questions 3 and 4. In addition, we noticed that the WiM extension (P12) could be extended to other issues, such as collaboration for improving remote expert control and awareness of the on-site environment, hence its link with research question 4. Last, the **exploration of users' feelings in a PAR environment (P13)** helps to popularize immersive AR in society.

11.2 Implications for Practice

In addition to the research contributions, this thesis contributes to various practice domains. The most obvious is, of course, the designers of immersive AR applications, who will be able to design based on the third (C3), sixth (C6), eighth (C8), eleventh (C11), thirteenth (C13), and fourteenth (C14) contributions to improve the usability of their applications. They will also be able to draw inspiration from our seventh (C7), twelfth (C12), and fifteenth (C15) contributions if their applications come close to the use cases we have developed.

A second practice domain is the legal and ethical department, which defines technology-related laws and regulations within society. Exploring the PAR in Chapter 10 highlighted a critical societal issue. The fact that, with current technologies, we were able to create such an environment that was found overwhelming and oppressing by the participants demonstrates the importance of putting this topic at the top of the agenda in the various domains. While our work has opened up some exciting avenues for computer science, it is essential that the legal and ethical fields also take up the issue.

Finally, as already discussed, our contributions to the Evocative Framework For Text Analysis - Mediality Models (EFFaTA-MeM) web application have enabled a small number of experts to benefit from a text analysis support application. Compared to many others, this application supports analysts during the analysis process without replacing them. Rather than automating everything, we proposed an application in which the user controls what happens.

11.3 General Limitations

This section outlines five limitations of the research discussed in this thesis. These limitations apply across multiple chapters and provide a broader perspective than what is covered in the individual discussions of each chapter. We encourage readers to review the dedicated chapter discussions for a more detailed exploration of each contribution and limitation of the associated research questions.

Alternative Research Designs

The first limitation concerns the choice of research questions. First, **why did we choose to explore each level of design?** Indeed, dwelling on the different levels prevented us from studying them intensely. Nevertheless, the four-level division of the design process is a structure that results from the long research process that led to this thesis. Initially, our focus was solely on research questions 3 and 4, but we were soon confronted with more fundamental issues illustrated by research questions 1 and 2. This also explains why the latter were addressed through a literature review. The primary objective was to provide a basis for the design of research questions 3 and 4. Moreover, research question 5 was chosen for two main reasons. The first reason was that we felt this was a complication we would eventually encounter with the two use cases we were interested in. The second reason is that we thought we were responsible for this new technology and wanted to contribute to its popularization by keeping the user at the heart of the design process. The latter is valid for all our research questions since we have always adopted this context-centric approach. Still, this choice is particularly evident with the PAR use case, which will likely affect everyone.

Second, **why did we choose these specific research questions?** Indeed, at each level of design, we only focused on a small part, deliberately leaving various assets, canonical tasks, application types, and cross-application problems to one side. The answer to the first point partly explains this choice. Research questions 1, 2, and 5 addressed specific issues in developing research questions 3 and 4. For the latter, on the other hand, the choice was dictated by the opportunities available to us. The two projects that funded this thesis offered an ideal environment, providing the domain experts needed for a context-centric approach. In addition, collaboration is seen by some as a key challenge for immersive augmented reality [Billinghurst, 2021], a view reinforced by our observations. Note that we use a different use case for research question 5 because we need a use case with sufficient information clutter to study the issue properly.

Third, **why did we choose to study two use cases at the application level?** On one hand, focusing on just one would have enabled a more intensive analysis. Conversely, focusing on more than two would have provided greater coverage of this design level. Selecting more than two use cases would not have been possible within the 4-year timeframe of this thesis. We chose two rather than one because they bring a different vision, with support to a web application, playing on different levels of Milgram's spectrum, and a fully immersive application, presenting different challenges. In the end, comparing the two will bring exciting observations on the usage of each case. Additionally, as already explained, these use cases result from opportunities that evolved during the four years dedicated to the thesis.

Independent Research Questions

The second limitation concerns the independence of each research design adopted to address the various use cases. We already discussed the implications between the four levels of design. However, **why a comparison between the development of prototypes for research questions 3, 4, and 5 was not further developed?** We needed to lay the foundations for each research question before comparing the guidelines obtained. Evaluation with the EFFaTA-MeM project team's internal experts showed that our approach to answering research question 3 needed to be refined. For research question 4, unfortunately, we have not yet evaluated the prototype. Finally, research question 5 was not concerned with designing a PAR environment but with techniques for mitigating the information clutter. In conclusion, the comparison between a cross-terminal application and a fully immersive application will occur in future work, as discussed in the next chapter.

Preliminary Contributions for Research Questions 3 and 4

The third limitation concerns the preliminary state of research questions 3 and 4. This thesis is a step in our research, which will be continued after that. In addition to the problems of time and difficult access to experts already discussed, research questions 3 and 4 are less explored than the other three, as we have adopted a more systematic approach than is more traditionally found in the literature. In other words, we were interested in the design of each element from the smaller to the bigger, which led to the elaboration of the four levels of design. Then, **why not adopt a more traditional approach from the outset?** During these four years of research, we have discovered that the literature on application development, in particular, tends to focus on combining elements and evaluating the application, compared to a more conventional approach without AR. So, there is a gap between the different levels of design in the literature, not least because there is no standard like WCAG[1] for the web. In this thesis, we adopted an approach to validate elements from the smaller (i.e., assets) to the bigger ones to build immutable rules to base the design of future applications. Of course, this approach is more time-consuming, but it complements those widely used in literature.

Quality of Reviewed Papers

The fourth limitation concerns the quality of review papers. During our research, we decided to refrain from applying quality criteria to the papers we reviewed. This is particularly true for research questions 1 and 2 since, for the other three, we have selected the papers most representative of the fields. The first reason was that we had chosen a systematic approach to the literature to bring together all the literature on the two research questions. In addition, we wanted to leave it up to the reader to decide which criteria to apply to avoid any subjectivity on our part in selecting one criterion over another. The definition of objective criteria is a research focus in its own right and too far removed from the focus of this thesis. Nevertheless, we recognize that these criteria can help resolve contradictions in the literature between several papers — a situation we encountered, for example, in defining decision trees for research question 1.

Limited Multimodal Interactions

As explained in the introduction to this thesis, another major challenge in the field of immersion is the interaction challenge, which is intrinsically linked to the user interface challenge. An interface depends on both its visual elements, which influence how we interact with them, and its interactions, which influence how we represent the interactive elements. In this respect, the literature on MR technologies is fairly unanimous in favoring multimodality, enabling an action to be carried out through different interaction solutions [Billinghurst, 2021; Kim et al., 2018; Ong et al., 2008]. The use of multimodality makes it easier to personalize the application and adapt to changing environments. When building the prototypes presented in this thesis, we chose to limit interaction recognition to what was available in the device, namely native gesture and speech recognition. This choice was strongly influenced by a concern for the performance of the prototypes, which were already pushing the headsets to the limit of their resources. Nonetheless, when evaluating WiM, we allowed users to use their hands to fill in the filter bar rather than using the keyboard. Some users also tried to use the native speech recognition integrated into the headset. However, the recognition performance was so bad that they reverted to using the keyboard. Testing multimodal interaction with a better speech recognition engine would be part of the future work of this thesis.

---

[1] https://www.w3.org/WAI/standards-guidelines/wcag/

CHAPTER 12

FUTURE WORK

The limitations discussed in Chapter 11 paved the way for several future work directions. In this chapter, we present five major leads for further research. They propose new contributions for which the research presented in this thesis can serve as a basis.

12.1 Framework based on the Four Levels of Design

This thesis has been built around the division into four levels of design, allowing us to adopt a "divide and conquer" approach to building an application with a context-centric approach. However, this structure is an element that became clear to us at the end of the thesis, even though it was implicitly present throughout our research. As a result, it would be interesting to take this division a step further to form a methodological framework in its own right. The impact of the lowest levels (i.e., assets) on the highest levels (i.e., cross-application) seems intuitive enough. Still, are there also implications from the highest to the lowest levels? If so, what is the role of each reading direction, and how can they be combined? In addition, we applied the Munzner [2015]'s framework of What, Why, and How questions extended with the Where, When, and Who at the different levels, but a better interweaving of the two approaches may be possible.

12.2 Further Research Specific to each Level of Design

This thesis has only touched on a tiny fraction of the issues associated with each level of design. A logical follow-up to this thesis would be to continue exploring them.

Asset Level

As explained in Chapter 6, we have defined avenues for future research into text characteristics. We recommend focusing future research efforts on (1) better defining the rules around parameters under the text appearance and segmentation label, (2) clarifying the range of optimal text size, (3) defining which results obtained on Video See-Through (VST) (resp. Optical See-Through (OST)) displays can also be applied on OST (resp. VST) displays without repeating the study, and (4) defining a taxonomy of text usage and associate readability rules to each. One approach that is beginning to emerge is the development of automatic adaptation algorithms. Nevertheless, work often tends to be limited to the use of the camera, but immersive Augmented Reality (AR) systems are usually equipped with a wide range of sensors. Using this data for asset adaptability would be an exciting avenue worth exploring. In addition, the hardware development of displays can also address some limitations encountered or at least reduce their impacts. This automatic adaptation is in line with what is currently being done at a higher level, with context-aware applications that react automatically to adapt to the context in which the user is evolving. This raises the question of how an application can be customized if behaviors are automatically managed without the user having to do anything.

Moreover, the text is not the only asset that needs to be studied. To name just a few, there are 3D models, icons, images, videos, etc. All have their parameters that designers can play with to enrich their applications. Of course, these parameters are also linked to contextual constraints, which may be the same as for text or may be new. Ultimately, the main objective would be to define a set of guidelines that would give rise to WCAG-style[1] regulations for the web. These guidelines would ensure the accessibility of all applications that comply with them. This regulation would be a way of bridging the gap that exists between the scientific and practice sectors.

Last, some works propose approaches to select the right assets for the job [Gattullo et al., 2020, 2017; Laviola et al., 2021; Tainaka et al., 2023]. As we mentioned in the introduction to this thesis, documentation, for example, is enriched with 3D models and other assets rather than the usual textual form. Each asset has its advantages and weaknesses when conveying information. It is, therefore, essential to define the context in which each asset is to be used to optimize the usability of each application.

Canonical Tasks Level

Regarding the canonical tasks level, the objective is threefold. First, it is necessary to define an exhaustive list of canonical tasks to be studied. Second, it is essential to identify and devise new techniques adapted to all these tasks. Finally, define which methods each task should use in a given context. Defining the canonical tasks and contextual elements to be considered is an issue in its own right, given the infinite number of use cases and sheer number of factors impacting the user experience. The most classical approach, on which the background of this thesis is based, is the classification of verbs used in the lexical fields associated with the major domains in which immersive technologies are used.

---

[1] https://www.w3.org/WAI/standards-guidelines/wcag/

Among these, tasks that recur almost systematically are selection and navigation. We addressed the former in part in Chapter 10, where we developed different remote information selection approaches to mitigate information clutter in the context of Pervasive Augmented Reality (PAR). Our results showed that they do not adopt the same approach depending on the user's context. If the information was not too far away, some participants tended to move to find the information in the real world. If, on the other hand, the information was more distant and hidden, participants tended to seek it out through the World-in-Miniature (WiM). This first example highlights the impact of context on these elementary tasks.

We found the same reasoning in Chapter 7, dedicated to the study of visual idioms supporting navigation tasks to off-screen points of interest (POIs). We identified a series of tasks used for evaluation, highlighting that the visual idioms must not be the same depending on the reason for making an action. Identifying the motivations for each canonical action within a catalog and adapting the visualizations accordingly would, in the long term, ensure consistency between applications and guarantee their usability. Consistency in techniques is an issue that participants raised while discussing the situated application strategy in Chapter 10.

Application Level

Considering the application level, studying every conceivable application is impossible. It is, therefore, necessary to determine a classification of applications to explore them by category. The most important thing is to determine general guidelines that can be shared between applications. An example of an application category will be mind-map applications that present various models and uses. On the other hand, contextual elements specific to an application will require a particular study that cannot be generalized. This is the case, for example, with expert applications based on particular behaviors linked to training or the expert's domain. In this respect, further exploration of the collaboration concerning research question 3 and the evaluation of research question 4 must be carried out. There is a relatively strong focus on collaborative applications, and rightly so. Our observations align with the literature on the fact that this is a primary feature of the new approach provided by immersion technologies. In this respect, a great deal of work has been carried out on the various cues that must be provided to improve verbal and non-verbal communication. However, further work is needed to enhance the effectiveness of each of these techniques.

