
SAARLAND UNIVERSITY

Faculty of Mathematics and Computer Science
Department of Computer Science
Dissertation



Visual-haptic Perception in the Digitally Augmented World

Dissertation zur Erlangung des Grades des
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Fakultät für Mathematik und Informatik
der Universität des Saarlandes

vorgelegt von
Denise Kahl, M.Sc.
Saarbrücken
2023

Date of the Colloquium:	November 29, 2023
Dean:	Prof. Dr. Jürgen Steimle
Reporter:	Prof. Dr. Antonio Krüger Prof. Dr. Niels Henze
Chairman of the Examination Board:	Prof. Dr. Anna Maria Feit
Scientific Assistant:	Dr. André Zenner

Notes on Style:

Much of the work presented in this dissertation was done in collaboration with students and researchers. Therefore, the scientific plural “we” is used throughout the document. Internet sources are given as URLs.

Acknowledgements

In the past years I have been actively supported by numerous people, without whom the completion of this work would not have been possible. I would like to express special thanks to some of them and apologize to those I have forgotten.

First of all, my sincere thanks go to my supervisor Antonio Krüger for making this work possible for me. His trust over the last years as well as his valuable feedback on this work have been incredibly beneficial. A big thank you also goes to Niels Henze, who agreed to review this work on short notice.

I would also like to express my sincere thanks to my colleagues, who have ensured a pleasant working atmosphere over the past years. Special thanks go to André Zenner and Martin Feick, who have supported me with valuable feedback on the design of studies. Another big thanks goes to Felix Kosmalla, who explained to me how to use the 3D printers and the laser cutter and was always there to help me with technical problems.

I would also like to thank the students whose work helped me in writing this dissertation. A special thanks goes to Marc Ruble, who was always available for feedback on my ideas and who assisted me in conducting various studies as part of his work as an assistant researcher.

I want to extend sincere gratitude to my friends and family who have been a constant source of motivation throughout the years. I am especially grateful to my husband, Gerrit Kahl, and my two daughters, Leni and Mona, who have always been there for me, even though I have not been able to spend as much time with them as they deserve in recent years. Thank you so much for your support and your unconditional love.

Abstract

In everyday life, we are confronted with a growing amount of digital content that is integrated into our surroundings. Visual elements, such as digital advertising or information boards, change our perception of the environment and make it increasingly difficult to perceive personally meaningful information.

In this work, we investigate how visual augmentations of the environment affect our visual and haptic perception of reality and explore how visual attention can be directed as subtly as possible toward personally relevant information in real-world environments.

We present a framework to evaluate visual stimuli for gaze guidance in instrumented environments and explore stimuli suitable for gaze guidance in real-world settings using it. Moreover, we explore the potential of using overlays displayed in Optical See-through Augmented Reality glasses to guide visual attention using subtle visual cue stimuli.

Additionally, we introduce a framework to investigate perceptual changes in physical objects interacted with that may result from overlaying them with digital augmentations. We investigate the extent to which the overlying virtual model can differ from the underlying physical object without significantly affecting the feeling of presence, the usability, and the performance. We provide results in terms of shape and size differences and demonstrate the influence of environmental lighting conditions.

Zusammenfassung

Täglich werden wir mit einer wachsenden Anzahl digitaler Inhalte in unserer Umgebung konfrontiert. Visuelle Elemente, wie digitale Werbung oder Informationstafeln, verändern unsere Wahrnehmung der Umwelt und erschweren es zunehmend, persönlich bedeutsame Informationen wahrzunehmen.

In dieser Arbeit untersuchen wir, wie visuelle Erweiterungen der Umgebung unsere visuelle und haptische Wahrnehmung der Realität beeinflussen und wie der Blick möglichst subtil auf persönlich relevante Informationen in realen Umgebungen gelenkt werden kann.

Wir stellen ein Framework vor, um visuelle Stimuli für die Blickrichtungslenkung in instrumentierten Umgebungen zu evaluieren und erforschen damit geeignete Stimuli zur Lenkung der visuellen Aufmerksamkeit in realen Umgebungen. Darüber hinaus untersuchen wir das Potenzial der Verwendung von Overlays in Optical See-through Augmented Reality Brillen, um den Blick durch subtile visuelle Reize zu lenken.

Ebenso führen wir ein Framework ein, um Wahrnehmungsveränderungen bei physikalischen Objekten zu untersuchen, die sich aus der Überlagerung mit digitalen Erweiterungen ergeben können. Wir untersuchen, inwieweit sich das überlagernde virtuelle Modell vom darunterliegenden physikalischen Objekt unterscheiden kann, ohne das Gefühl der Präsenz, die Benutzerfreundlichkeit sowie die Leistung signifikant zu beeinträchtigen. Wir liefern Ergebnisse in Bezug auf Form- und Größenunterschiede und zeigen den Einfluss der Umgebungsbeleuchtung auf.

Relevant Publications

A number of publications were published during the dissertation period. Several chapters of this thesis contain ideas, figures, tables, results and further text sections already published in other works. Below is a list of the relevant publications including information on where they have been incorporated into this thesis.

Full Conference Papers

- [69] Denise Kahl, Marc Ruble and Antonio Krüger. 2022. The Influence of Environmental Lighting on Size Variations in Optical See-through Tangible Augmented Reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 121–129. (appears in Chapter 1 and Chapter 4)
- [68] Denise Kahl, Marc Ruble and Antonio Krüger. 2021. Investigation of Size Variations in Optical See-through Tangible Augmented Reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 147–155. (appears in Chapter 4)

Preprints

- [65] Denise Kahl and Antonio Krüger. 2023. Using Abstract Tangible Proxy Objects for Interaction in Optical See-through Augmented Reality. *arXiv Preprint arXiv:2308.05836*. (appears in Chapter 4)

Extended Abstracts

- [2] Dmitry Alexandrovsky, Susanne Putze, Valentin Schwind, Elisa D Mekler, Jan David Smeddinck, Denise Kahl, Antonio Krüger and Rainer Malaka. 2021. Evaluating User Experiences in Mixed Reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, 1–5. (appears in Chapter 4 and Chapter 5)
- [23] Tim Düwel, Nico Herbig, Denise Kahl and Antonio Krüger. 2020. Combining Embedded Computation and Image Tracking for Composing Tangible Augmented Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, 1–7. (appears in Chapter 2)

Workshop Papers

- [66] Denise Kahl, Marc Ruble and Antonio Krüger. 2021. Evaluating User Experience in Tangible Augmented Reality. In *Workshop on Evaluating User Experiences in Mixed Reality at CHI '21*. ACM. (appears in Chapter 4 and Chapter 5)
- [67] Denise Kahl, Marc Ruble and Antonio Krüger. 2021. Identification of Everyday Proxies for Tangible Augmented Reality. In *Workshop on Everyday Proxy Objects for Virtual Reality at CHI '21*. ACM. (appears in Chapter 4)

Master's and Bachelor's Theses

Several master's and bachelor's students have been supervised by the author of this thesis. Below is the list of finished works related to this thesis, including the references to the papers above if their results have been included in one of these.

- [62] Philipp Jonczyk. 2023. Adaptive Subtle Gaze Direction in Augmented Reality. Master's thesis, Saarland University, Germany. (appears in Chapter 3)
- [129] Marc Ruble. 2022. Tangible Augmented Reality for Virtual Scene Authoring. Master's thesis, Saarland University, Germany. (appears in Chapter 2)
- [128] Marc Ruble. 2020. Investigation of Possible Size Differences Between Virtual Overlay and Physical Prop in Optical See-through Tangible Augmented Reality. Bachelor's thesis, Saarland University, Germany. Included in [68]. (appears in Chapter 4)
- [24] Tim Düwel. 2018. TAROC: A Tangible Augmented Reality System for Object Configuration. Bachelor's thesis, Saarland University, Germany. Included in [23]. (appears in Chapter 2)
- [130] Nadja Rutsch. 2015. Methoden zur Blickrichtungssteuerung am projizierten Einkaufsregal. Bachelor's thesis, Saarland University, Germany. (appears in Chapter 3)

Other Publications

In addition to the publications mentioned above, further publications were published during the dissertation period. Below you will find a list of these publications as well as a list of the finished works of supervised master's and bachelor's students that are not related to this dissertation.