Moreover, our research in Chapter 8 highlights the importance of putting the cross-terminal aspect of applications back at the center of discussions. What is the role of each type of terminal as defined in Weiser [1991]'s vision? Literature has shown the importance of collaboration within immersive technologies by demonstrating a positive impact on the major psychological components (presence, engagement, awareness) affecting collaboration between two people. What is more, these technologies break up space by virtually teleporting each user to the others. This is, nevertheless, a hypothesis we would like to verify, but one that is currently supported by the literature. But are these the only advantages of this technology? We are convinced that it is not as 3D visualizations and workspace were also enhanced, but a more in-depth study needs to be conducted. In our view, interaction is one way of answering these questions. We believe that conventional interfaces offer a different kind of control over data but are complementary to the freedom of movement within data offered by immersive analytics. That is why, in the following of our research, we will try to associate user tasks with a type of terminal by analyzing the interactions required to carry out these tasks.

Cross-Application Concerns Level

Finally, at the last level, we studied information clutter. Once again, our results indicated that it was still possible to iterate on the prototype we had set up. An evaluation of the latest versions of the strategies will be carried out shortly. Beyond that, other issues can be addressed, such as security. An exhaustive list should be able to be determined through studies carried out on conventional computing, the problems at this level being transversal to the devices used. Nevertheless, it is not clear whether the impact of these problems is the same on usability in the different levels of mixed reality. How do presence, engagement, and awareness affect the techniques to be adopted to mitigate these concerns? Ultimately, the question of what rules apply from conventional computing studies to immersive technologies is a problem we find at every level of design. We need to re-establish the basic rules we have taken for granted until now.

12.3 Questionnaire for the design of applications with a context-centric approach

The principle of context-centric application design consists of eliminating visualization and interaction idioms that do not match the constraints of our environment, users, tasks, etc. To help designers in the future, it would be interesting to provide help in the shape of a form, for example, that forces them to think about their applications in terms of constraints. At the start of this thesis, we created a first version, which we shelved before evaluation due to lack of content (see Appendix B). We could describe various elements of an application but not yet link them with contextual constraints, thus guidelines. Based on the results we propose and future works we identified, as discussed in the first two sections of this chapter, we believe there is the basis to start such a project. We can cite the work of Palmarini et al. [2017] and Gattullo et al. [2019], which points in this direction. This project would be an excellent way of bringing together the scientific and practice worlds - A gap we raised in Chapter 6.

12.4 Pervasive Augmented Reality Scenario in Daily Life

We have already discussed it, but we consider the study of PAR in everyday life to be a major issue that needs to be addressed from a transdisciplinary point of view to maintain control over the future development of this kind of technology. Society is constantly evolving, but research must support this evolution by trying to make the most of it and denounce any possible abusive lures. Under no circumstances should it be a brake, but rather an essential player in helping to establish a shared ethical guideline that can protect citizens. As we discussed in Chapter 10, there is a significant risk of abuse and manipulation by oneself or others, which can have a major impact on the users' psychology, as well as on their perception and understanding of the real world. Even for more sporadic and specialized applications such as those discussed in Chapters 8 and 9, there are also ethical and legal risks to consider, such as user safety. This is why we insist on continuing to explore such scenarios. For example, at the level of computer science, it is possible to determine guidelines on how to get information to users without putting them in an unpleasant situation, as we did in Chapter 10. Another research direction will be to address the issue of multiple sources of information and how to ensure visual cohesion. Moreover, the evolution of technology will eventually enable more precise analyses in this context since current technologies make it challenging to carry out long-term studies, particularly the battery and the helmet's weight.

12.5 Development Environment for Immersive Mixed Reality

Although not the focus of our research, the final point we would like to address is the development environment around immersive Mixed Reality (MR) technologies. During the development phases of the various prototypes discussed in this thesis, we encountered the difficulty of this kind of development. Other developers and researchers at conferences also shared these difficulties with us. First, development frameworks that aim to offer developers all-in-one tools are not necessarily compatible with all devices. The OpenXR initiative has tried to solve this problem, but its implementation is taking time. Second, the fastest way to develop in immersive MR is to use a game engine (if we exclude the no-code solution limited to basic elements). The problem is that these development environments are better suited to video games, which have the advantage of 3D development but are less optimized for MR. The other approach is to develop from a lower-level solution, requiring more time and requirements. Third, MR applications are real-time, relying on numerous sensors, which implies a high computational load. Either this load is carried out at the Head-Mounted Display (HMD) level, which more quickly implies latency, as the helmet must be wearable and therefore not carry heavy hardware. Or the experience is streamed to the headset, which depends on factors such as network speed and image compression. In conclusion, to enable the expansion of these technologies, there is a need to explore further standardized solutions, such as OpenXR, specialized game engines for immersive MR, and the development of optimized frameworks that embrace the needs of these kinds of applications.

CHAPTER 13

CONCLUSION

We addressed the challenges of user interfaces within immersive Augmented Reality (AR) environments. While on desktop and mobile, the workspace is in 2D and limited to the size of the screen, in immersive AR, the workspace is in 3D and limited to all the environment. It means rethinking the place and usefulness of 2D and 3D interfaces, including new constraints related to the real environment, such as depth, occlusion, and situated and embedded visualizations. Indeed, the principle of AR is to place information directly onto the real environment. So, there is a vital need to identify good and bad design practices to prevent cognitive and visual overload and prevent generating overwhelming environments.

This thesis focuses on visualization to support the user's task. This main objective was pursued by delving into four design levels, aware of the contextual factors intrinsic to the application's deployment. On the asset level, we entailed examining text parameters contingent on contextual factors and the specificities of immersive devices. On the canonical task level, we concentrated on assessing the influence of context on visualization choices for navigation tasks to real and virtual off-screen points of interest (POIs). On the application level, we studied user behavior using immersive AR either as a complement to conventional text analysis applications for experts in humanities or for remote collaborative engagement in industrial maintenance. Finally, on the cross-application level, we explored strategies to mitigate information clutter in a Pervasive Augmented Reality (PAR) scenario.

Our results take the form of guidelines. We conducted a literature review on text legibility to determine text parameters, the contextual elements influencing them, and the guidelines to be respected. Similarly, we also carried out a survey of the literature on visualization idioms for navigation. In this, we highlighted the visualization idioms, the types of tasks used to evaluate them, and, thus, the best idioms for each task.

We then explored how immersive metaphors can support experts in humanities in text analysis. Our results indicated that metaphors can reinforce principles implicit in text analysis, but the choice of metaphors remains relatively complicated to determine and highly context-dependent. Moreover, our results echoed observations in the literature regarding the animation and realism of metaphor implementation. This exploration also enabled us to raise interest in cross-terminal applications where each terminal plays its role, being complementary but not directly redundant, with collaboration as a key element of the immersive AR. Looking at the collaboration scenario, we have not yet had time to evaluate the prototype proposed to address the issues identified from the interview with a remote user. This assessment will be carried out in future work.

For the last level, our results demonstrated the interest of three of our proposed strategies to mitigate the information clutter in a PAR scenario. However, a lot of feedback was given to refine the implementation of these strategies. The three strategies were (1) situated information on demand with gaze interactions, (2) reducing or increasing the amount of virtual elements in the user's environment, and (3) situated applications. The strategy that did not receive sufficient participant approval was (4) transparency management for non-readable content. Next, based on participants' feedback, we evaluated using a World-in-Miniature (WiM) combined with an information filter mechanism to access distant information. The results demonstrated the efficiency of this technique.

Finally, this thesis has contributed to the challenges mentioned above through different levels of design. These levels form an element that emerges from the thesis and may lead to a fully-fledged framework in future works. Otherwise, this thesis makes fifteen contributions to research and various practice domains. In addition, this thesis is written from and extends seven publications and five student works.

Part V

Appendices

APPENDIX A

INDIVIDUAL ROLE IN PUBLICATIONS

The strategy followed in this thesis was to write and submit an article for each significant contribution that was mature enough. The content of the thesis is based on and extends the publications and student works resulting from collaborations between the author of this thesis, his supervisor, project partners, other researchers, and students. For each publication, the individual role of the author of this thesis is detailed in Table A.1. The roles are described using the taxonomy presented in Table A.2, reproduced from Brand et al. [2015].

Documents	**Type**	**Roles**
Cauz and Cleve [2019]	Poster	Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Data curation, Writing - Original Draft, Visualization
Linden et al. [2020]	Paper	Software, Resources, Validation
Cauz et al. [2021]	Paper	Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data curation, Writing - Original Draft, Visualization
Cauz et al. [2023]	Paper	Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data curation, Writing - Original Draft, Visualization
Septon et al. [2023]	Paper	Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data curation, Writing - Original Draft, Visualization
André et al. [2023]	Poster	Resources, Supervision, Validation
Cauz et al. [2024]	Paper	Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Data curation, Writing - Original Draft, Visualization
Cauz [2019]	Master's thesis	Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data curation, Writing - Original Draft, Visualization
Hincq [2022]	Master's thesis	Resources, Supervision, Validation
Dejardin [2022]	Master's thesis	Resources, Supervision, Validation
Thielemans [2023]	Master's thesis	Resources, Supervision, Validation
Sanfilippo [2023]	Student work	Resources, Supervision, Validation

Table A.1: Roles for each publication and student work.

Term	**Definition**
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims
Methodology	Development or design of methodology; creation of models
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
Data curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse
Writing - Original Draft	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)
Writing - Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or postpublication stages
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
Project Administration	Management and coordination responsibility for the research activity planning and execution
Funding acquisition	Acquisition of the financial support for the project leading to this publication.

Table A.2: Contributor role taxonomy. (Source: Brand et al. [2015])

APPENDIX

B

From Real Environment and Tasks to Augmented Reality, a Draft Questionnaire

This appendix presents the first version of a form that aims to help the transition from a physical environment to an augmented reality version to optimize the efficiency of user tasks. Based on a series of forms, the approach prioritizes the users' needs and tasks rather than the technology itself. After an analysis, the analyst can determine the appropriate direction for developing an effective and tailored augmented reality application for their specific use cases. Unfortunately, these forms are only at the stage of describing the environment and are not yet linked to design guidelines.

Lexicon

Artifact	A real element of the environment.
Environment	The physical environment where the user is involved. It can be divided into numbered subspaces to specify its constraints more clearly.
Subspace	A subpart of the environment with specific constraints against the other subspaces or to divide into more understanding parts of the environment.
User	The person who will use the augmented reality system in the environment to perform tasks. To represent the user in the forms, the analyst can consider that the user has the combination of all constraints that the real users have or create a series of sheets for each family of users.
Remote user	An external person whom a user can contact through videoconferencing for help.

A. Short Description of Tasks

A description of each task must be provided to determine the environment and the constraints on the system that must be described in the next forms. All tasks described in the same form must evolve in the same environment and must highlight the elements to describe later (i.e., data, artifacts, etc.).

Task number

Task description

Task number

Task description

Task number

Task description

Task number

Task description

B. Environment Description

This form describes the properties of the environment. In the case of a large environment (see question B.2), it can be interesting to complete this form for all subspaces and the global environment to specify the constraints more precisely. In the case of a medium environment, the different parts of the environment can also be the subject of different forms of completion if they are too different from each other.

1) Name ☐

To indicate if this form concerns the environment or a subspace.

2) Size of environment ☐ small ☐ medium ☐ large

Small: The user doesn't need to move to cover the entire environment. The head movement is sufficient.

Medium: The user must move in a small area (e.g., 3m x 3m) to cover the entire environment. The head movement is no longer sufficient to access all the parts of the environment.

Large: The user must move in a large area (e.g., more than one room) to cover the entire environment. The user must know the entrance/exit of each subspace and the sequence of subspaces to reach the correct part of the environment.

3) Can be mapped in ☐ 2D ☐ 3D

If one of the dimensions is unimportant, the environment is considered in 2D. For example, a production line or a coffee maker is in 3D, while a physical panel of control composed of buttons is in 2D because the depth of each button doesn't matter.

4) The environment is ☐ outdoor ☐ indoor ☐ both

5) Number of subspaces ☐

6) Number of artifacts ☐

7) The environment is ☐ spacious ☐ narrow

In a spacious environment, the user has the possibility to move without constraints and take distance from his/her point of interest to evaluate the whole system. In a narrow environment, the user has some mobility constraints on one or more 3D axes or does not have the possibility to take distance from the point of interest.

8) The lighting is ☐ shining ☐ normal ☐ dark

The most common environments are considered normal, with lighting by the sun for outdoor environments or artificial lighting for indoor environments. However, sometimes the user must work in a non-sufficiently lit environment or in a too-shining environment for business reasons. It can be the case if the user works at night, for example, and has little access to additional lighting.

9) We have or we can produce

☐ The map of the environment in a format usable by a computer.

E.g., described as a graph of nodes usable by path algorithms.

☐ The 3D model of the environment on computer.

☐ Physical markers (e.g., QR code) to easily recognize the position of the user.

10) Map of the environment

Annotate the position of each subspace/artifact with unique identifiers. It is recommended to use capital letters for subspaces and numbers for artifacts.

C. Data Description

This form describes the data existing in the environment. Data is virtual information that characterizes the environment. For example, in a product line composed of a series of tanks, the quantity of liquid in each tank represents a series of data.

1) Identifier

Each data's identifier must be unique to reference them in the next forms.