Full Conference Papers:

- [8] Jochen Bauer, Reiner Wichert, Christoph Konrad, Michael Hechtel, Simon Dengler, Simon Uhrmann, Mouzhi Ge, Peter Poller, Denise Kahl, Bruno Ristok, and others. 2022. ForeSight – User-Centered and Personalized Privacy and Security Approach for Smart Living. In *International Conference on Human-Computer Interaction*. Springer, 18–36.
- [147] Carsten Ullrich, Matthias Aust, Niklas Kreggenfeld, Denise Kahl, Christopher Prinz, Simon Schwantzer. 2015. Assistance- and Knowledge-services for Smart Production. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*. ACM, 1–4.
- [86] Christian Lander, Marco Speicher, Denise Paradowski, Norine Coenen, Sebastian Biewer, and Antonio Krüger. 2015. Collaborative Newspaper: Exploring an Adaptive Scrolling Algorithm in a Multi-user Reading Scenario. In *Proceedings of the 4th International Symposium on Pervasive Displays*. ACM, 163–169.
- [71] Gerrit Kahl and Denise Paradowski. 2013. A Privacy-aware Shopping Scenario. In *Proceedings of the Companion Publication of the 2013 International Conference on Intelligent User Interfaces (IUI)*. ACM, 107–108.
- [121] Denise Paradowski and Antonio Krüger. 2013. Modularization of Mobile Shopping Assistance Systems. In *2013 5th International Workshop on Near Field Communication (NFC)*. IEEE, 1-6.

Workshop Papers

- [146] Carsten Ullrich, Matthias Aust, Roland Blach, Michael Dietrich, Christoph Igel, Niklas Kreggenfeld, Denise Kahl, Christopher Prinz, Simon Schwantzer. 2015. Assistenz-und Wissensdienste für den Shopfloor. In *Proceedings of DeLFI Workshops 2015*.
- [102] Markus Löchtefeld, Sven Gehring, Denise Paradowski and Antonio Krüger. 2014. Filtered Reality – Keeping Your Peripheral Vision Clean. In *Workshop on Peripheral Interaction: Shaping the Research and Design Space at CHI 2014*. ACM.

Book Chapters

- [84] Antonio Krüger, Wolfgang Maaß, Denise Paradowski, Sabine Jansen. 2014. Empfehlungssysteme und integrierte Informationsdienste zur Steigerung der Wertschöpfung im stationären Handel. In *Wertschöpfung im Handel*. Kohlhammer Verlag.
- [106] Christian Malewski, Janina Wiesmann, Denise Paradowski and Andreas Grote. 2013. Fächerverbindendes Arbeiten mit Karten-APIs. In *Geoinformation im Geographieunterricht*. Monsenstein und Vannerdat, 158-175.

Master's and Bachelor's Theses

- Hafiz Zee Waqar Irtaza. 2023. Guidance using Augmented Reality Filtering. Master's thesis, Saarland University, Germany.
- Michal Jurek. 2023. Smarthome Webshop Assistant. Bachelor's thesis, Saarland University, Germany.
- Thorsten Willems. 2022. Entwicklung und Evaluation einer mobilen Augmented Reality Applikation zur Einkaufsunterstützung basierend auf Makronährstoffen mithilfe einer Lebensmittel Ontologie. Bachelor's thesis, Saarland University, Germany.
- Samantha Ruppert. 2021. Movement Detecting Infant Monitoring Device. Bachelor's thesis, Saarland University, Germany.
- Philipp Jonczyk. 2020. Genetic Algorithm for Decentralized Multi-Agent Multi-Goal-Pathfinding. Bachelor's thesis, Saarland University, Germany.
- Alla Domke. 2017. An Interaction Concept for Performance Support using Smart Glasses. Master's thesis, Saarland University, Germany.
- Alexander Cullmann. 2014. Motion Gestures for Mobile Payment. Master's thesis, Saarland University, Germany.
- Patrick Bender. 2014. RFID-basiertes Framework zur Einkaufsunterstützung mittels Gestensteuerung. Bachelor's thesis, Saarland University, Germany.
- Yannic Haupenthal. 2013. YAVA: Yet Another Vegan App. Bachelor's thesis, Saarland University, Germany.
- Michael Barz. 2012. Smartphone zu Smartphone Bezahlung per Near Field Communication. Bachelor's thesis, Saarland University, Germany.

Contents

1	Introduction	1
1.1	Problem Statement and Motivation	1
1.1.1	Perception in an Augmented World	1
1.1.2	Controlling Perception through Visual Augmentations	3
1.1.3	Perception of Augmented Physical Objects	4
1.2	Research Questions	5
1.3	Methods and Approach	6
1.4	Contributions to the Field	7
1.5	Thesis Outline	8
2	Background and Related Work	11
2.1	Visual Perception	11
2.1.1	Anatomy of Vision	11
2.1.2	Visual Attention	15
2.1.3	Directing Attention using Visual Cues	17
2.1.4	Eye Tracking for Measuring Gaze Behavior	18
2.1.5	Summary	20
2.2	Gaze Direction	20
2.2.1	Overt Gaze Direction	20
2.2.2	Subtle Gaze Direction	23
2.2.3	Summary	30
2.3	Augmented Reality	30
2.3.1	Tangible Augmented Reality	31
2.3.2	Optical See-through Augmented Reality	35
2.3.3	Summary	35
2.4	Haptic Perception	36
2.4.1	Anatomy of the Somatosensory System	36
2.4.2	Haptic Exploration of Objects	37
2.4.3	Intersensory Interaction	38
2.4.4	Summary	39
2.5	Proxy Interaction	39
2.5.1	Proxies in Virtual Reality	39
2.5.2	Proxies in Video See-through Augmented Reality	42
2.5.3	Proxies in Optical See-through Augmented Reality	43
2.5.4	Object Tracking	44
2.5.5	Summary	45

3	Understanding the Perception of Visual Cues for Gaze Guidance	47
3.1	Peripheral Perception of Visual Cues	47
3.2	Perception of Visual Stimuli for Guiding Gaze	54
3.2.1	Subtle Visual Cues for Gaze Guidance at Projections	54
3.2.2	Gaze Guidance in Instrumented Environments	65
3.2.3	Subtle Gaze Guidance in Augmented Reality	78
3.3	Summary	90
4	Understanding Visual-haptic Perception of AR Proxy Interaction	93
4.1	Measuring Presence in Tangible AR	94
4.2	Perception upon Interaction with a Divergent Proxy Object	96
4.2.1	Basic Approach for Evaluating AR Proxy Interaction	96
4.2.2	Investigation of Size Variations	97
4.2.3	Influence of Environmental Lighting on Size Variations	112
4.2.4	Investigation of Shape Variations	128
4.3	Summary	141
5	Frameworks for Controlling Visual and Haptic Perception	143
5.1	Framework for Adaptive Gaze Guidance	143
5.1.1	Concept	144
5.1.2	Prototypical Implementation	146
5.1.3	Conclusion	151
5.2	Framework for AR Proxy Interaction	152
5.2.1	Concept	152
5.2.2	Prototypical Implementation	154
5.2.3	Conclusion	165
5.3	Summary	166
6	Conclusion	167
6.1	Summary	167
6.2	Major Contributions	169
6.3	Future Work	171
6.4	Closing Remarks	174

A	Landscape Images	175
B	Questionnaires	179
B.1	Presence Questionnaires	180
B.1.1	AR Presence Questionnaires	180
B.1.2	TAR Presence Questionnaires	182
B.2	Usability Questionnaires	184
B.2.1	Size Perception Questionnaires	184
B.2.2	Lighting Perception Questionnaire	186
B.2.3	Shape Perception Questionnaire	187
B.3	Concluding Questionnaires	188
B.3.1	Size Variations Study	188
B.3.2	Lighting Variations Study	191
B.3.3	Shape Variations Study	194
	List of Figures	197
	List of Tables	201
	List of Abbreviations	202
	Bibliography	205

Chapter 1

Introduction

This chapter first explains the motivation for this thesis by identifying the problems to be investigated along with possible solutions. Then, the research questions and the approach to answer these questions are presented. In addition, the contributions of the thesis are described and its structure is explained.

1.1 Problem Statement and Motivation

In the following, we address the problems triggered by the multitude of visual information in our environment, which we are already confronted with and which will become even more critical in the future. Furthermore, we address potential technical solutions to be investigated, which are the motivation for the research conducted in this thesis.

1.1.1 Perception in an Augmented World

In our daily lives, we are constantly surrounded by multitudes of visual information. Besides posters or information signs, more and more digital displays are integrated into the environment, and these are designed to try to attract our attention. In addition to large monitors and information boards, there are also more and more small systems, such as digital price tags, which, as Mark Weiser already predicted in 1991, integrate themselves almost invisibly into our environment [158]. Especially in complex environments, like supermarkets and airports,



Figure 1.1: Digital advertising in Tokyo causing visual information overload.¹

or in big cities, we are flooded with a huge amount of information (see Figure 1.1). The human is not able to process all visual information in such situations, which inevitably leads to ignoring much of the information presented [16, 26, 117]. This can lead to truly important information being overlooked, while many things that are unimportant in the current situation are processed. This results unavoidably in overstraining, especially in stressful situations.

In an ideal world without information overload, each person would be presented only with information that is of interest to him or her at the moment, and only as much as he or she can consume without stress. However, such a highly personalized system would inhibit our further development, as we would never be confronted with new ideas, e.g. new products via advertising. In addition, interests of external groups, such as advertisers, must be taken into account. A world without information overload is therefore not desirable and would also be impossible or very difficult to implement in many situations. In a supermarket, for example, we are inevitably confronted with product information on the packages, but it is impossible to imagine doing without it here. The same is true for other sources of information that are firmly integrated into our daily lives.

¹Source: <https://authory.com/WilsondaSilva/Information-Overload> (last accessed: 2023-08-15)

What is needed, therefore, are suitable approaches and strategies to help people locate the most important information for them in the environment.

In this work, we investigate to what extent visual augmentations in the environment can be used to direct the visual attention of individuals to information that is relevant to them. In particular, possibilities are considered that allow for a subtle and less disturbing directing of attention instead of further flooding the user with visual information via additional obvious presentations, e.g. arrows or red borders.

1.1.2 Controlling Perception through Visual Augmentations

Some studies have already been conducted in related work that mainly used computer screens to test whether gentle guidance of gaze direction, e.g. by saliency modulation [37, 47, 111, 112, 151], blur [50, 101] or by subtle visual cues [6, 109, 110, 138] is possible. Research on the use of subtle visual stimuli in the real world or in real-world environments is limited [14, 43, 45]. However, if a system for everyday assistance is to be created on this basis, it is essential to investigate this in more detail. We have therefore conducted studies on the visual guidance of attention by subtle cue stimuli using projected real-sized environments and an instrumented environment.

In densely populated public spaces, it is not practical to use visual stimuli in the environment to direct attention. Such stimuli are visible to everyone in the vicinity, making it difficult to ensure that only the desired person responds to the stimulus. There is also a risk that personal information about the user could be revealed to others. For example, highlighting the product a person is looking for in a supermarket would compromise their privacy. In such crowded public spaces, where instrumented environments are not suitable for directing the attention of individuals, smartphones or Augmented Reality (AR) glasses can be used to draw the user's attention as gently as possible to the products of interest. In this regard, viewing digital content on a private device provides better protection of private information and ensures that other people's perception is not flooded with meaningless information. We therefore investigate whether and which methods can be used to guide gaze direction by displaying subtle visual cues in AR glasses. However, the visual augmentation of reality, e.g., with the help of AR glasses, can also influence the haptic perception of real objects with

which one interacts, which is why more detailed investigations are needed in this area as well.

1.1.3 Perception of Augmented Physical Objects

When reality is superimposed with virtual content, the physical environment must also be taken into account. For example, by overlaying physical items, the visual perception of these items changes. At the same time, the visual overlay can also have an impact on the haptic perception. For example, if the physical object is overlaid with a completely different virtual object, it would likely result in a misperception when trying to grasp the object.

Especially in the field of Tangible Augmented Reality (TAR), where physical objects are used to manipulate virtual content, the influence of the visual augmentations on the visual-haptic perception of the physical objects plays a crucial role. AR glasses, which are experiencing growing interest and increased use, enable this form of natural interaction. One notable advantage of AR glasses is that they offer a hands-free experience that doesn't require holding additional devices, such as a smartphone, to interact with virtual content. With Optical See-through AR (OST AR) glasses, the virtual content overlaid on the environment is somewhat transparent, so that the physical objects behind it remain visible to a certain degree. This is likely to influence the perception of the environment.

When a physical object serves as a substitute for interacting with virtual content through overlaying it with its associated virtual representation, the physical object is referred to as a proxy object. This approach avoids direct interaction with the virtual content. Instead, one can intuitively grab and move a physical counterpart in order to interact with the virtual content. A simple example would be the use of a wooden figure as a proxy object to control a virtual game figure on a virtual game board. It is reasonable to assume that such a proxy object used to interact with virtual content would ideally be an exact visual replica of the virtual object. However, given the large number of possible applications, it would not be feasible to create and store suitable objects for every use case. Therefore, it is necessary to investigate to what extent the physical object which is used for interaction can differ from its virtual counterpart.

Some research exists on how the appearance of virtual objects affects the perception of underlying proxy objects in Virtual Reality (VR) [9, 21, 133] and in Video See-through AR (VST AR) [79, 85]. There is little research related to proxy

interaction in OST AR [15, 55]. To our knowledge, it has not yet been investigated how the visualization of virtual objects on top of physical proxy objects affects the visual and haptic perception of these objects in OST AR.

We therefore investigate to what extent it is possible to use a physical object for interaction that is different from the virtual model and that can ideally represent a variety of virtual objects in OST AR. Using a different physical proxy object for interaction leads to visual discrepancies between the virtual and physical object in size, shape, and texture. We determine whether a deviation in size and shape is feasible without significant drawbacks for the usability, performance, and sense of presence.

When determining possible deviations, the environmental lighting has to be taken into account, since the illumination level influences the perception of contrast and color of the virtual overlay [27, 32, 33, 34]. Therefore, we also investigate the influence of illumination on possible differences between the physical proxy object and its virtual overlay.

1.2 Research Questions

This thesis explores the following main research question, to which we would like to contribute:

How can the visual-haptic perception of reality be influenced by digital augmentations?

We divide this question into the following sub-questions, which we address in this thesis:

- RQ1** How can a person's gaze direction be influenced by a digitally augmented reality?
- RQ2** How can we test which actuators and sensors are suitable for directing visual attention?
- RQ3** How do digital extensions of reality in the form of overlays on real 3D objects influence the visual-haptic perception of these objects?
- RQ4** How can we test how much virtual overlays can differ from their physical 3D counterparts used for interaction?

With **RQ1** we investigate to what extent digital augmentations of the real world can direct people's visual attention. We want to find out which visual cues are suitable for directing visual attention to specific objects. We also investigate which sensors are suitable for determining gaze direction. Furthermore, we explore how visual attention can be directed as subtly as possible and how this can be done in crowded environments.

With **RQ2** we want to answer how it is possible to determine suitable sensors and actuators for gaze guidance. Here we investigate which approach can be used to determine which actuators or cue stimuli are suitable for directing attention to predefined locations and which sensors are suitable for gaze direction detection in real instrumented environments.

With **RQ3** we investigate how digital extensions of reality in the form of overlays on real 3D objects influence the visual-haptic perception of these objects. Here, we determine how much the virtual objects can differ in size and shape from their physical counterparts that are interacted with, without seriously affecting user acceptance. In addition, we investigate the influence of environmental lighting on the perception of virtual objects in OST AR.

With **RQ4** we want to determine what approach can be used to test how much virtual overlays can differ from the physical 3D objects used for interaction. We also want to determine how evaluations in the form of questionnaires can be carried out in OST AR without disrupting the AR experience.

1.3 Methods and Approach

In order to answer the research questions stated above, we proceeded as follows. After an extensive literature review, we identified questions that we wanted to answer with experimental studies. In order to conduct these studies, two frameworks had to be developed first. The first framework we needed was to display visual stimuli in the environment and measure people's response to the stimuli in terms of changes in gaze direction (see Section 5.1). The second framework we required was to allow virtual overlays in OST AR to be placed precisely onto real objects and to be able to interact with them (see Section 5.2). After conceptualization, the two frameworks were implemented prototypically so that we could perform the studies. Every study was carefully designed and hypotheses were formulated (see Chapters 3 and 4). The prototypical framework

was then tailored to suit the specific requirements of each study. Participants were recruited in advance for the experiments, which all took place in a laboratory setting. During the studies, the participants had to complete various questionnaires and several measurements were taken. After completion, the results were thoroughly analyzed to gain valuable insights.

1.4 Contributions to the Field

This thesis makes several contributions to different areas of Human-Computer Interaction (HCI).

It contributes to the field of (Subtle) Gaze Guidance by showing that it is possible to subtly guide gaze in real and real-sized environments. Various visual cues for directing visual attention in real-world environments were investigated, and both static and adaptive visualization approaches were explored. To our knowledge, we are the first to investigate different approaches for subtle visual attention guidance directly in OST AR, making an important contribution to the fields of (Subtle) Gaze Guidance and Optical See-through Augmented Reality.

In addition, this work contributes to the areas of Proxy Interaction, Tangible Augmented Reality, and Optical See-through Augmented Reality by demonstrating how visual augmentations superimposed on physical objects affect the visual and haptic perception of these objects in OST AR. To our knowledge, there has been no previous research in OST AR exploring these issues. This work provides insights into the extent to which the virtual overlay can differ in size and shape from the underlying physical object without significantly worsening usability, performance, and the feeling of presence.