2) Type ☐ Number ☐ Graph ☐ Text
☐ Image ☐ Movie

A graph is a matrix of numbers, while a number or a list of numbers are independent values not represented as a graph.

3) Record history of data points ☐ yes ☐ no

Sometimes, the user needs to refer to a history of past recorded values. For a graph, past values mean past graphs and not evolution in the graph.

4) Minimum value

5) Maximum value

A blank indicates that there is no minimum and/or maximum value.

6) The text is ☐ short ☐ medium ☐ long

Short: A text composed of only a few words.

Medium: A text composed of a few sentences.

Long: A text composed of at least two paragraphs.

7) Dynamicity of data ☐ static ☐ dynamic

Static data does not change and can be loaded in a built version of the application. Dynamic data depends on external sources and must be loaded with the help of network interfaces.

8) The access API is ☐ HTTP ☐ Bluetooth

Describe the type of API to obtain the data in case of dynamic data.

9) The loading station is

Indicate where the AR headset can perceive the flow of data. If there is no restriction, note "everywhere" or "environment". If the user must be in a particular space to access the data, note the identifier of the subspace.

D. Artifact Description

This form describes one artifact of the environment. The Form B "Environment Description" and the Form C "Data Description" must be completed before this form. One form by artifact must be completed.

1) Identifier ☐

It must be the same as at B.10.

2) Name ☐

3) Identifier of associated data ☐ ☐ ☐ ☐ ☐

The identifier of each data must correspond to the value in C.1.

4) Type of values ☐ binary ☐ discrete ☐ continuous

Indicate how the interactor updates the data. “binary” means that there are two states (e.g., a button). “discrete” means that the interactor has a list of defined values, like a set of enumerated buttons. “continuous” means that the data can be updated on a continuous list of values, like a wheel.

5) Off value for "binary" ☐

6) On value for "binary" ☐

7) List values for "discrete" ☐ ☐ ☐ ☐ ☐

8) Min. value for "continuous" ☐

9) Max. value for "continuous" ☐

10) Type of interactor

☐ push/pull

☐ displacement/rotation

E. Task Description

This form describes the scenarios of the user in the environment. It must be completed once every step of the tasks and for all scenarios in the environment.

1) Task number

It must correspond to Form A.

2) Time scheduling

Indicate the order of the actions. If two or more actions have the same value, their order is not critical.

3) Precondition(s)

Precondition(s) to perform this action.

4) Postcondition(s)

Postcondition(s) that must be encountered to prevent error or to pass to the next step.

5) In case of error(s)

6) Identifier of artifact

It must correspond to D.1.

7) Value(s) to assign

F. User Description

This appendix describes the equipment and the user's behavior in the environment.

1) Equipment of the user ☐ helmet ☐ gloves ☐ tablet ☐ keyboard

2) Using the AR system ☐ sometimes ☐ regular ☐ always

3) Duration of use ☐ short time ☐ long time

G. Remote User Description

This appendix describes the equipment and the abilities of the remote user to help the user.

1) Equipment of the remote user

- ☐ Computer/tablet to access task/environment information.
- ☐ VR headset.

2) Remote user's abilities

- ☐ The remote user and the user must be able to talk to each other in audio.
- ☐ The remote user and the user must be able to see each other in video.
- ☐ The remote user must be able to draw in the environment of the user.
- ☐ The remote user must be able to share documents with the user.
- ☐ The remote user must be able to synchronise his view with the user's one.

  By using a VR headset, the remote user takes virtually the position of the user.

ACRONYMS

AR	Augmented Reality.
AV	Augmented Virtuality.
BCI	Brain-Computer Interaction.
CAR	Conventional Augmented Reality.
CSCW	Computer-Supported Collaborative Work.
DoF	Degrees of Freedom.
DR	Diminished Reality.
EFFaTA-MeM	Evocative Framework For Text Analysis - Mediality Models.
FLARACC	Federated Learning and Augmented Reality for Advanced Control Centers.
fps	frames per second.
HHD	Hand-Held Display.
HMD	Head-Mounted Display.
HUD	Head-Up Display.
HWD	Head-Worn Display.
IS	Information System.
MfR	Modified Reality.
MLR	Multivocal Literature Review.
MR	Mixed Reality.
MR-CSCW	Computer-Supported Collaborative Work combined with Mixed Reality fields.
OST	Optical See-Through.
PAR	Pervasive Augmented Reality.
POI	point of interest.

SAR	Spatial Augmented Reality.
SLR	Systematic Literature Review.
SUS	System Usability Scale.
SWAVE	Spherical Wave-Based Guidance.
UEQ	User Experience Questionnaire.
UI	User Interface.
UX	User Experience.
VE	virtual environment.
VR	Virtual Reality.
VST	Video See-Through.
WiM	World-in-Miniature.
WIMP	Windows-Icon-Menu-Pointer.
xR	Extended Reality.

BIBLIOGRAPHY

Adams, R. J., Smart, P., and Huff, A. S. (2017). Shades of Grey: Guidelines for Working with the Grey Literature in Systematic Reviews for Management and Organizational Studies. **International Journal of Management Reviews**, 19(4):432–454.

Agati, S. S., Bauer, R. D., da S. Hounsell, M., and Paterno, A. S. (2020). Augmented Reality for Manual Assembly in Industry 4.0: Gathering Guidelines. In **2020 22nd Symposium on Virtual and Augmented Reality (SVR)**, pages 179–188.

Agić, A., Mandić, L., and Loknar, N. S. (2022). Legibility of Typefaces and Preferences of Text/Background Color Variations in Virtual Environment. In **Proceedings - The Eleventh International Symposium GRID 2022**, pages 835–841.

Agrawala, M. and Stolte, C. (2001). Rendering effective route maps: improving usability through generalization. In **Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH '01)**, pages 241–249.

Aguilar, J., Jerez, M., and Rodríguez, T. (2018). CAMeOnto: Context Awareness Meta Ontology Modeling. **Applied Computing and Informatics**, 14(2):202–213.

Ahmad, A. M., Goldiez, B. F., and Hancock, P. A. (2005). Gender Differences in Navigation and Wayfinding using Mobile Augmented Reality. **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, 49(21):1868–1872.

Amburn, C. R., Vey, N. L., Boyce, M. W., and Mize, J. R. (2015). The Augmented REality Sandtable (ARES). techreport, American Army Research Lab.

André, M., Bayet, A., Jetzen, T., Luycx, P., Cauz, M., and Dumas, B. (2023). Engineering User Interfaces with Beat Gestures. In **Companion Proceedings of the 2023 ACM SIGCHI Symposium on Engineering Interactive Computing Systems**, pages 76–78.

Arefin, M. S., Phillips, N., Plopski, A., Gabbard, J. L., and Swan, J. E. (2022). The Effect of Context Switching, Focal Switching Distance, Binocular and Monocular Viewing, and Transient Focal Blur on Human Performance in Optical See-Through Augmented Reality. **IEEE Transactions on Visualization and Computer Graphics**, 28(5):2014–2025.

Argelaguet, F. and Andujar, C. (2013). A survey of 3d object selection techniques for virtual environments. **Computers & Graphics**, 37(3):121–136.

Arntz, A., Keßler, D., Borgert, N., Zengeler, N., Jansen, M., Handmann, U., and Eimler, S. C. (2020). **Navigating a Heavy Industry Environment Using Augmented Reality - A Comparison of Two Indoor Navigation Designs**, volume 12191, pages 3–18.

Attfield, S., Kazai, G., Lalmas, M., and Piwowarski, B. (2011). Towards a science of user engagement (Position Paper). In **WSDM Workshop on user Modeling for Web Applications**.

Azuma, R. T. (1997). A Survey of Augmented Reality. **Presence: Teleoperators and Virtual Environments**, 6(4):355–385.

Bane, R. and Hollerer, T. (2004). Interactive Tools for Virtual X-Ray Vision in Mobile Augmented Reality. In **Third IEEE and ACM International Symposium on Mixed and Augmented Reality**, pages 231–239.

Bangor, A., Kortum, P. T., and Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. **International Journal of Human-Computer Interaction**, 24(6):574–594.

Bangor, A., Kortum, P. T., and Miller, J. T. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. **Journal of Usability Studies**, 4(3):114–123.

Barata, J. (2021). The fourth industrial revolution of supply chains: A tertiary study. **Journal of Engineering and Technology Management**, 60:101624.

Barreiros, C., Veas, E., and Pammer, V. (2018). Can a Green Thumb Make a Difference?: Using a Nature Metaphor to Communicate the Sensor Information of a Coffee Machine. **IEEE Consumer Electronics Magazine**, 7(3):90–98.

Barthes, R. (1966). Introduction à l'analyse structurale des récits. **Communications**, 8(1):1–27.

Baudisch, P. and Rosenholtz, R. (2003). Halo: a technique for visualizing off-screen objects. In **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**, pages 481–488.

Bell, B., Höllerer, T., and Feiner, S. (2002). An annotated situation-awareness aid for augmented reality. In **Proceedings of the 15th annual ACM symposium on User interface software and technology (UIST '02)**, pages 213–216.

Billinghurst, M. (2021). Grand Challenges for Augmented Reality. **Frontiers in Virtual Reality**, 2.

Billinghurst, M., Bowskill, J., and Morphett, J. (1998). WearCom: A Wearable Communication Space. In **Proceedings of CVE**, volume 98, pages 123–130.

Binetti, N., Wu, L., Chen, S., Kruijff, E., Julier, S., and Brumby, D. P. (2021). Using visual and auditory cues to locate out-of-view objects in head-mounted augmented reality. **Displays**, 69:102032.

Biocca, F., Owen, C., Tang, A., and Bohil, C. (2007). Attention issues in spatial information systems: Directing mobile users' visual attention using augmented reality. **Journal of Management Information Systems**, 23(4):163–184.

Biocca, F., Tang, A., Owen, C., and Xiao, F. (2006). Attention funnel: omnidirectional 3d cursor for mobile augmented reality platforms. In **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**, pages 1115–1122.

Bioy, A. and Bourgeois, F. (2003). **La communication entre soignant et soigné: repères et pratiques**. Editions Bréal.

Biswas, N., Singh, A., and Bhattacharya, S. (2022). AroundArrow: Off-screen POI visualization for handheld augmented reality in vertically dense regions. In **2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 570–575.

Blanc-Goldhammer, D. R. and MacKenzie, K. J. (2018). The effects of natural scene statistics on text readability in additive displays. **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, 62(1):1281–1285.

Borg, O., Casanova, R., and Bootsma, R. J. (2015). Reading from a Head-Fixed Display during Walking: Adverse Effects of Gaze Stabilization Mechanisms. **PLOS ONE**, 10(6):e0129902.

Bork, F., Eck, U., and Navab, N. (2019). Birds vs. fish: Visualizing out-of-view objects in augmented reality using 3d minimaps. In **2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 285–286.

Bork, F., Schnelzer, C., Eck, U., and Navab, N. (2018). Towards efficient visual guidance in limited field-of-view head-mounted displays. **IEEE Transactions on Visualization and Computer Graphics**, 24(11):2983–2992.

Bottani, E. and Vignali, G. (2019). Augmented reality technology in the manufacturing industry: A review of the last decade. **IISE Transactions**, 51(3):284–310.

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

Brand, A., Allen, L., Altman, M., Hlava, M., and Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. **Learned Publishing**, 28(2):151–155.

Brooke, J. (2013). Sus: a retrospective. **Journal of usability studies**, 8(2):29–40.

Buchmann, V., Billinghurst, M., and Cockburn, A. (2008). Directional interfaces for wearable augmented reality. In **Proceedings of the 9th ACM SIGCHI New Zealand Chapter's International Conference on Human-Computer Interaction: Design Centered HCI**, pages 47–54.

Buchner, A., Mayr, S., and Brandt, M. (2009). The advantage of positive text-background polarity is due to high display luminance. **Ergonomics**, 52(7):882–886.

Burigat, S. and Chittaro, L. (2007). Navigation in 3d virtual environments: Effects of user experience and location-pointing navigation aids. **International Journal of Human-Computer Studies**, 65(11):945–958.

Burigat, S. and Chittaro, L. (2011a). On the effectiveness of overview+detail visualization on mobile devices. **Personal and Ubiquitous Computing**, 17(2):371–385.

Burigat, S. and Chittaro, L. (2011b). Visualizing references to off-screen content on mobile devices: A comparison of arrows, wedge, and overview+detail. **Interacting with Computers**, 23(2):156–166.

Burigat, S., Chittaro, L., and Gabrielli, S. (2006). Visualizing locations of off-screen objects on mobile devices: a comparative evaluation of three approaches. In **Proceedings of the 8th conference on Human-computer interaction with mobile devices and services**, pages 239–246.

Burigat, S., Chittaro, L., and Vianello, A. (2012). Dynamic visualization of large numbers of off-screen objects on mobile devices: an experimental comparison of wedge and overview+detail. In **Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI ’12)**, pages 93–102.