This thesis also makes several technical contributions achieved by the two frameworks developed during the dissertation period. The framework for adaptive guidance of visual attention in real-world environments allows other researchers to investigate suitable visual stimuli for gaze guidance and suitable sensors for gaze direction measurement, and to evaluate different visualization methods appropriate for their use cases. The framework for proxy interaction in OST AR enables investigations into the influence of the visual overlay superimposed on a physical object on visual and haptic perception when interacting with it. This allows other researchers to also investigate proxy interactions in OST AR, e.g. with respect to possible differences in the material or texture. The specifications

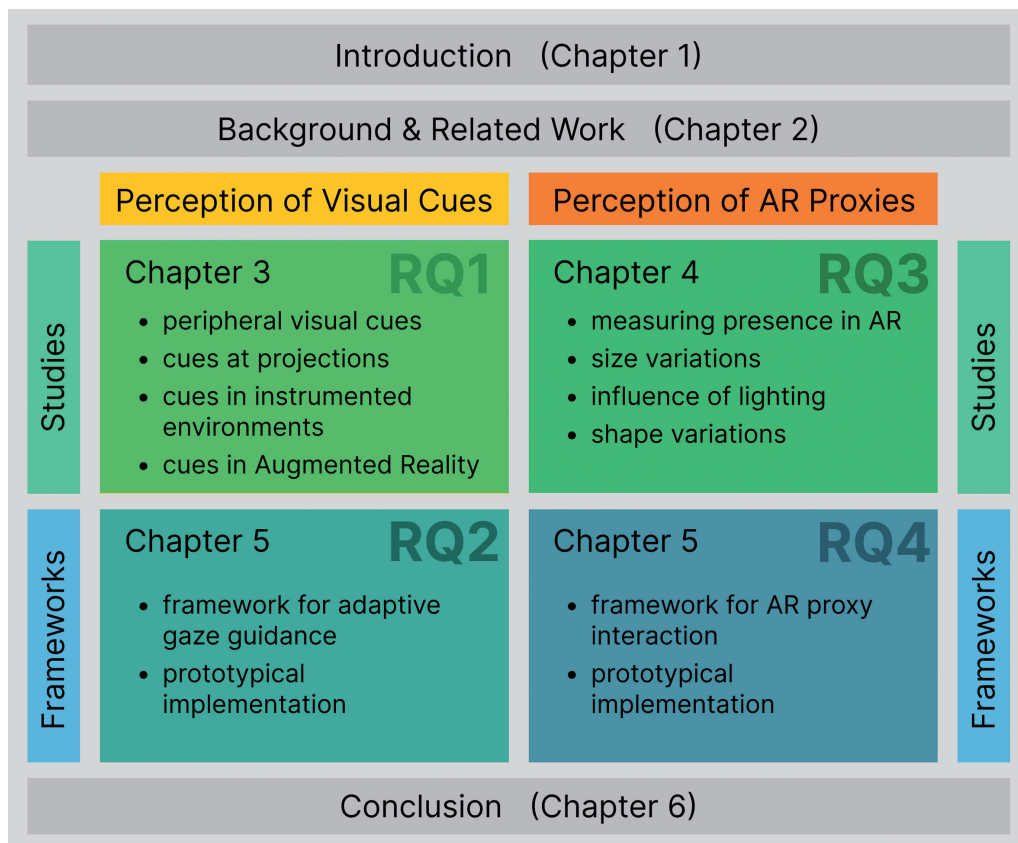


Figure 1.2: Presentation of the structure of the thesis. Two main topics were investigated, the perception of visual cues and the perception of AR proxy objects. A framework was implemented for each of the two topics, which forms the basis for the studies conducted.

of the frameworks assist other researchers in implementing the frameworks for their own purposes, adapted to the available hardware and software for tracking gaze and objects and for visualizing the digital augmentations.

1.5 Thesis Outline

The structure of the remainder of this dissertation is as follows (see also Figure 1.2): In Chapter 2, the fundamentals on which this thesis is based are explained. Furthermore, related work is presented and the importance of the research of this thesis is highlighted. Chapter 3 presents our research on gaze direction guidance using visual cue stimuli. This chapter also provides answers to research question **RQ1**. Subsequently, in Chapter 4 we describe our investiga-

tions on the perception of proxy objects in OST AR and show how they answer research question **RQ3**. Chapter 5 explains the frameworks we developed in detail. This chapter provides answers to research questions **RQ2** and **RQ4**. Finally, Chapter 6 summarizes our main contributions and possible future work that can be built upon this thesis.

Chapter 2

Background and Related Work

This chapter serves to provide a basic understanding of the concepts that are important for understanding this thesis. Additionally, we present relevant work that is closely related to the thesis. First, we explain in detail how visual perception works and how visual attention can be directed to predefined objects and locations. We then present related work in the field of gaze guidance, where studies on the guidance of visual attention have been conducted. Next, we address the foundations of AR and haptic perception. Finally, we present related work in the field of proxy interactions, where studies have been conducted on proxy interaction in VR, VST AR, and OST AR.

2.1 Visual Perception

To control what information people should focus on, it is important to understand how human visual perception works. We therefore first give an overview of the anatomical properties of the visual system. We then explain visual attention and the role it plays in the perception process, and show how it can be directed.

2.1.1 Anatomy of Vision

Visual perception is the end result of many complex processes. In addition to the eye, nerves, synapses and the brain are also involved in this process. Only

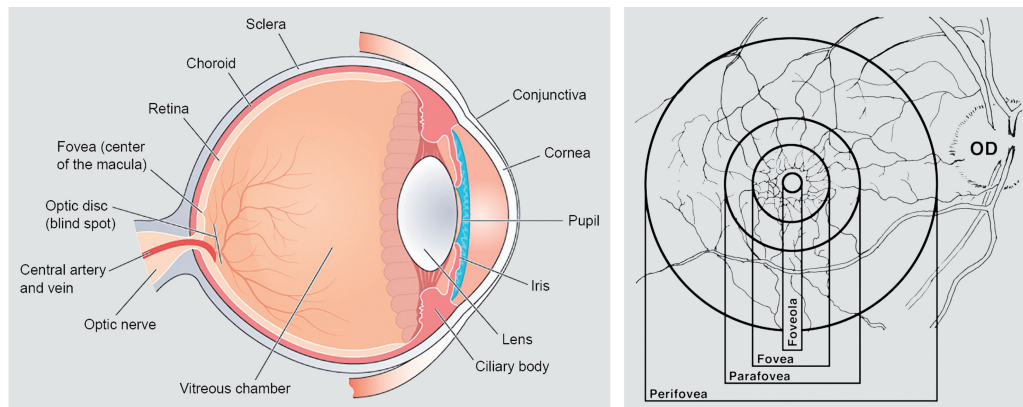


Figure 2.1: The human eye with the cornea, lens, optic nerve, retina, and fovea (left). Adapted from [30]. The central human retina (macula) with an idealized view of its blood vessels and the optic disc (OD) (right). The macula is subdivided into the central-most foveola, surrounded by the fovea, parafovea and wider perifovea. Adapted from [54].

through the interaction of all processes is a visual impression created, which is why we present the individual processes in more detail below.

The Human Eye The human eye is the basis for perceiving objects in the environment. Figure 2.1 (left) represents a simplified cross-section of this complex sense organ. The light rays reflected from objects in the environment strike the human eye. They are then projected through the lens at the front of the eye onto the back of the eye, creating an image of the environment on the retina [42, 105]. The retina is divided into different areas. The area in the center of the retina is called the foveola. Around this area are the foveal area, the parafoveal area, and the perifoveal area [54, 81] (see Figure 2.1, right). All of these areas contain millions of different receptors, which exist in varying amounts in each area.

Functioning of the Receptors Receptors are light-sensitive photoreceptors that process and transmit light information to the human brain [42, 105]. When the light hits the retina of the eye, the receptors located there are stimulated. A chemical process is triggered by which the light energy is converted into electrical energy. This conversion is called transduction [42]. The resulting electrical energy is then transported from the receptors via the optic nerve, which begins at the back of the eye, through complex neural pathways to the brain.