Buttner, A., Grunvogel, S. M., and Fuhrmann, A. (2020). The influence of text rotation, font and distance on legibility in VR. In **2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)**, pages 662–663.

Büschel, W., Lehmann, A., and Dachselt, R. (2021). MIRIA: A Mixed Reality Toolkit for the In-Situ Visualization and Analysis of Spatio-Temporal Interaction Data. In **Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems**, Article 470, pages 1–15.

Caggianese, G., Gallo, L., and Neroni, P. (2015). User-Driven View Management for Wearable Augmented Reality Systems in the Cultural Heritage Domain. In **2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)**.

Cárdenas-Robledo, L. A., Hernández-Uribe, Ó., Reta, C., and Cantoral-Ceballos, J. A. (2022). Extended reality applications in industry 4.0. – a systematic literature review. **Telematics and Informatics**, 73:101863.

Carmo, M. B., Afonso, A. P., Ferreira, A., Claudio, A. P., and Silva, G. (2016). Poi awareness, relevance and aggregation for augmented reality. In **2016 20th International Conference Information Visualisation (IV)**, pages 300–305.

Carmo, M. B., Afonso, A. P., Melo, M., Rocha, B., and Botan, V. (2020). Augmented reality with maps for off-screen POI awareness. In **2020 24th International Conference Information Visualisation (IV)**, pages 454–459.

Carter, J. A. (1986). A taxonomy of user-oriented functions. **International Journal of Man-Machine Studies**, 24(3):195–292.

Caudell, T. P. and Mizell, D. W. (1992). Augmented reality: An application of heads-up display technology to manual manufacturing processes. In **Hawaii international conference on system sciences (HICSS)**, pages 659–669.

Cauz, M. (2019). Augmented Reality visualization and interaction for maintenance. Master's thesis, University of Namur.

Cauz, M., Albert, J., Wallemacq, A., Linden, I., and Dumas, B. (2021). Shock wave: a graph layout algorithm for text analyzing. In **Proceedings of the 21st ACM Symposium on Document Engineering**, Article 32, pages 1–4.

Cauz, M., Clarinval, A., and Dumas, B. (2024). Text Readability in Augmented Reality: A Multivocal Literature Review [Article under Review Process after Minor Revisions]. **Virtual Reality**.

Cauz, M. and Cleve, A. (2019). Interacting with overlaid information in augmented reality systems for maintenance: A preliminary review. In **2019 13th International Conference on Research Challenges in Information Science (RCIS)**, pages 1–2.

Cauz, M., Septon, T., and Dumas, B. (2023). Interaction Techniques to Control Information Clutter in a Pervasive Augmented Reality Scenario. In **Proceedings of the Internation Conference on Human-Computer Interaction (HCI) - Late Breaking Papers: Virtual, Augmented and Mixed Reality**, pages 3–21.

Chandler, T., Cordeil, M., Czauderna, T., Dwyer, T., Glowacki, J., Goncu, C., Klapperstueck, M., Klein, K., Marriott, K., Schreiber, F., and Wilson, E. (2015). Immersive analytics. In **2015 Big Data Visual Analytics (BDVA)**, pages 1–8.

Chang, C.-C., Liu, C.-J., and Liu, K.-H. (2019). The legibility threshold of Chinese characters on the augmented reality display - A preliminary study. In **50th Nordic Ergonomics and Human Factors Society Conference**, pages 69–72.

Cheliotis, K., Liarokapis, F., Kokla, M., Tomai, E., Pastra, K., Anastopoulou, N., Bezerianou, M., Darra, A., and Kavouras, M. (2023). A systematic review of application development in augmented reality navigation research. **Cartography and Geographic Information Science**, 50(3):249–271.

Chenechal, M. L., Duval, T., Gouranton, V., Royan, J., and Arnaldi, B. (2016). Vishnu: virtual immersive support for HelpiNg users an interaction paradigm for collaborative remote guiding in mixed reality. In **2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)**, pages 9–12.

Cheng, Y. F., Yin, H., Yan, Y., Gugenheimer, J., and Lindlbauer, D. (2022). Towards understanding diminished reality. In **CHI Conference on Human Factors in Computing Systems**, pages 1–16.

Cheok, A. D. and Karunanayaka, K. (2018). **Virtual Taste and Smell Technologies for Multisensory Internet and Virtual Reality**. Springer International Publishing.

Chewar, C. M. and McCrickard, D. S. (2002). Dynamic route descriptions: tradeoffs by usage goals and user characteristics. In **Proceedings of the 2nd international symposium on Smart graphics**, pages 71–78.

Chittaro, L. (2006). Visualizing information on mobile devices. **Computer**, 39(3):40–45.

Chittaro, L. and Burigat, S. (2004). 3d location-pointing as a navigation aid in virtual environments. In **Proceedings of the working conference on Advanced visual interfaces**, pages 267–274.

Chittaro, L., Gatla, V., and Venkataraman, S. (2005). The interactive 3d breakaway map: a navigation and examination aid for multi-floor 3d worlds. In **2005 International Conference on Cyberworlds (CW'05)**, pages 8 pp.–66.

Chittaro, L., Ranon, R., and Ieronutti, L. (2003). Guiding visitors of web3d worlds through automatically generated tours. In **Proceedings of the eighth international conference on 3D Web technology**, Web3D03, pages 27–38.

Chung, S., Joh, H., Lee, E., and Oh, U. (2021). Panocue: An efficient visual cue with a omnidirectional panoramic view for finding a target in 3d space. In **2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 218–223.

Chénéchal, M. L., Duval, T., Gouranton, V., Royan, J., and Arnaldi, B. (2015). The Stretchable Arms for Collaborative Remote Guiding. In Imura, M., Figueroa, P., and Mohler, B., editors, **ICAT-EGVE 2015 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments**. The Eurographics Association.

Clarinval, A., Linden, I., Wallemacq, A., and Dumas, B. (2018). Evoq: a Visualization Tool to Support Structural Analysis of Text Documents. In **Proceedings of the ACM Symposium on Document Engineering 2018**, Article 27, pages 1–10.

Cockburn, A., Karlson, A., and Bederson, B. B. (2009). A review of overview+detail, zooming, and focus+context interfaces. **ACM Computing Surveys**, 41(1):1–31.

Cook, K. A. and Thomas, J. J. (2005). Illuminating the Path: The Research and Development Agenda for Visual Analytics. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

Cooper, R. (1989). Modernism, Post Modernism and Organizational Analysis 3: The Contribution of Jacques Derrida. **Organization Studies**, 10(4):479–502.

Czerwinski, M., Tan, D. S., and Robertson, G. G. (2002). Women take a wider view. In **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)**, pages 195–202.

Danyluk, K., Ens, B., Jenny, B., and Willett, W. (2021). A design space exploration of worlds in miniature. In **Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)**, pages 1–15.

Darken, R. and Cevik, H. (1999). Map usage in virtual environments: orientation issues. In **Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)**, VR-99, pages 133–140.

Davies, R. (2015). Industry 4.0: Digitalisation for productivity and growth.

Debernardis, S., Fiorentino, M., Gattullo, M., Monno, G., and Uva, A. E. (2013). Text Readability in Head-Worn Displays: Color and Style Optimization in Video versus Optical See-Through Devices. **IEEE Transactions on Visualization and Computer Graphics**, 20(1):125–139.

Dejardin, G. (2022). CSApp, A Mixed Reality Game for Cybersecurity Awareness. Master's thesis, University of Namur.

Derrida, J. (1967). **L'écriture et la différence**. Editions du Seuil, Paris.

Dewitz, B., Geiger, C., and Steinicke, F. (2021). Virtual Visus-Vision Acuity and Text Legibility in Virtual Environments. In **GI VR/AR Workshop**.

Dey, A., Billinghurst, M., Lindeman, R. W., and Swan, J. E. (2018). A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. **Frontiers in Robotics and AI**, 5.

Dingler, T., Kunze, K., and Outram, B. (2018). VR Reading UIs: Assessing Text Parameters for Reading in VR. In **Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems**.

Duchowski, A. T. (2018). Gaze-based interaction: A 30 year retrospective. **Computers & Graphics**, 73:59–69.

Dünser, A., Billinghurst, M., Wen, J., Lehtinen, V., and Nurminen, A. (2012). Exploring the use of handheld ar for outdoor navigation. **Computers & Graphics**, 36(8):1084–1095.

Eades, P. (1984). A heuristic for graph drawing. **Congressus numerantium**, 42(11):149–160.

Elmqvist, N. and Tsigas, P. (2007). View-projection animation for 3d occlusion management. **Computers & Graphics**, 31(6):864–876.

Elmqvist, N. and Tsigas, P. (2008). A Taxonomy of 3D Occlusion Management for Visualization. **IEEE Transactions on Visualization and Computer Graphics**, 14(5):1095–1109.

Elvins, T. T., Nadeau, D. R., and Kirsh, D. (1997). Worldlets—3d thumbnails for wayfinding in virtual environments. In **Proceedings of the 10th annual ACM symposium on User interface software and technology - UIST '97**, pages 21–30.

Ens, B., Bach, B., Cordeil, M., Engelke, U., Serrano, M., Willett, W., Prouzeau, A., Anthes, C., Büschel, W., Dunne, C., Dwyer, T., Grubert, J., Haga, J. H., Kirshenbaum, N., Kobayashi, D., Lin, T., Olaosebikan, M., Pointecker, F., Saffo, D., Saquib, N., Schmalstieg, D., Szafir, D. A., Whitlock, M., and Yang, Y. (2021). Grand challenges in immersive analytics. In **Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems**, pages 1–17.

Erickson, A., Kim, K., Bruder, G., and Welch, G. (2020a). A review of visual perception research in optical see-through augmented reality. In **International Conference on Artificial Reality and Telexistence, Eurographics Symposium on Virtual Environments (ICAT-EGVE)**, pages 1–9.

Erickson, A., Kim, K., Bruder, G., and Welch, G. F. (2020b). Effects of dark mode graphics on visual acuity and fatigue with virtual reality head-mounted displays. In **2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)**, pages 1–9.

Erickson, A., Kim, K., Lambert, A., Bruder, G., Browne, M. P., and Welch, G. F. (2021). An extended analysis on the benefits of dark mode user interfaces in optical see-through head-mounted displays. **ACM Transactions on Applied Perception**, 18(3):1–22.

Evangelista, A., Manghisi, V. M., Laera, F., Gattullo, M., Uva, A. E., and Fiorentino, M. (2021). **CompassbAR: A Technique for Visualizing Out-of-View Objects in a Mixed Reality Environment**, pages 141–148.

Falk, J., Eksvard, S., Schenkman, B., Andren, B., and Brunnstrom, K. (2021). Legibility and readability in augmented reality. In **2021 13th International Conference on Quality of Multimedia Experience (QoMEX)**, pages 231–236.

Farr, A. C., Kleinschmidt, T., Yarlagadda, P., and Mengersen, K. (2012). Wayfinding: A simple concept, a complex process. **Transport Reviews**, 32(6):715–743.

Feiner, S. and Shamash, A. (1991). Hybrid user interfaces: Breeding virtually bigger interfaces for physically smaller computers. In **Proceedings of the 4th annual ACM symposium on User interface software and technology - UIST '91**, pages 9–17.

Feiner, S. K. and Seligmann, D. D. (1992). Cutaways and ghosting: satisfying visibility constraints in dynamic 3D illustrations. **The Visual Computer**, 8(5-6):292–302.

Fiorentino, M., Debernardis, S., Uva, A. E., and Monno, G. (2013). Augmented Reality Text Style Readability with See-Through Head-Mounted Displays in Industrial Context. **Presence: Teleoperators and Virtual Environments**, 22(2):171–190.

Forsell, C. and Johansson, J. (2010). An heuristic set for evaluation in information visualization. In **Proceedings of the International Conference on Advanced Visual Interfaces**, AVI ’10, pages 199–206.

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. **Perception & Psychophysics**, 8(4):215–221.

Fowler, J. J. and Kobourov, S. G. (2012). Planar Preprocessing for Spring Embedders. In **Proceedings of the 20th international conference on Graph Drawing**, pages 388–399.

Fuchs, H., State, A., and Bazin, J.-C. (2014). Immersive 3D Telepresence. **Computer**, 47(7):46–52.

Fukatsu, S., Kitamura, Y., Masaki, T., and Kishino, F. (1998). Intuitive control of “bird's eye” overview images for navigation in an enormous virtual environment. In **Proceedings of the ACM symposium on Virtual reality software and technology**, pages 67–76.

Fukushima, S., Hamada, T., and Hautasaari, A. (2020). Comparing World and Screen Coordinate Systems in Optical See-Through Head-Mounted Displays for Text Readability while Walking. In **2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)**, pages 649–658.

Funk, M., Boldt, R., Pfleging, B., Pfeiffer, M., Henze, N., and Schmidt, A. (2014). Representing indoor location of objects on wearable computers with head-mounted displays. In **Proceedings of the 5th Augmented Human International Conference (AH ’14)**, pages 1–4.