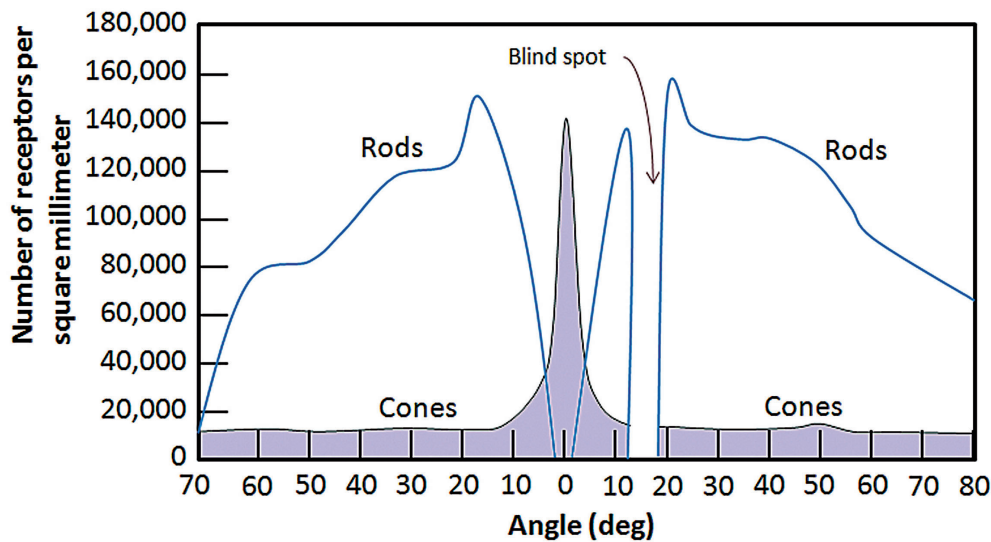


Figure 2.2: Distribution of rods and cones on the retina [96].

Information Processing in the Brain The processing of visual information takes place in the brain in the so-called occipital lobe. The electrical energy created by the arrival of the light rays on the retina is transmitted via the nerve pathways to this brain region, where it is finally processed. Only after processing in the brain is the visual process complete and the person perceives an image of his or her surroundings. Just because the gaze or the attention is directed to a certain object, that does not mean that this object is also perceived. Only by his or her knowledge can the human being find a connection to what is looked at, recognize what it is about and carry out a suitable action.

Peripheral and Foveal Vision Humans can only see in detail and sharpness in the center of their gaze. Towards the edge, the details and sharpness decrease significantly [42, 105]. This is due to the different distribution of receptor types on the retina. In the center, the foveal area, there are exclusively so-called cones, which are responsible for high detail resolution and color representation [96]. This allows the human eye to see in detail and sharpness in the center of its gaze. Outside the fovea are mainly the so-called rods (see Figure 2.2). This type of receptor does not have high detail resolution and cannot perceive colors, but is only sensitive to brightness values. To obtain a detailed and sharp image of a section of their environment, humans must therefore look directly at that section [42].

Perception of Colors and Visual Resolution in the Periphery How we perceive the color of objects in our environment depends on the wavelengths of light that are reflected from the objects into our eyes [28, 42]. The color we see depends on how the objects selectively reflect some wavelengths (selective reflection). The color of transparent objects such as glass and liquids is generated by selective transmission, because only some wavelengths can pass through the transparent objects [28].

The cones in the retina are specialized for color vision, because they contain light-sensitive photopigments. According to a well-known color theory, the trichromatic (three-colored) theory, there are three different types of receptors [28]. The first are sensitive to short wavelengths and respond predominantly to blue perceived stimuli. Receptors of the second type are sensitive to medium-wavelength light, which is reflected from yellow-green objects, for example. And the third type of receptors is most sensitive to long wavelengths reflected from orange-red visual stimuli [28]. Most stimuli activate two or three of the above receptor types, and we perceive color based on the relative stimulation strength of the receptor types. Despite the limited number of cone types, we can distinguish millions of colors in this way.

Figure 2.2 shows that there are significantly more cones in the foveal area than in the peripheral area. Many studies have investigated the extent to which color perception is possible in the periphery. The results show that the areas for color perception differ for individual colors. Overall, results vary with respect to color ranges in studies from 48° to 56° for blue, from 33° to 45° for red, and from 24° to 35° for green [78].

Many studies have also been conducted to determine the visual resolution in the periphery. The visual acuity at a distance from 0° to 30° from the fixation point was investigated here. Various shapes or Landolt rings were used in the experiments [75]. In all of them, visual acuity was found to decrease with increasing distance from the fixation point. However, Kerr [75] has shown that acuity in the periphery does not decrease as rapidly as previously thought.

Field of View Another limitation in the visual perception of the environment is the limited human field of view (FOV). Human beings can only view a small section of their environment at any given time [42]. Figure 2.3 illustrates the FOV, which extends horizontally over a maximum range of 200° . The vertical range covers a total of about 135° , of which 80° belong to the lower part and 55° belong

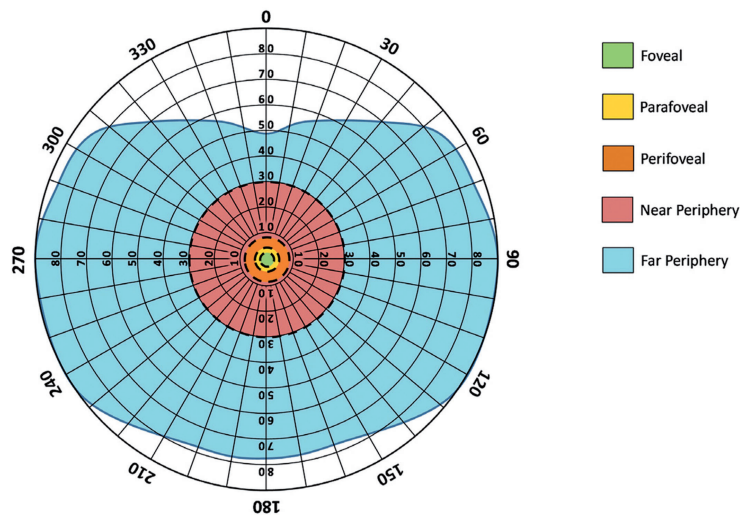


Figure 2.3: Illustration of the human visual field. Central visual field, near, middle and far peripheral vision are from 0° to 8° , 8° to 30° , 30° to 60° and beyond 60° respectively. Adapted from [139].

to the upper part of the FOV [105]. Thus, changes in the viewing direction are necessary to be able to perceive the surroundings completely.

2.1.2 Visual Attention

To overcome the limitation of the FOV, it is necessary for humans to make use of eye and head movements. Since they are able to see sharply only in the center of their vision, it is necessary for them to direct their gaze to a specific stimulus [42, 105]. The process of focusing on a specific object while blocking out (not paying attention to) other stimuli in the environment is called attention. William James, the first professor of psychology at Harvard, wrote an entire chapter on visual attention in his 1890 book *Principles of Psychology* [60]. His statements were based on personal observations and not on experiments. According to James, we are confronted with many sensory impressions. He said that we focus only on certain things that we want to perceive and exclude others, so that we can perceive only a fraction of all experiences. Many of his statements have been confirmed in various studies over time.

There are different varieties of attention [160]. If one concentrates on a certain object even though there are many distractors, one speaks of *focused attention*. During visual scanning of the environment, people have the ability to pay atten-

tion selectively to certain stimuli, which is called *selective attention*. The process of moving attention from one object to the next is called *attention switch*. If someone is able to process two tasks truly simultaneously, one speaks of *divided attention*. Whether, and to what extent, tasks can be processed in parallel depends on the perceptual load required to process the individual tasks. A high-load task requires more of the person's total available perceptual capacity, making distractions by other stimuli less likely [87, 88]. However, the "fuel" will never be completely used up, so it will always be possible to respond to unexpected dangerous emergencies [160].

Many different visual stimuli simultaneously affect the human being. If, however, he or she is supposed to look at only one of them intensively, the orientation of attention plays a significant role. This is often influenced by the objective of the respective person (top-down process) [3, 42]. For example, humans direct their attention and gaze to a door handle when they have the goal of opening the associated door. Thus, attention is directed to an object in order to achieve the observer's goal. However, the focus of attention and gaze is not always guided by a conscious decision. In addition to the conscious decision to take a closer look at and process a certain area, there are also unconscious processes that lead a person to take a closer look at a certain area and process it in detail. Such a bottom-up process occurs, for example, by capturing attention through salient visual features in the scene, such as color, orientation, motion, and depth [80]. Koch and Ullman [80] have combined all of these features into a single topographically oriented map. This so-called saliency map shows how strongly individual regions in a scene stand out from their surrounding areas and thus attract attention.

Since the peripheral area of the eye functions similarly to an alarm system, new, potentially interesting and relevant stimuli occurring in the periphery are noticed by the visual system and targeted as the next fixation goals. This function is performed by the transient cells, which are strongly represented in the periphery. These cells react very strongly to movements. This includes, for example, suddenly appearing objects [3]. Visual peripheral stimuli, e.g. in the form of flashing lights, therefore involuntarily attract the attention of humans.