Funk, M., Kosch, T., Kettner, R., Korn, O., and Schmidt, A. (2016). motionEAP: An Overview of 4 Years of Combining Industrial Assembly with Augmented Reality for Industry 4.0. In **Proceedings of the 16th International Conference on Knowledge Technologies and Data-driven Business**, pages 1–4.

Gabbard, J. L., Mehra, D. G., and Swan, J. E. (2019). Effects of AR Display Context Switching and Focal Distance Switching on Human Performance. **IEEE Transactions on Visualization and Computer Graphics**, 25(6):2228–2241.

Gabbard, J. L., Swan, J. E., and Hix, D. (2006). The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. **Presence: Teleoperators and Virtual Environments**, 15(1):16–32.

Gabbard, J. L., Swan, J. E., Hix, D., Kim, S.-J., and Fitch, G. (2007). Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications. In **2007 IEEE Virtual Reality Conference**, pages 35–42.

Gabbard, J. L., Swan, J. E., Zedlitz, J., and Winchester, W. W. (2010). More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In **2010 IEEE Virtual Reality Conference (VR)**, pages 79–86.

Gao, L., Bai, H., Lindeman, R., and Billinghurst, M. (2017). Static local environment capturing and sharing for MR remote collaboration. In **SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications on - SA '17**, Article 17, pages 1–6.

Garousi, V., Felderer, M., and Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. **Information and Software Technology**, 106:101–121.

Gattullo, M., Dammacco, L., Ruospo, F., Evangelista, A., Fiorentino, M., Schmitt, J., and Uva, A. E. (2020). Design preferences on industrial augmented reality: a survey with potential technical writers. In **2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**. IEEE.

Gattullo, M., Evangelista, A., Uva, A. E., Fiorentino, M., and Gabbard, J. L. (2022). What, How, and Why are Visual Assets Used in Industrial Augmented Reality? A

Systematic Review and Classification in Maintenance, Assembly, and Training (From 1997 to 2019). **IEEE Transactions on Visualization and Computer Graphics**, 28(2):1443–1456.

Gattullo, M., Scurati, G. W., Fiorentino, M., Uva, A. E., Ferrise, F., and Bordegoni, M. (2019). Towards augmented reality manuals for industry 4.0: A methodology. **Robotics and Computer-Integrated Manufacturing**, 56:276–286.

Gattullo, M., Uva, A. E., Fiorentino, M., and Gabbard, J. L. (2015). Legibility in Industrial AR: Text Style, Color Coding, and Illuminance. **IEEE Computer Graphics and Applications**, 35(2):52–61.

Gattullo, M., Uva, A. E., Fiorentino, M., and Monno, G. (2014). Effect of Text Outline and Contrast Polarity on AR Text Readability in Industrial Lighting. **IEEE Transactions on Visualization and Computer Graphics**, 21(5):638–651.

Gattullo, M., Uva, A. E., Fiorentino, M., Scurati, G. W., and Ferrise, F. (2017). From Paper Manual to AR Manual: Do We Still Need Text? **Procedia Manufacturing**, 11:1303–1310.

Gauglitz, S., Nuernberger, B., Turk, M., and Höllerer, T. (2014a). In touch with the remote world: remote collaboration with augmented reality drawings and virtual navigation. In **Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology**, pages 197–205.

Gauglitz, S., Nuernberger, B., Turk, M., and Höllerer, T. (2014b). World-stabilized annotations and virtual scene navigation for remote collaboration. In **Proceedings of the 27th annual ACM symposium on User interface software and technology**, pages 449–459.

Gay-Bellile, V., Lothe, P., Bourgeois, S., Royer, E., and Naudet Collette, S. (2010). Augmented reality in large environments: Application to aided navigation in urban context. In **2010 IEEE International Symposium on Mixed and Augmented Reality**, pages 225–226.

Gebhardt, C., Hecox, B., van Opheusden, B., Wigdor, D., Hillis, J., Hilliges, O., and Benko, H. (2019). Learning cooperative personalized policies from gaze data. In **Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology**, pages 197–208.

Gershon, N. and Eick, S. (1995). Foreword of the conference. In **IEEE Symposium Information Visualization (InfoVis)**, pages vii–viii.

Ghani, S., Riche, N. H., and Elmqvist, N. (2011). Dynamic insets for context-aware graph navigation. **Computer Graphics Forum**, 30(3):861–870.

Gibson, H., Faith, J., and Vickers, P. (2013). A survey of two-dimensional graph layout techniques for information visualisation. **Information Visualization**, 12(3-4):324–357.

Gonçalves, T., Afonso, A. P., Carmo, M. B., and Pombinho, P. (2011a). Evaluation of HaloDot: Visualization of relevance of off-screen objects with over cluttering prevention on mobile devices. In **Human-Computer Interaction – INTERACT 2011**, pages 300–308.

Gonçalves, T., Afonso, A. P., Carmo, M. B., and Pombinho, P. (2011b). Halodot: Visualization of the relevance of off-screen objects. **Proceedings of SIACG**, pages 117–120.

Gonzalez-Calleros, J. M., Guerrero-Garcia, J., Vanderdonckt, J., and Munoz-Arteaga, J. (2009). Towards Canonical Task Types for User Interface Design. In **2009 Latin American Web Congress**, pages 63–70.

Gonçalves, T., Afonso, A. P., Carmo, M. B., and Pombinho, P. (2013). Comparison of off-screen visualization techniques with representation of relevance on mobile devices. In **Electronic Workshops in Computing**, pages 1–10.

Google Fonts (2023). Using type in AR & VR. https://fonts.google.com/knowledge/using_type_in_ar_and_vr [Accessed 19 June 2023].

Google for Developers (2017). Designing screen interfaces for vr (google i/o '17). https://www.youtube.com/watch?v=ES9jArHRFHQ&ab_channel=GoogleforDevelopers [Accessed 19 June 2023].

Grout, C., Rogers, W., Apperley, M., and Jones, S. (2015). Reading text in an immersive head-mounted display: An investigation into displaying desktop interfaces in a 3D virtual environment. In **Proceedings of the 15th New Zealand Conference on Human-Computer Interaction**.

Grubert, J. (2021). Mixed reality interaction techniques. https://arxiv.org/abs/2103.05984.

Grubert, J., Langlotz, T., Zollmann, S., and Regenbrecht, H. (2017). Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. **IEEE Transactions on Visualization and Computer Graphics**, 23(6):1706–1724.

Gruenefeld, U., Ali, A. E., Boll, S., and Heuten, W. (2018a). Beyond halo and wedge: visualizing out-of-view objects on head-mounted virtual and augmented reality devices. In **Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services**, pages 1–11.

Gruenefeld, U., Ali, A. E., Heuten, W., and Boll, S. (2017a). Visualizing out-of-view objects in head-mounted augmented reality. In **Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services**, MobileHCI '17, pages 1–7.

Gruenefeld, U., Ennenga, D., Ali, A. E., Heuten, W., and Boll, S. (2017b). Eyesee360: designing a visualization technique for out-of-view objects in head-mounted augmented reality. In **Proceedings of the 5th Symposium on Spatial User Interaction**, SUI '17, pages 109–118.

Gruenefeld, U., Koethe, I., Lange, D., Weiss, S., and Heuten, W. (2019a). Comparing techniques for visualizing moving out-of-view objects in head-mounted virtual reality. In **2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)**, pages 742–746.

Gruenefeld, U., Lange, D., Hammer, L., Boll, S., and Heuten, W. (2018b). Flyingarrow: Pointing towards out-of-view objects on augmented reality devices. In **Proceedings of the 7th ACM International Symposium on Pervasive Displays (PerDis ’18)**, pages 1–6.

Gruenefeld, U., Prädel, L., and Heuten, W. (2019b). Improving search time performance for locating out-of-view objects in augmented reality. In **Proceedings of Mensch und Computer 2019 (MuC ’19)**, pages 481–485.

Gugenheimer, J., McGill, M., Huron, S., Mai, C., Williamson, J., and Nebeling, M. (2020). Exploring Potentially Abusive Ethical, Social and Political Implications of Mixed Reality Research in HCI. In **Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems**.

Gupta, K., Lee, G. A., and Billinghurst, M. (2016). Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. **IEEE Transactions on Visualization and Computer Graphics**, 22(11):2413–2422.

Gustafson, S., Baudisch, P., Gutwin, C., and Irani, P. (2008). Wedge: clutter-free visualization of off-screen locations. In **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**, pages 787–796.

Gustafson, S. G. and Irani, P. P. (2007). Comparing visualizations for tracking off-screen moving targets. In **CHI '07 Extended Abstracts on Human Factors in Computing Systems**, pages 2239–2404.

Gutwin, C. and Greenberg, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. **Computer Supported Cooperative Work (CSCW)**, 11(3-4):411–446.

Günther, S., Kratz, S., Avrahami, D., and Mühlhäuser, M. (2018). Exploring Audio, Visual, and Tactile Cues for Synchronous Remote Assistance. In **Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference**, pages 339–344.

Harada, Y. and Ohyama, J. (2021). Quantitative evaluation of visual guidance effects for 360-degree directions. **Virtual Reality**, 26(2):759–770.

Hartson, R. and Pyla, P. S. (2012a). Chapter 1: Introduction. In **The UX Book: Process and guidelines for ensuring a quality user experience**, pages 1–46. Morgan Kaufmann.

Hartson, R. and Pyla, P. S. (2012b). Chapter 13: Rapid evaluation methods. In **The UX Book: Process and guidelines for ensuring a quality user experience**, pages 467–501. Morgan Kaufmann, Boston.

Heilig, M. (1961). Sensorama simulator. **U.S. Patent 3 050 870A**.

Hein, P., Bernhagen, M., and Bullinger, A. C. (2020). Two is better than one. improved attention guiding in AR by combining techniques. **IEEE Computer Graphics and Applications**, 40(5):57–66.

Henderson, S. J. and Feiner, S. (2008). Opportunistic controls: Leveraging natural affordances as tangible user interfaces for augmented reality. In **Proceedings of the 2008 ACM symposium on Virtual reality software and technology**, pages 211–218.

Hendrickson, B. and Leland, R. (1995). Proceedings of the 1995 acm/ieee conference on supercomputing (cdrom) - supercomputing '95. In **A multilevel algorithm for partitioning graphs**, pages 28–es, New York, New York, USA. ACM Press.

Henze, N. and Boll, S. (2010). Push the study to the app store: evaluating off-screen visualizations for maps in the android market. In **Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (MobileHCI '10)**, pages 373–374.

Henze, N., Poppinga, B., and Boll, S. (2010). Experiments in the wild: public evaluation of off-screen visualizations in the android market. In **Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10)**, pages 675–678.

Hertel, J., Karaosmanoglu, S., Schmidt, S., Braker, J., Semmann, M., and Steinicke, F. (2021). A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review. In **2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)**. IEEE.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design Science in Information Systems Research. **Management Information Systems Quarterly**, 28(1):5180–5190.

Higuch, K., Yonetani, R., and Sato, Y. (2016). Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**. ACM.

Hincapie-Ramos, J. D., Ivanchuk, L., Sridharan, S. K., and Irani, P. (2014). SmartColor: Real-Time Color Correction and Contrast for Optical See-Through Head-Mounted Displays. In **2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)**.

Hincapie-Ramos, J. D., Ivanchuk, L., Sridharan, S. K., and Irani, P. P. (2015). SmartColor: Real-Time Color and Contrast Correction for Optical See-Through Head-Mounted Displays. **IEEE Transactions on Visualization and Computer Graphics**, 21(12):1336–1348.

Hincq, L. (2022). Application de technique de réalité augmentée dans la maintenance industrielle. Master's thesis, University of Namur.

Hoffman, D. M., McKenzie, C., Koprowski, B., Iqbal, A., and Balram, N. (2019). Aligning content rendering resolution and feature size with display capability in near-eye display systems. **Journal of the Society for Information Display**, 27(4):207–222.

Hu, S., Malloch, J., and Reilly, D. (2021). A comparative evaluation of techniques for locating out-of-view targets in virtual reality. **Proceedings of Graphics Interface 2021**.

Huang, W., Billinghurst, M., Alem, L., and Kim, S. (2018). HandsInTouch: sharing gestures in remote collaboration. In **Proceedings of the 30th Australian Conference on Computer-Human Interaction**, pages 396–400.

Human Performance Research Group (1980). Nasa task load index. Technical report, NASA Ames Research Center, Moffett Field, California.

Höllerer, T., Feiner, S., Hallaway, D., Bell, B., Lanzagorta, M., Brown, D., Julier, S., Baillot, Y., and Rosenblum, L. (2001). User interface management techniques for collaborative mobile augmented reality. **Computers & Graphics**, 25(5):799–810.

International Standards Organization (2018). ISO 9241-11:2018 Ergonomics of human-system interaction, Part 11: Usability: Definitions and concepts. Technical report, International Standards Organization.

Jaakkola, E. (2020). Designing conceptual articles: four approaches. **AMS Review**, 10(1–2):18–26.