If we want someone to perceive something specific in the environment, his or her attention must first be directed to it. We distinguish between overt and covert attention shifts. Overt attention shifts involve a movement of the eyes, head, or body to shift the focus to an object of interest [122, 160]. In a covert attention shift, on the other hand, there is no physical movement that can be easily measured.

Instead, there is a change of mental focus on an object already in the FOV, which one might otherwise not even have noticed. The phenomenon of overlooking an object that is prominent in the visual field is known as *inattention blindness* [103]. Another way a lack of attention can affect perception is *change blindness* [125, 126]. This is the difficulty of recognizing differences in scenes that are obvious when one knows where they are. If an observer is looking somewhere else at the time of the change, or if the object that is changing is obscured at the time of the change, changes usually remain unnoticed. This must be taken into account in studies on attention control.

2.1.3 Directing Attention using Visual Cues

The first investigations into whether visual attention can be directed by visual cue stimuli were conducted very early on by John Jonides [63]. He found through experiments that peripheral visual cue stimuli automatically attract a person's attention. Jonides conducted his series of experiments on a computer. One experimental run included several identically structured tasks in which participants were asked to press a specific key on the keyboard depending on the target stimulus presented. The procedure for the tasks with peripheral cues was such that the participants first had to fixate a point in the center of the screen. This was then hidden and instead a cue in the form of an arrow was displayed in the periphery for a short moment. This was followed by the display of the actual task, which consisted of a visual search task [143]. Here, up to eight letters were arranged in a circle and the participants had to determine whether an "L" or an "R" could be identified among the letters. If they found an "L", they had to press the left arrow key on the keyboard as quickly as possible, and if they found an "R", they had to press the right arrow key. A distinction was made between valid and non-valid test conditions. In the valid experimental conditions, the position of the cue stimulus corresponded to the position of the searched target, i.e., the "R" or "L"; in the non-valid experimental conditions, the cue stimulus was displayed at the position of a different letter. Among other things, the time taken by the participants to solve the search task was measured.

John Jonides found that a valid test set-up results in faster reaction times than a non-valid test set-up. This can be attributed to the fact that in the case of a valid test condition, attention was already directed to the location of the target stimulus beforehand by the cue stimulus. In this way, the task was made easier for the participant and his or her performance improved. Peripheral exogenous

cue stimuli thus automatically attract a person's attention, even under high levels of cognitive load. This fundamental property can be used for the purpose of gaze direction control.

2.1.4 Eye Tracking for Measuring Gaze Behavior

Eye tracking can be used to check whether attention could be directed to an object for a longer period of time, i.e. whether a shift of gaze direction has occurred. Eye tracking makes it possible to recognize the direction of the user's gaze and thus also to identify gaze shifts. In this way, it can be precisely determined whether a presented cue has directed the gaze direction. There are different eye tracking methods. Today, the use of video-based eye trackers is the most common method for determining gaze direction with a high degree of accuracy. Here, the position of the corneal reflection of an infrared light relative to the pupil is measured. This method is used for head-mounted eye trackers as well as for eye trackers embedded in the environment and allows tracking of the eyes in real time [17]. Combining images of the eye with images of the FOV makes it possible to determine what someone is looking at. During a specific task, an eye tracker can therefore be used to determine where, how and in what order the gaze is directed. Due to the anatomy of the eye, we can only see a small part of our visual field very sharply, so we tend to move our eyes to what we are processing. Due to this so-called eye-mind link [64], eye tracking represents a suitable tool for the investigation of visual attention [17].

There is a wide range of video-based eye trackers that differ from each other in many ways. The selection of the appropriate eye tracker should therefore be adapted to the task. Besides the distinction between stationary, portable or head-mounted eye trackers, there are also differences in the freedom of movement. While some allow head movements, others depend on the head being held in a fixed position, e.g. with the help of a chin rest. Additionally, eye trackers differ in their sampling rate, which is measured in Hertz (Hz). The fastest commercial eye trackers can measure the eye position up to 2000 times per second (2000 Hz) while some eye tracking glasses, for example, only manage 50 measurements per second (50 Hz). Again, the actual task plays a big role in the selection. For example, if the task is only to determine what someone is looking at, an inexpensive eye tracker with 50 Hz may already be sufficient [17]. Eye trackers record the position data of the eyes in x and y coordinates, which must be analyzed later on. For

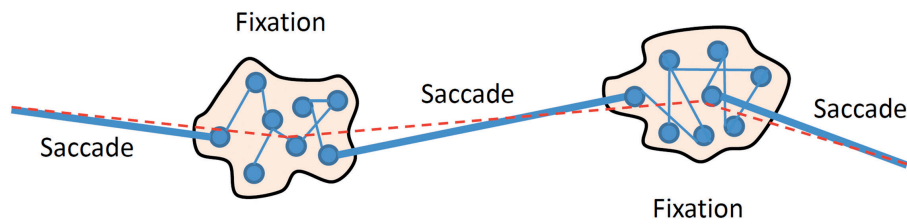


Figure 2.4: Illustration of fixations and saccades. The eye jitters around the area the human is focusing on. These micromovements are abstracted to fixations. A saccade occurs when the focus switches from one area of interest to the next [83].

this purpose, it is possible to use standard software or to carry out the analyses oneself. To do this, it is important to know what types of eye movements exist.

Eye movements are divided into fixations and saccades. In a fixation, the eyes are fixed on a visual target and are perceiving visual information. The fixations are usually very short, because it is necessary to move the eyes regularly to get a lot of high quality information from the complete environment. The length of a single fixation depends on a variety of factors. In addition to the type of visual stimulus, the target and the difficulty of the task as well as the experience and attention of the person play a significant role. In general, the duration of a fixation is about 180–330 milliseconds [17].

If someone focuses on something of interest, the eye jitters around the area of interest. When the eye movement changes rapidly from one fixation point to the next, this is called a saccade (see Figure 2.4). During a saccade, no visual input is received, so that one is virtually blind during this period, which is known as *saccadic suppression* [108]. The length of a saccade also depends on the task and is between 30 and 50 milliseconds [17]. A typical length of a saccade is between 2° , e.g. for reading, and 5° , e.g. for scanning the environment [157]. This information is of significant use when manually evaluating eye tracking raw data.

The accuracy of the data depends not only on the eye tracker used but also on how precisely the system has been calibrated. A calibration to the respective user is essential in order to be able to make correct statements about where someone has looked. In this process, the user usually has to look at different known points in the target environment one after the other. The positions of the center of the pupil(s) when looking at the respective points in the environment are used to calculate the calibration. The calibration is validated again in a test run afterwards to ensure that it was successful [17].

2.1.5 Summary

The aspects of visual perception presented in this section demonstrate the fundamentals of why and in what way the direction of human gaze can be specifically controlled. In addition to the biological aspects that are necessary to perceive an image of the environment, the psychological aspects also play an important role. In the foveal visual field, colors and details can be recognized, while the peripheral visual field is mainly used for orientation. In addition, the latter also reacts like an alarm system to visual stimuli. This property can be used to direct the person's gaze to certain objects. Eye trackers can be used to measure the direction of gaze and thus check whether directing attention was successful.

2.2 Gaze Direction

There are basically two ways to direct the user's gaze to a particular object using visual means. One can either use obvious visual stimuli that are clearly perceived (overt gaze direction), or one can try to subtly direct the gaze direction so that the visual stimuli are not perceived at all, or at least as little as possible (subtle gaze direction) [45]. We first briefly address the issue of obvious gaze guidance. Subsequently, we deal intensively with the issue of subtle gaze guidance, since our goal is to minimize stimulus overload and not to overwhelm the subjects with further obvious stimuli.

2.2.1 Overt Gaze Direction

Overt gaze direction refers to the use of obvious visual cues to direct the user's gaze. These stimuli include, for example, colored arrows and highlight boxes, which may also move or flash. These stimuli are all clearly visible to the human eye. Overt gaze direction is used a lot in the area of AR assistance systems, where obvious cues are used to point out the next work step or the appropriate tool, for example. In the following, some methods for directing attention with obvious visual cue stimuli are presented as examples.