Jo, H., Hwang, S., Park, H., and hee Ryu, J. (2011). Aroundplot: Focus+context interface for off-screen objects in 3d environments. **Computers & Graphics**, 35(4):841–853.

Joachimczak, M., Liu, J., and Ando, H. (2017). Real-time mixed-reality telepresence via 3D reconstruction with HoloLens and commodity depth sensors. In **Proceedings of the 19th ACM International Conference on Multimodal Interaction**, pages 514–515.

Joshi, N., Lakier, M., Vogel, D., and Zhao, J. (2022). A Design Framework for Contextual and Embedded Information Visualizations in Spatial Augmented Reality. In **Proceedings of Graphics Interface**.

Joshi, R., Hiwale, A., Birajdar, S., and Gound, R. (2019). Indoor navigation with augmented reality. In **Lecture Notes in Electrical Engineering**, pages 159–165.

Julier, S., Lanzagorta, M., Baillot, Y., Rosenblum, L., Feiner, S., Hollerer, T., and Sestito, S. (2000). Information filtering for mobile augmented reality. In **Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)**, pages 3–11.

Kalkofen, D., Tatzgern, M., and Schmalstieg, D. (2009). Explosion Diagrams in Augmented Reality. In **2009 IEEE Virtual Reality Conference**, pages 71–78. IEEE.

Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., and Tachibana, K. (2000). Virtual object manipulation on a table-top AR environment. In **Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)**, pages 111–119.

Kaufeld, M., Mundt, M., Forst, S., and Hecht, H. (2022). Optical see-through augmented reality can induce severe motion sickness. **Displays**, 74:102283.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., and Melançon, G. (2008). **Visual Analytics: Definition, Process, and Challenges**, pages 154–175. Springer Berlin Heidelberg, Berlin, Heidelberg.

Khan, A., Matejka, J., Fitzmaurice, G., and Kurtenbach, G. (2005). Spotlight: directing users' attention on large displays. In **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**, pages 791–798.

Kijima, R. and Ojika, T. (1997). Transition between virtual environment and workstation environment with projective head mounted display. In **Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality**, pages 130–137.

Kim, K., Billinghurst, M., Bruder, G., Duh, H. B.-L., and Welch, G. F. (2018). Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017). **IEEE Transactions on Visualization and Computer Graphics**, 24(11):2947–2962.

Kim, K., Erickson, A., Lambert, A., Bruder, G., and Welch, G. (2019a). Effects of Dark Mode on Visual Fatigue and Acuity in Optical See-Through Head-Mounted Displays. In **Symposium on Spatial User Interaction**, Article 9, pages 1–9.

Kim, S., Lee, G., Huang, W., Kim, H., Woo, W., and Billinghurst, M. (2019b). Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration. In **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems**, pages 1–13.

Kim, S., Lee, G., Sakata, N., and Billinghurst, M. (2014). Improving co-presence with augmented visual communication cues for sharing experience through video conference. In **2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)**, pages 83–92.

Kitchenham, B. A., Budgen, D., and Brereton, P. (2015). **Evidence-Based Software Engineering and Systematic Reviews**. Chapman and Hall/CRC.

Kleiber, M., Alexander, T., Winkelholz, C., and Schlick, C. M. (2012). User-centered design and evaluation of an integrated AR-VR system for tele-maintenance. In **2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)**, pages 1443–1448. IEEE.

Klose, E. M., Mack, N. A., Hegenberg, J., and Schmidt, L. (2019). Text Presentation for Augmented Reality Applications in Dual-Task Situations. In **2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)**, pages 636–644.

Kobayashi, S., Kanari, K., and Sato, M. (2021). An Examination of View-Settings for Long Texts in VR Reading. In **Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters**, Article 10, pages 1–2.

Kobayashi, S., Kanari, K., and Sato, M. (2022). Effects of Font Type and Weight on Reading in VR. In **SIGGRAPH Asia 2022 Posters**.

Kobourov, S. G. (2013). Force-Directed Drawing Algorithms. In Tamassia, R., editor, **Handbook of Graph Drawing and Visualization**, pages 398–423. Chapman and Hall/CRC.

Koide, H., Kanari, K., and Sato, M. (2022). An Examination on Reduction of Displayed Character Shake while Walking in Place with AR Glasses. In **2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)**, pages 786–787.

Kojic, T., Ali, D., Greinacher, R., Moller, S., and Voigt-Antons, J.-N. (2020). User Experience of Reading in Virtual Reality — Finding Values for Text Distance, Size and Contrast. In **2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)**, pages 1–6.

Kojic, T., Vergari, M., Möller, S., and Voigt-Antons, J.-N. (2022). Assessing User Experience of Text Readability with Eye Tracking in Virtual Reality. In **Virtual, Augmented and Mixed Reality: Design and Development**, pages 199–211.

Kolbe, T. H. (2004). Augmented videos and panoramas for pedestrian navigation. In **Proceedings of the 2nd Symposium on Location Based Services & TeleCartography 2004**.

Korkut, E. H. and Surer, E. (2023). Visualization in virtual reality: a systematic review. **Virtual Reality**, 27(2):1447–1480.

Kraus, M., Schafer, H., Meschenmoser, P., Schweitzer, D., Keim, D. A., Sedlmair, M., and Fuchs, J. (2020). A comparative study of orientation support tools in virtual reality environments with virtual teleportation. In **2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)**, pages 227–238.

Kruijff, E., Orlosky, J., Kishishita, N., Trepkowski, C., and Kiyokawa, K. (2019). The Influence of Label Design on Search Performance and Noticeability in Wide Field of View Augmented Reality Displays. **IEEE Transactions on Visualization and Computer Graphics**, 25(9):2821–2837.

Kundu, S. N., Muhammad, N., and Sattar, F. (2017). Using the augmented reality sandbox for advanced learning in geoscience education. In **2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE)**, pages 13–17. IEEE.

Lagasse, M., Labour, M., and Schmitt, D. (2023). REMIND-Contrast : Qualifier l'expérience de visite au musée par les construits personnels des visiteurs. **Revue des Interactions Humaines Médiatisées (RHIM)**, 24(1).

Langlotz, T., Nguyen, T., Schmalstieg, D., and Grasset, R. (2014). Next-Generation Augmented Reality Browsers: Rich, Seamless, and Adaptive. **Proceedings of the IEEE**, 102(2):155–169.

Lanier, J. (1992). Virtual reality: the promise of the future. **Interact Learn Int**, 8(4):275–279.

Latvala, O.-M., Keranen, T., Noponen, S., Lehto, N., Sailio, M., Valta, M., and Olli, P. (2017). Visualizing network events in a muggle friendly way. In **2017 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)**, pages 1–4.

Laviola, E., Gattullo, M., Manghisi, V. M., Fiorentino, M., and Uva, A. E. (2021). Minimal ar: visual asset optimization for the authoring of augmented reality work instructions in manufacturing. **The International Journal of Advanced Manufacturing Technology**, 119(3–4):1769–1784.

LaViola JR., J. J., Kruijff, E., McMahan, R. P., Bowman, D. A., and Poupyrev, I. (2017). **3D User Interfaces: Theory and Practice, 2sd edition**. Pearson Education, Limited.

Lawrence, A., Houghton, J., Thomas, J., and Weldon, P. (2014). Where Is the Evidence? Realising the Value of Grey Literature for Public Policy & Practice, A Discussion Paper. **Swinburne Institute for Social Research, Melbourne, Australia.**

Lee, G., Kim, S., Lee, Y., Dey, A., Piumsomboon, T., Norman, M., and Billinghurst, M. (2017a). [POSTER] Mutually Shared Gaze in Augmented Video Conference. In **2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)**.

Lee, G. A., Teo, T., Kim, S., and Billinghurst, M. (2017b). Mixed reality collaboration through sharing a live panorama. In **SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications on - SA '17**, Article 14, pages 1–4.

Lee, H., Bang, S., and Woo, W. (2023). Effects of coordinate system and position of AR notification while walking. **Virtual Reality**, 27(2):829–848.

Lee, J. Y., Seo, D. W., and Rhee, G. (2008). Visualization and interaction of pervasive services using context-aware augmented reality. **Expert Systems with Applications**, 35(4):1873–1882.

Lejeune, C. (2010). Montrer, calculer, explorer, analyser. ce que l’informatique fait (faire) à l’analyse qualitative. In **Logiciels pour l’analyse Qualitative: Innovations Techniques et Sociales (Recherches Qualitatives)**, volume 9 (Hors Série), pages 15–32.

Leusmann, J. (2021). A literature review on distant object selection methods. Master’s thesis, University of Stuttgart.

Levinew, M., Marchon, I., and Hanley, G. (1984). The placement and misplacement of you-are-here maps. **Environment and Behavior**, 16(2):139–157.

Leykin, A. and Tuceryan, M. (2004). Determining text readability over textured backgrounds in augmented reality systems. In **Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry - VRCAI '04**, pages 436–439.

Li, R. (2016). **Effects of Visual Variables on the Perception of Distance in Off-Screen Landmarks: Size, Color Value, and Crispness**, pages 89–103.

Li, X. and Chen, C.-H. (2021). **Effects of Visual Cue Design and Gender Differences on Wayfinding Using Mobile Devices**, pages 15–24.

Lin, Y.-T., Liao, Y.-C., Teng, S.-Y., Chung, Y.-J., Chan, L., and Chen, B.-Y. (2017). Outside-in: Visualizing out-of-sight regions-of-interest in a 360° video using spatial picture-in-picture previews. In **Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)**, pages 255–265.

Lincoln, P., Welch, G., Nashel, A., Ilie, A., State, A., and Fuchs, H. (2009). Animatronic Shader Lamps Avatars. In **2009 8th IEEE International Symposium on Mixed and Augmented Reality**, pages 27–33.

Linden, I., Wallemacq, A., Dumas, B., Deville, G., Clarinval, A., and Cauz, M. (2020). Text as Semantic Fields: Integration of an Enriched Language Conception in the Text Analysis Tool Evoq®. In **Research Challenges in Information Science (RCIS)**, volume 385, pages 543–548, Cham.

Lindlbauer, D., Feit, A. M., and Hilliges, O. (2019). Context-aware online adaptation of mixed reality interfaces. In **Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology**, pages 147–160.

Lu, F. (2021). [DC] Glanceable AR: Towards an Always-on Augmented Reality Future. In **2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)**, pages 717–718.

Lu, F. and Bowman, D. A. (2021). Evaluating the Potential of Glanceable AR Interfaces for Authentic Everyday Uses. In **2021 IEEE Virtual Reality and 3D User Interfaces (VR)**, pages 768–777.

Lu, F. and Xu, Y. (2022). Exploring Spatial UI Transition Mechanisms with Head-Worn Augmented Reality. In **CHI Conference on Human Factors in Computing Systems**, Article 550, pages 1–16.

Lystbæk, M. N., Pfeuffer, K., Grønbæk, J. E. S., and Gellersen, H. (2022a). Exploring Gaze for Assisting Freehand Selection-based Text Entry in AR. **Proceedings of the ACM on Human-Computer Interaction**, 6(ETRA):1–16.

Lystbæk, M. N., Rosenberg, P., Pfeuffer, K., Grønbæk, J. E., and Gellersen, H. (2022b). Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality. **Proceedings of the ACM on Human-Computer Interaction**, 6(ETRA):1–18.

M. Stanney, K., Hash, P., and Dryer, D. (1995). **The Task, Interaction, and Display (TID) Taxonomy for Human-Virtual Environment Interaction**, pages 499–504. Elsevier.

MacIntyre, B., Hill, A., Rouzati, H., Gandy, M., and Davidson, B. (2011). The Argon AR Web Browser and standards-based AR application environment. In **2011 10th IEEE International Symposium on Mixed and Augmented Reality**, pages 65–74. IEEE.

Maimone, A., Yang, X., Dierk, N., State, A., Dou, M., and Fuchs, H. (2013). General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. In **2013 IEEE Virtual Reality (VR)**, pages 23–26.

Maiti, A., Jadliwala, M., and Weber, C. (2017). Preventing shoulder surfing using randomized augmented reality keyboards. In **2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)**, pages 630–635.

Manghisi, V. M., Gattullo, M., Fiorentino, M., Uva, A. E., Marino, F., Bevilacqua, V., and Monno, G. (2017). Predicting Text Legibility over Textured Digital Backgrounds for a Monocular Optical See-Through Display. **Presence: Teleoperators and Virtual Environments**, 26(1):1–15.

Mann, S. (1999). Mediated reality. **Linux Journal**, 1999(59):5–es.

Marques, B., Carvalho, R., Dias, P., and Santos, B. S. (2019). Pervasive augmented reality for indoor uninterrupted experiences. In **Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers**, pages 141–144.

Marriott, K., Schreiber, F., Dwyer, T., Klein, K., Riche, N. H., Itoh, T., Stuerzlinger, W., and Thomas, B. H., editors (2018). **Immersive Analytics**. Springer International Publishing.