Seeliger et al. [132] used a simulated industrial assembly task to investigate the effect of different overt visual cue stimuli in AR with HMDs. They examined a total of 8 different stimuli designed to assist participants in assembling screws with nuts and washers. As techniques they used different types of arrows,

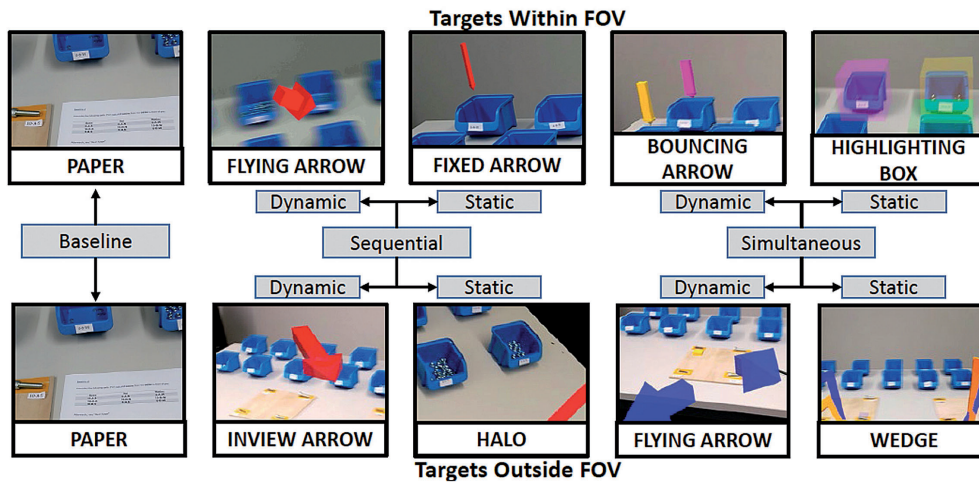


Figure 2.5: Experimental conditions investigated by Seeliger et al. [132]. Visual cues displayed for targets within the FOV (top) and displayed for targets outside the FOV (bottom).

highlighting boxes, halos and wedges. In addition, there were two baseline conditions where a sheet of paper specified which parts from which boxes were needed for assembly (see Figure 2.5). A distinction was made between stimuli pointing to targets within the FOV and stimuli pointing to targets outside the FOV. The cue stimuli were either simultaneous or sequential. A total of 12 participants took part in the study, whose eye movements were recorded using the built-in eye tracking camera of the HoloLens 2². The results of the conducted study showed that the visual stimuli changed the participants' attention. The dynamic stimuli directed attention to the targets better than the static visual stimuli. With these, participants focused more on the stimulus itself than on the actual target. The simultaneous stimuli produced the same effect.

Another approach to directing attention using obvious visual cue stimuli was explored by Biocca et al. [13]. They used attention funnels, which were designed to direct participants' visual attention to the target objects they were looking for (see Figure 2.6). In their study, in which fourteen students participated, the participants had to search for selected objects that were located on tables around them. For each run, participants had to search for and retrieve one selected object. They were supported by the displayed attention funnels, visual highlighting in the form of a 3D bounding box, or verbal descriptions of the object. Among other things, the search time, the error rate, and the mental workload were measured.

²<https://www.microsoft.com/de-de/hololens> (last accessed: 2023-08-15)

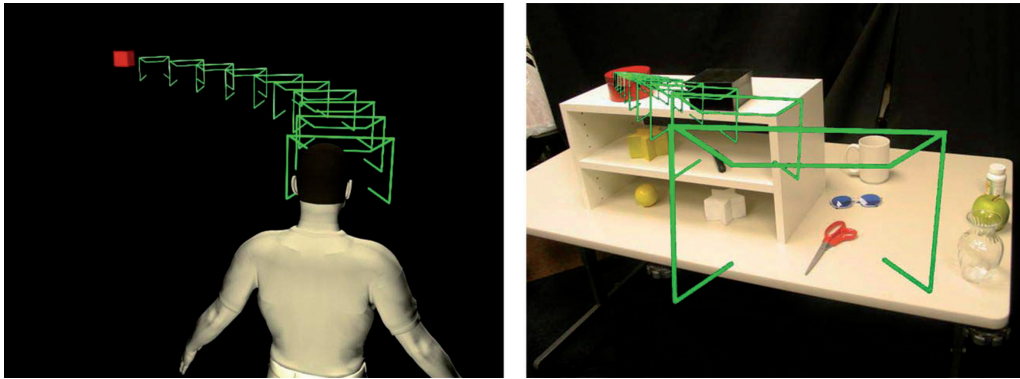


Figure 2.6: Attention funnels directing the gaze of the user to the red target objects [13].

Biocca et al. [13] found in their study that the attention funnel leads to faster search and retrieval times than the comparative conditions and significantly reduces the mental workload compared to the other conditions. However, the use of attention funnels also leads to visual clutter; hence they are not suitable for every purpose.

Khan et al. [76] used spotlights to direct attention on wall-sized displays. In their study with 12 participants, they showed that the performance of a target search task supported by spotlights was significantly better than when only the cursor to be searched was displayed. They obtained the same result when they tested on a standard desktop PC monitor.

Smith et al. [135] investigated an approach to directing attention in immersive 360° environments. They made the environment appear yellow-green when the subject's gaze was no longer on the selected point-of-interest (POI) and displayed everything normally again once the gaze was within 20° of the POI. In this way, they attempted to draw participants' gaze back to the POI. The results of the study showed that while this appeared to work quite well, the method did not lead to a significant improvement in overall gaze performance over the condition without image modulation.

Since in the case of overt gaze guidance the visual stimuli are clearly visible to the participants, guiding the direction of gaze works in most cases. However, the obvious cues can be perceived as very distracting, especially in crowded or visually cluttered environments, which is why more and more research is being done in the area of more subtle methods for directing gaze.



Figure 2.7: Visualization of several modulation thresholds used by Veas et al. from zero modulation (left) to full modulation (right) [151].

2.2.2 Subtle Gaze Direction

In contrast to the obvious stimuli, there is the possibility of directing the gaze through subtle visual stimuli. The methods used for this purpose should be perceived by the viewers as little as possible or not at all and therefore should not interfere with the execution of the task at hand. There are two variants of Subtle Gaze Direction that can be distinguished. In the first variant, mostly area-wide changes are made to the viewed image or video material in order to increase the saliency at the areas of interest and thus to draw attention to these locations. In the second variant, visual cues in the peripheral vision are used to direct the gaze to areas of interest. As soon as there is a change in the direction of gaze to the target region, the stimuli are usually suppressed so that they cannot be perceived in the foveal view. For this purpose, the phenomenon of saccade suppression is additionally exploited, i.e., the stimulus is blanked out during a saccade so that this sudden image change is not perceived.

Gaze Guidance Through Subtle, Area-wide Image Modifications

Over the years, a number of methods have been developed to artificially increase the saliency of areas of interest by modifying the image, thus subtly directing the eye to these areas. Well-known methods include color adjustments to the image and the elimination of distracting regions.

Veas et al. [151] have increased visual saliency of interesting areas in videos to draw attention to these areas. To do this, they use the Saliency Modulation Technique introduced by Mendez et al. [112]. In this algorithm, visual saliency is adjusted based on lightness, red-green and blue-yellow color contrasts. In their studies, the participants had to watch original and modulated videos (see Figure 2.7). The analysis of the recorded eye tracking data showed that the participants looked more often and longer at the areas of interest in the modulated video material, in which the areas of interest had an increased saliency. This also had a positive effect on participants' recall of objects within these areas.

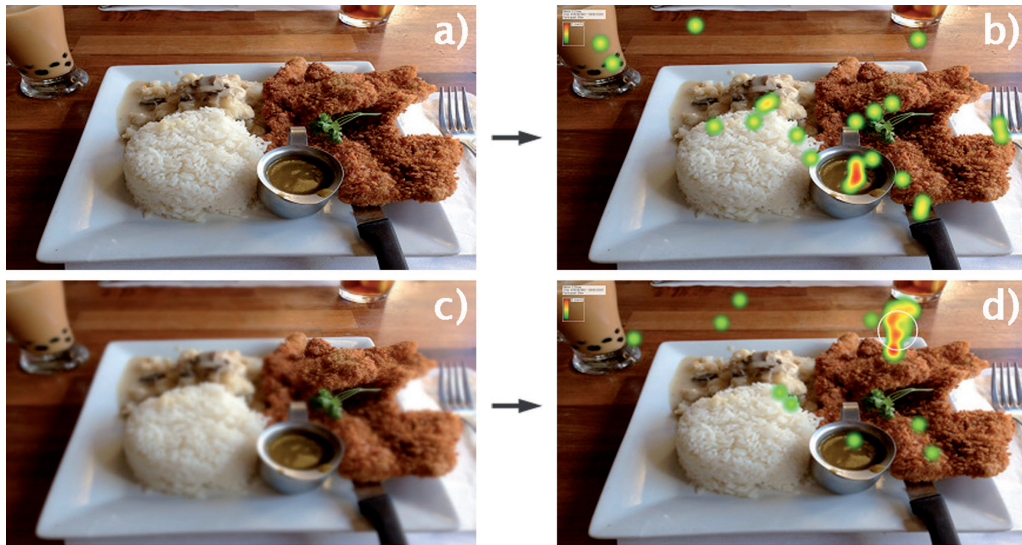


Figure 2.8: Sample image of the experiment by Hata et al. with no blur a) and maximum blur c) and the corresponding heatmaps b) and d). The white circle in d) indicates the unblurred region [50].