Matsuda, K. (2016). Hyper-reality. http://hyper-reality.co/. http://hyper-reality.co/.

McGuffin, M. J., Tancau, L., and Balakrishnan, R. (2003). Using deformations for browsing volumetric data. In **IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control**, pages 401–408.

Mendes, D., Medeiros, D., Sousa, M., Cordeiro, E., Ferreira, A., and Jorge, J. A. (2017). Design and evaluation of a novel out-of-reach selection technique for VR using iterative refinement. **Computers & Graphics**, 67:95–102.

Meurice, L. and Cleve, A. (2016). DAHLIA 2.0: A Visual Analyzer of Database Usage in Dynamic and Heterogeneous Systems. In **2016 IEEE Working Conference on Software Visualization (VISSOFT)**, pages 76–80. IEEE.

Miau, D. and Feiner, S. (2016a). Personalized compass: A compact visualization for direction and location. In **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**, CHI'16. ACM.

Miau, D. and Feiner, S. (2016b). Personalized compass: A demonstration of a compact visualization for direction and location. In **Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI '16)**, pages 3731–3734.

Microsoft (2023a). Color, light, and materials. https://learn.microsoft.com/en-us/windows/mixed-reality/design/color-light-and-materials [Accessed 19 June 2023].

Microsoft (2023b). Comfort. https://learn.microsoft.com/en-us/windows/mixed-reality/design/comfort [Accessed 19 June 2023].

Microsoft (2023c). Typography. https://learn.microsoft.com/en-us/windows/mixed-reality/design/typography [Accessed 19 June 2023].

Microsoft (2023d). Understanding performance for mixed reality. https://learn.microsoft.com/en-us/windows/mixed-reality/develop/advanced-concepts/understanding-performance-for-mixed-reality [Accessed 2 September 2023].

Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. **IEICE Transactions on Information and Systems**, E77-D(12):1321–1329.

Miyashita, T., Meier, P., Tachikawa, T., Orlic, S., Eble, T., Scholz, V., Gapel, A., Gerl, O., Arnaudov, S., and Lieberknecht, S. (2008). An augmented reality museum guide. In **2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality**, pages 103–106.

Munafo, J., Diedrick, M., and Stoffregen, T. A. (2016). The virtual reality head-mounted display oculus rift induces motion sickness and is sexist in its effects. **Experimental Brain Research**, 235(3):889–901.

Munzner, T. (2015). **Visualization Analysis and Design**. AK Peters Visualization Series. CRC Press.

Narzt, W., Pomberger, G., Ferscha, A., Kolb, D., Müller, R., Wieghardt, J., Hörtner, H., and Lindinger, C. (2005). Augmented reality navigation systems. **Universal Access in the Information Society**, 4(3):177–187.

Newman, J., Bornik, A., Pustka, D., Echtler, F., Huber, M., Schmalstieg, D., and Klinker, G. (2007). Tracking for distributed mixed reality environments. In **Proceedings of the IEEE Virtual Reality Workshop on Trends and Issues in Tracking for Virtual Environments**.

Oculus Developers (2023a). Vr accessibility design: Captions and subtitles. https://developer.oculus.com/resources/design-accessible-vr-captions/ [Accessed 19 June 2023].

Oculus Developers (2023b). VR Accessibility Design: User Experience (UX) and User Interface (UI). https://developer.oculus.com/resources/design-accessible-vr-ui-ux/#text [Accessed 19 June 2023].

Oda, O., Elvezio, C., Sukan, M., Feiner, S., and Tversky, B. (2015). Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In **Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology**.

Ong, S. K., Yuan, M. L., and Nee, A. Y. C. (2008). Augmented reality applications in manufacturing: a survey. **International Journal of Production Research**, 46(10):2707–2742.

Orlosky, J., Kiyokawa, K., and Takemura, H. (2013). Towards intelligent view management: A study of manual text placement tendencies in mobile environments using video see-through displays. In **2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)**, pages 281–282.

Orlosky, J., Kiyokawa, K., Toyama, T., and Sonntag, D. (2015). Halo Content: Context-aware Viewspace Management for Non-invasive Augmented Reality. In **Proceedings of the 20th International Conference on Intelligent User Interfaces**, pages 369–373.

Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P. L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P. A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S. B., Kohli, P., Lutchyn, Y., Keskin, C., and Izadi, S. (2016). Holoportation: Virtual 3D Teleportation in Real-time. In **Proceedings of the 29th Annual Symposium on User Interface Software and Technology**, pages 741–754, New York, NY, USA.

Osmers, N. and Prilla, M. (2020). Getting out of Out of Sight: Evaluation of AR Mechanisms for Awareness and Orientation Support in Occluded Multi-Room Settings. In **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI ’20)**, pages 1–11.

Oulasvirta, A., Estlander, S., and Nurminen, A. (2008). Embodied interaction with a 3d versus 2d mobile map. **Personal and Ubiquitous Computing**, 13(4):303–320.

O’Hagan, J., Gugenheimer, J., Bonner, J., Mathis, F., and McGill, M. (2023). Augmenting People, Places & Media: The Societal Harms Posed by Everyday Augmented Reality, and the Case for Perceptual Human Rights. In **Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia**, MUM ’23.

Palmarini, R., Erkoyuncu, J. A., and Roy, R. (2017). An Innovative Process to Select Augmented Reality (AR) Technology for Maintenance. **Procedia CIRP**, 59:23–28.

Pampel, S. M., Lamb, K., Burnett, G., Skrypchuk, L., Hare, C., and Mouzakitis, A. (2018). An investigation of the effects of driver age when using novel navigation systems in a head-up display. **Presence: Teleoperators and Virtual Environments**, 27(1):32–45.

Perea, P., Morand, D., and Nigay, L. (2017). Halo3D: a technique for visualizing off-screen points of interest in mobile augmented reality. In **Proceedings of the 29th Conference on l'Interaction Homme-Machine**, pages 43–51.

Perea, P., Morand, D., and Nigay, L. (2019). Spotlight on off-screen points of interest in handheld augmented reality. In **Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces**, pages 43–54.

Petford, J., Carson, I., Nacenta, M. A., and Gutwin, C. (2019). A comparison of notification techniques for out-of-view objects in full-coverage displays. In **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19)**, Article 58, pages 1–13.

Piret, A., Nizet, J., and Bourgeois, E. (1996). **L'analyse structurale: une méthode d'analyse de contenu pour les sciences humaines**. De Boaeck Supérieur, Brussels.

Piumsomboon, T., Dey, A., Ens, B., Lee, G., and Billinghurst, M. (2019a). The effects of sharing awareness cues in collaborative mixed reality. **Frontiers in Robotics and AI**, 6.

Piumsomboon, T., Lee, G. A., Hart, J. D., Ens, B., Lindeman, R. W., Thomas, B. H., and Billinghurst, M. (2018). Mini-me: An adaptive avatar for mixed reality remote collaboration. In **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems**, pages 1–13.

Piumsomboon, T., Lee, G. A., Irlitti, A., Ens, B., Thomas, B. H., and Billinghurst, M. (2019b). On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems**, pages 1–17.

Plopski, A., Hirzle, T., Norouzi, N., Qian, L., Bruder, G., and Langlotz, T. (2022). The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-worn Extended Reality. **ACM Computing Surveys**, 55(3):1–39.

Poupyrev, I. and Ichikawa, T. (1999). Manipulating Objects in Virtual Worlds: Categorization and Empirical Evaluation of Interaction Techniques. **Journal of Visual Languages & Computing**, 10(1):19–35.

Prouzeau, A., Wang, Y., Ens, B., Willett, W., and Dwyer, T. (2020). Corsican Twin: Authoring In Situ Augmented Reality Visualisations in Virtual Reality. In **Proceedings of the International Conference on Advanced Visual Interfaces**, pages 1–9. ACM.

Psychoactive agency (2023). Mediated reality. https://mediatedreality.info/.

Radkowski, R., Herrema, J., and Oliver, J. (2015). Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. **International Journal of Human-Computer Interaction**, 31(5):337–349.

Rasmussen, T., Feuchtner, T., Huang, W., and Grønbæk, K. (2022). Supporting workspace awareness in remote assistance through a flexible multi-camera system and augmented reality awareness cues. **Journal of Visual Communication and Image Representation**, 89:103655.

Rasmussen, T. A. and Gronbak, K. (2019). Tailorable Remote Assistance with RemoteAssistKit: A Study of and Design Response to Remote Assistance in the Manufacturing Industry. In **Collaboration Technologies and Social Computing**, pages 80–95.

Rauschnabel, P. A., Felix, R., Heller, J., and Hinsch, C. (2024). The 4C framework: Towards a holistic understanding of consumer engagement with augmented reality. **Computers in Human Behavior**, 154:108105.

Regenbrecht, H., Zwanenburg, S., and Langlotz, T. (2022). Pervasive Augmented Reality—Technology and Ethics. **IEEE Pervasive Computing**, 21(3):84–91.

Renkewitz, H., Kinder, V., Brandt, M., and Alexander, T. (2008). Optimal Font Size for Head-Mounted-Displays in Outdoor Applications. In **2008 12th International Conference Information Visualisation**, pages 503–508.

Renner, P., Blattgerste, J., and Pfeiffer, T. (2018). A path-based attention guiding technique for assembly environments with target occlusions. In **2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)**, pages 671–672.

Renner, P. and Pfeiffer, T. (2017a). Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In **2017 IEEE Symposium on 3D User Interfaces (3DUI)**, pages 186–194.

Renner, P. and Pfeiffer, T. (2017b). Evaluation of attention guiding techniques for augmented reality-based assistance in picking and assembly tasks. In **Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion (IUI '17)**, pages 89–92.

Renner, P. and Pfeiffer, T. (2017c). [poster] augmented reality assistance in the central field-of-view outperforms peripheral displays for order picking: Results from a virtual reality simulation study. In **2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)**, pages 176–181.

Renner, P. and Pfeiffer, T. (2020). Ar-glasses-based attention guiding for complex environments: requirements, classification and evaluation. In **Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '20)**, pages 1–10.

Rhyne, T.-M. (2003). Does the difference between information and scientific visualization really matter? **IEEE Computer Graphics and Applications**, 23(3):6–8.

Rhyne, T.-M. (2008). Visualization and the larger world of computer graphics. In **ACM SIGGRAPH 2008 classes**, Article 97, pages 1–4.

Rolland, J. P. and Fuchs, H. (2000). Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization. **Presence: Teleoperators and Virtual Environments**, 9(3):287–309.

Rosilius, M., Wirsing, B., von Eitzen, I., Wilhelm, M., Schmitt, J., Engelmann, B., and Brautigam, V. (2021). Evaluation of Visual Requirements and Software-Design for Immersive Visibility in Industrial Applications. In **2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 234–239.

Rothe, S., Buschek, D., and Hußmann, H. (2019). Guidance in Cinematic Virtual Reality-Taxonomy, Research Status and Challenges. **Multimodal Technologies and Interaction**, 3(1):19.

Rovira, A., Fatah gen Schieck, A., Blume, P., and Julier, S. (2020). Guidance and surroundings awareness in outdoor handheld augmented reality. **PLOS ONE**, 15(3):e0230518.

Rzayev, R., Ugnivenko, P., Graf, S., Schwind, V., and Henze, N. (2021). Reading in VR: The Effect of Text Presentation Type and Location. In **Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems**, Article 531, pages 1–10.

Rzayev, R., Woźniak, P. W., Dingler, T., and Henze, N. (2018). Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking. In **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems**, pages 1–9.

Sanchez, C. R., Fernandez, F. J. G., Vena, L. C. S., Carpio, J., and Castro, M. (2011). Industrial Telemaintenance: Remote Management Experience From Subway to Industrial Electronics. **IEEE Transactions on Industrial Electronics**, 58(3):1044–1051.

Sanfilippo, U. (2023). État de l'art: Visualization de graphes 3D en réalité augmentée. Unpublished technical report, University of Namur.

Scharff, L. F. V., Albert J. Ahumada, J., and Hill, A. L. (1999). Discriminability measures for predicting readability. In Rogowitz, B. E. and Pappas, T. N., editors, **Human Vision and Electronic Imaging IV**, volume 3644, pages 270–277.

Schinke, T., Henze, N., and Boll, S. (2010). Visualization of off-screen objects in mobile augmented reality. In **Proceedings of the 12th international conference on Human computer interaction with mobile devices and services**, pages 313–316.

Schlosser, P. D., Matthews, B., and Sanderson, P. M. (2021). Head-worn displays for healthcare and industry workers: A review of applications and design. **International Journal of Human-Computer Studies**, 154:102628.

Schmalstieg, D. and Hollerer, T. (2016). **Augmented reality: principles and practice**. Addison-Wesley Professional.

Schopfel, J. and Farace, D. (2010). Grey literature. In Press, C., editor, **Encyclopedia of Library and Information Sciences, Third Edition**, chapter Grey literature. CRC Press.

Schrepp, M. (2023). User experience questionnaire handbook. Technical report, Team UEQ.

Schwajda, D., Friedl, J., Pointecker, F., Jetter, H.-C., and Anthes, C. (2023). Transforming graph data visualisations from 2D displays into augmented reality 3D space: A quantitative study. **Frontiers in Virtual Reality**, 4.

Schwerdtfeger, B., Frimor, T., Pustka, D., and Klinker, G. (2006). **Mobile Information Presentation Schemes for Supra-adaptive Logistics Applications**, pages 998–1007.

Schwerdtfeger, B. and Klinker, G. (2008). Supporting order picking with augmented reality. In **2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality**, pages 91–94.

Schwerdtfeger, B., Reif, R., Günthner, W. A., and Klinker, G. (2011). Pick-by-vision: there is something to pick at the end of the augmented tunnel. **Virtual Reality**, 15(2–3):213–223.

Schwind, V., Reinhardt, J., Rzayev, R., Henze, N., and Wolf, K. (2018). Virtual reality on the go?: a study on social acceptance of VR glasses. In **Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct**, MobileHCI '18, pages 111–118.

Scurati, G. W., Gattullo, M., Fiorentino, M., Ferrise, F., Bordegoni, M., and Uva, A. E. (2018). Converting maintenance actions into standard symbols for augmented reality applications in industry 4.0. **Computers in Industry**, 98:68–79.

Septon, T., Cauz, M., and Dumas, B. (2023). Extending the World-In-Miniature Metaphor to Access Situated Information in a Pervasive Augmented Reality Environment. In **IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 543–548.

Seraji, M. R. and Stuerzlinger, W. (2022a). HybridAxes: An Immersive Analytics Tool With Interoperability Between 2D and Immersive Reality Modes. In **2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 155–160.

Seraji, M. R. and Stuerzlinger, W. (2022b). XVCollab: An Immersive Analytics Tool for Asymmetric Collaboration across the Virtuality Spectrum. In **2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 146–154.

Sereno, M., Wang, X., Besancon, L., Mcguffin, M. J., and Isenberg, T. (2022). Collaborative Work in Augmented Reality: A Survey. **IEEE Transactions on Visualization and Computer Graphics**, 28(6):2530–2549.

Shimizu, Y., Ohnishi, A., Terada, T., and Tsukamoto, M. (2021). Gaze-Adaptive Subtitles Considering the Balance among Vertical/Horizontal and Depth of Eye Movement. In **2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)**, pages 127–132.

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In **Proceedings 1996 IEEE Symposium on Visual Languages**, VL-96, pages 336–343.

Si-Mohammed, H., Petit, J., Jeunet, C., Argelaguet, F., Spindler, F., Evain, A., Roussel, N., Casiez, G., and Lecuyer, A. (2020). Towards BCI-Based Interfaces for Augmented Reality: Feasibility, Design and Evaluation. **IEEE Transactions on Visualization and Computer Graphics**, 26(3):1608–1621.

Singh, K. and Balakrishnan, R. (2004). Visualizing 3d scenes using non-linear projections and data mining of previous camera movements. In **Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa**, pages 41–48. ACM.

Skarbez, R., Smith, M., and Whitton, M. C. (2021). Revisiting Milgram and Kishino's Reality-Virtuality Continuum. **Frontiers in Virtual Reality**, 2.

Sridharan, S. K., Hincapié-Ramos, J. D., Flatla, D. R., and Irani, P. (2013). Color correction for optical see-through displays using display color profiles. In **Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology**, pages 231–240.

Stafford, A., Piekarski, W., and Thomas, B. (2006). Implementation of god-like interaction techniques for supporting collaboration between outdoor AR and indoor tabletop users. In **2006 IEEE/ACM International Symposium on Mixed and Augmented Reality**, pages 165–172.

Stoakley, R., Conway, M. J., and Pausch, R. (1995). Virtual reality on a wim: Interactive worlds in miniature. In **CHI'9 5 MOSAIC OF CREATIVITY**.

Suomela, R. and Lehikoinen, J. (2000). Context compass. In **Digest of Papers. Fourth International Symposium on Wearable Computers**.

Sutherland, I. (1968). A head-mounted three dimensional display. In **Proceedings of the AFIPS Fall Joint Computer Conference, Part I**, pages 757–764.

Sutherland, I. E. (1965). The ultimate display. In **Proceedings of the Congress of the International Federation of Information Processing (IFIP)**, volume 2, pages 506–508.

Tainaka, K., Fujimoto, Y., Sawabe, T., Kanbara, M., and Kato, H. (2023). Selection framework of visualization methods in designing AR industrial task-support systems. **Computers in Industry**, 145:103828.

Tait, M. and Billinghurst, M. (2015). The Effect of View Independence in a Collaborative AR System. **Computer Supported Cooperative Work (CSCW)**, 24(6):563–589.

Tanaka, K., Kishino, Y., Miyamae, M., Terada, T., and Nishio, S. (2008). An Information Layout Method for an Optical See-Through Head Mounted Display Focusing on the Viewability. In **2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality**, pages 139–142. IEEE.

Tang, L. and Zhou, J. (2020). **Usability Assessment of Augmented Reality-Based Pedestrian Navigation Aid**, pages 581–591.

Tarawaneh, R. M., Keller, P., and Ebert, A. (2012). A general introduction to graph visualization techniques. **OpenAccess Series in Informatics**, 27:151–164.

Teo, T., Lawrence, L., Lee, G. A., Billinghurst, M., and Adcock, M. (2019). Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction. In **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems**, Paper 201, pages 1–14.

Teo, T., Lee, G. A., Billinghurst, M., and Adcock, M. (2018). Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In **Proceedings of the 30th Australian Conference on Computer-Human Interaction**, pages 406–410.

Teodosio, L. A. and Mills, M. (1993). Panoramic overviews for navigating real-world scenes. In **Proceedings of the first ACM international conference on Multimedia**, pages 359–364.

Thielemans, D. (2023). Virtualisation d'un manuel de maintenance industrielle en réalité augmentée. Master's thesis, University of Namur.

Tonnis, M. and Klinker, G. (2006). Effective control of a car driver's attention for visual and acoustic guidance towards the direction of imminent dangers. In **2006 IEEE/ACM International Symposium on Mixed and Augmented Reality**, pages 13–22.

Tonnis, M., Sandor, C., Klinker, G., Lange, C., and Bubb, H. (2005). Experimental evaluation of an augmented reality visualization for directing a car driver's attention. In **Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)**, pages 56–59.

Trapp, M., Schneider, L., Holz, N., and Döllner, J. (2009). Strategies for visualizing points-of-interest of 3d virtual environments on mobile devices. In **6th International Symposium on LBS & TeleCartography**.

Trapp, M., Schneider, L., Lehmann, C., Holz, N., and Döllner, J. (2011). Strategies for visualising 3d points-of-interest on mobile devices. **Journal of Location Based Services**, 5(2):79–99.

Trueba, R., Andujar, C., and Argelaguet, F. (2009). Complexity and Occlusion Management for the World-in-Miniature Metaphor. In **Smart Graphics**, pages 155–166.

Tsunajima, T. and Nishiuchi, N. (2020). Visual Environment Design of VR Space for Sequential Reading in Web Browsing. In **Communications in Computer and Information Science**, pages 131–136.

Van Krevelen, D. and Poelman, R. (2010). A Survey of Augmented Reality Technologies, Applications and Limitations. **The International Journal of Virtual Reality**, 9(2):1–20.

Veas, E., Grasset, R., Kruijff, E., and Schmalstieg, D. (2012). Extended Overview Techniques for Outdoor Augmented Reality. **IEEE Transactions on Visualization and Computer Graphics**, 18(4):565–572.

Vi, S., da Silva, T. S., and Maurer, F. (2019). User Experience Guidelines for Designing HMD Extended Reality Applications. In **Human-Computer Interaction – INTERACT 2019**, pages 319–341.

Viega, J., Conway, M. J., Williams, G., and Pausch, R. (1996). 3D magic lenses. In **Proceedings of the 9th annual ACM symposium on User interface software and technology - UIST '96**, pages 51–58.

Vyas, D. A. and Bhatt, D. (2017). Augmented Reality (AR) Applications: A survey on Current Trends, Challenges, & Future Scope. **International Journal of Advanced Research in Computer Science**, 8(5):2724–2730.

Wallgrun, J. O., Bagher, M. M., Sajjadi, P., and Klippel, A. (2020). A comparison of visual attention guiding approaches for 360° image-based vr tours. In **2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)**, pages 83–91.

Wang, P., Bai, X., Billinghurst, M., Zhang, S., Zhang, X., Wang, S., He, W., Yan, Y., and Ji, H. (2021). AR/MR Remote Collaboration on Physical Tasks: A Review. **Robotics and Computer-Integrated Manufacturing**, 72:102071.

Wang, Z., Gao, P., Ma, L., and Zhang, W. (2020). Effects of Font Size, Line Spacing, and Font Style on Legibility of Chinese Characters on Consumer-Based Virtual Reality Displays. In **Communications in Computer and Information Science**, pages 468–474. Springer International Publishing.

Ward, M., Barde, A., Russell, P. N., Billinghurst, M., and Helton, W. S. (2016). Visual cues to reorient attention from head mounted displays. **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, 60(1):1574–1578.

Watanabe, J., Kagami, S., and Hashimoto, K. (2015). [poster] photo billboarding: A simple method to provide clues that relate camera views and a 2d map for mobile pedestrian navigation. In **2015 IEEE International Symposium on Mixed and Augmented Reality**, pages 210–211.

Wei, C., Yu, D., and Dingler, T. (2020). Reading on 3D Surfaces in Virtual Environments. In **2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)**, pages 721–728.

Weise, M., Zender, R., and Lucke, U. (2019). A Comprehensive Classification of 3D Selection and Manipulation Techniques. In **Proceedings of Mensch und Computer 2019**, pages 321–332.

Weiser, M. (1991). The computer for the 21st century. **Scientific American**, 265(3):94–105.

Wettel, R. and Lanza, M. (2007). Visualizing Software Systems as Cities. In **2007 4th IEEE International Workshop on Visualizing Software for Understanding and Analysis**, pages 92–99.

Wettel, R. and Lanza, M. (2008). CodeCity: 3D visualization of large-scale software. In **Companion of the 13th international conference on Software engineering - ICSE Companion '08**, pages 921–922.

Wieland, J., Garcia, R. C. H., Reiterer, H., and Feuchtner, T. (2022). Arrow, bézier curve, or halos? – comparing 3d out-of-view object visualization techniques for handheld augmented reality. In **2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)**, pages 797–806.

Wijayawardena, A., Abeysekera, R., and Maduranga, M. (2023). A Systematic Review of 3D Metaphoric Information Visualization. **International Journal of Modern Education and Computer Science (IJMECS)**, 15(1):73–93.

Wingrave, C. A., Haciahmetoglu, Y., and Bowman, D. A. (2006). Overcoming World in Miniature Limitations by a Scaled and Scrolling WIM. In **3D User Interfaces (3DUI'06)**, pages 11–16.

Witmer, B. G. and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. **Presence: Teleoperators and Virtual Environments**, 7(3):225–240.

Woodward, J., Smith, J., Wang, I., Cuenca, S., and Ruiz, J. (2020). Examining the Presentation of Information in Augmented Reality Headsets for Situational Awareness. In **Proceedings of the International Conference on Advanced Visual Interfaces**, Article 57, pages 1–5.

Yadav, N. (2023). Website for their independent research. https://niteeshyadav.com/.

Yang, J., Sasikumar, P., Bai, H., Barde, A., Sörös, G., and Billinghurst, M. (2020). The effects of spatial auditory and visual cues on mixed reality remote collaboration. **Journal on Multimodal User Interfaces**, 14(4):337–352.

Ye, Z.-M., Chen, J.-L., Wang, M., and Yang, Y.-L. (2021). Paval: Position-aware virtual agent locomotion for assisted virtual reality navigation. In **2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)**, pages 239–247.

Yu, D., Liang, H.-N., Fan, K., Zhang, H., Fleming, C., and Papangelis, K. (2020). Design and evaluation of visualization techniques of off-screen and occluded targets in virtual reality environments. **IEEE Transactions on Visualization and Computer Graphics**, 26(9):2762–2774.

Yuan, J., Mansouri, B., Pettey, J., Ahmed, S., and Khaderi, S. (2018). The Visual Effects Associated with Head-Mounted Displays. **International Journal of Ophthalmology and Clinical Research**, 5(2):085.

Zellweger, P. T., Mackinlay, J. D., Good, L., Stefik, M., and Baudisch, P. (2003). City lights: contextual views in minimal space. In **CHI '03 extended abstracts on Human factors in computing systems - CHI '03**, pages 838–839.

Zhao, Y., Hu, M., Hashash, S., and Azenkot, S. (2017). Understanding Low Vision People's Visual Perception on Commercial Augmented Reality Glasses. In **Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems**, pages 4170–4181.