In addition to the Saliency Modulation Technique presented by Mendez et al. [112] that Veas et al. [151] used in their study, a number of other methods have been developed to modify the saliency maps of images. There are techniques that only adjust the color of the objects that should be in focus [107] as well as techniques similar to the Saliency Modulation Technique [112] that additionally make changes to the saturation, illumination and sharpness of the object of interest [47, 111, 163]. Hagiwara et al. [47] and Mechrez et al. [111] additionally made changes to the background to bring the areas of interest to the forefront. Gatys et al. [37] used convolutional neural networks to manipulate saliency maps and Yokomi et al. [164] used gradual brightness modulation on the areas of interest to direct gaze to these areas.

In addition to adjusting the color, saturation, illumination, and sharpness of images to direct attention to areas of interest, methods were also developed to remove distracting salient features. Hata et al. [50], for example, developed a system for adaptive subtle gaze direction in which parts of images were blurred to draw attention to the non-modulated areas, which have higher saliency due to the better resolution. They gradually increased the blurring until the gaze pointed in the desired direction and then slowly reduced the blurring again so that the modulation was as unnoticeable as possible. The eye tracking data



Figure 2.9: Example of the luminance modulation. Original image (middle). Image with white color values mixed in (left) and with black color values mixed in (right). Adapted from [6].

recorded during the study showed a clear focus on the non-modulated image areas (see Figure 2.8). The blurring was not perceived by the study participants.

In contrast to Hata et al. [50], Lukashova-Sanz and Wahl [101] used blurring to attenuate only individual image regions containing distracting salient features in the image. They conducted an experiment in VR to see if their method could improve visual search performance in VR. The most salient areas were slightly blurred and the intensity of the blur was adjusted to the saliency in the respective area. The participants' task was to find a Gabor Cross that was displayed at a pseudo-random position with a low saliency. Lukashova-Sanz and Wahl [101] found that blur can lead to a faster location of the provided targets. In addition, the use of blur reduced the rate of total failure by about 40%.

Fried et al. [31] took a different approach than Hata et al. [50] to eliminate the interfering image areas. Instead of blurring, they used inpainting to remove the disturbing salient features in the image.

Gaze Guidance Through Subtle Visual Cues

Bailey et al. [6] were among the first to try to direct human's gaze direction without them being obviously aware of it. Their method, which they call "Subtle Gaze Direction", forms the basis for many other studies, so it is explained in detail below.

In the Subtle Gaze Direction (SGD) method, Bailey et al. [6] used the attentiveness of peripheral vision to control the direction of subjects' gaze. As peripheral stimuli, they used circular changes in brightness or color at the target points in the image by alternately mixing black and white or red and blue color values

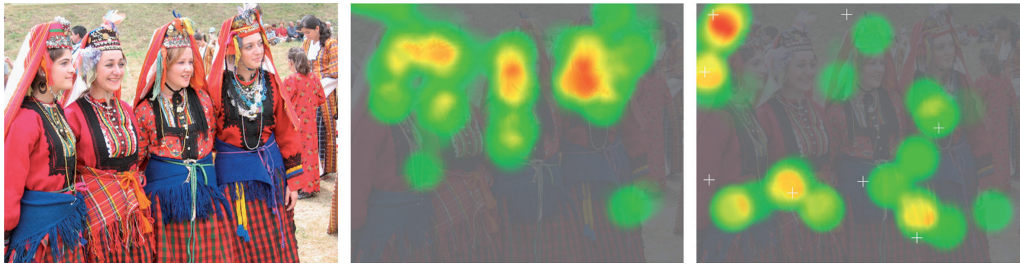


Figure 2.10: Gaze distribution for an unmodulated (middle) and modulated image (right). The white crosses mark the modulated regions. Adapted from [6].

among the pixels of the original image. The color change took place at a frequency of 10 Hz. They determined the necessary color intensity of the cues based on a preliminary study. Figure 2.9 shows an example of how the changes in brightness at an area of interest looked. The study took place in front of a 20 inch computer monitor. The participants sat at a distance of 75 cm from the monitor, so that a visual FOV on the monitor of approx. $30^\circ \times 24^\circ$ was given. The cue stimuli had a size of 0.76° of the visual view, which corresponds to about 1 cm on the screen. The change in brightness of the circles decreased outward to avoid creating sharp edges between the original image and the cue. For this purpose a Gaussian falloff function [165] was used. To ensure that the cue stimuli were never visible in the FOV, they were masked out as soon as gaze moved in the direction of the area of interest. A change in gaze was interpreted as goal-directed if the angle of the vector of the current saccade and the angle of the vector to the target were less than or equal to 10° . Blocking out the stimulus occurred during the saccade so that the change would go as unnoticed as possible. Participants were shown 40 different images in succession for 8 seconds each. In between, a white cross was displayed in the center of the screen on a black background. For half of the participants, the images were unchanged; for the other half, cues were displayed at preselected locations that were not visually prominent. The results of Bailey et al. showed that the gaze was guided by the modulations, but that it did not always reach the target region. Figure 2.10 visualizes the fixation points of the group with and without stimulus influence on an example image. Here it can be seen that the subjects of the test group more often fixated image locations where cues were shown. Bailey et al. also found with their study that luminance modulation (black-white) worked better than warm-cool modulation (red-blue), which is why they used luminance modulation exclusively for their subsequent studies [109, 110, 138].

In follow-up studies, McNamara et al. found that SGD can be successfully used to improve search task performance [109], as well as to contribute to a better understanding of narrative art [110]. Here, they use the SGD method to navigate the viewer through different scenes presented in a single image. The results showed that the participants who received the cue stimuli named fewer scenes in the wrong order than the group that did not receive cue stimuli and the fixations were also much better aligned to the individual scenes. Sridharan et al. [138] found that the use of SGD can also improve mammography training.

While Bailey et al. [6], McNamara et al. [109, 110], and Sridharan et al. [138] tested SGD exclusively on a computer screen, Booth et al. [14] conducted investigations in a real-world environment. For this purpose, they placed various objects on a table standing in front of the participants and projected cue stimuli onto the objects using a projector. According to Bailey et al. [6] they used luminance modulation by alternately projecting a proportion of white and black onto the target object. The modulation was performed at a frequency of 10 Hz. As with SGD, a Gaussian falloff was used to soften the edge of the stimulus. The size of the cue stimuli ranged from 2 to 4 centimeters, depending on the distance of the objects from the projector. Participants were shown different object sequences. During one object sequence, visual cues were displayed on the individual objects one after the other. At the end, it was investigated to what extent the scanpaths recorded with the eye tracker matched those given by the cue stimuli. The results showed that more than one error occurred in the viewing order in only 18% of all trials. Booth et al. thus demonstrated that by projecting cue stimuli onto real objects, the viewing order of these objects can be manipulated.

Waldin et al. [155], like Bailey et al. [6] used color changes in the image to direct the subjects' gaze to the areas of interest. Their so-called "Flicker Observer Effect" was created by changing the pixel values at the areas of interest back and forth between the original pixel values and black. The flicker region was round and the flicker was softened toward the edges by using a dithering mask for the flicker. Waldin et al., in contrast to Bailey et al., use a high frequency flicker that should only be visible in peripheral view but not in foveal view. Ideal intensity values and size of the stimulus were determined in preliminary studies. The task of the participants of the study was to recognize several characters in a cluttered image. An image from the series "Where's Waldo" was used for this purpose. The regions where the target characters were located were modulated. The results of the study showed that the participants were able to find the characters

within a few seconds, which is significantly faster than when this is done without support. Thus, the Flicker Observer Effect has a positive influence on search task performance. However, Waldin et al. also found that the flickering was noticed by the participants. In the foveal FOV, the flicker was rated as acceptable by the participants, but in the peripheral FOV it was rated as distracting, so that it cannot be classified as subtle in this area.

Dorr et al. [22] presented two other techniques to subtly guide gaze direction using peripheral cue stimuli. They displayed small red squares and looming patterns in videos at a distance of 12° to the gaze direction. These were displayed 250 ms before the estimated next saccade. The stimuli were faded out after 120ms or when the subjects made a saccade. The study took place on a 22 inch display at a distance of 50 cm, so that a FOV of about $44^\circ \times 33^\circ$ was spanned. The red rectangle had a size of $1^\circ \times 1^\circ$ and its luminance was proportional to local spatial contrast. Using the looming pattern, a square of 2° size was enlarged with an increasing factor of 1 to 3 over a period of 60ms. This was intended to mimic the visual expansion of an approaching object, thus stimulating the visual alarm system. Dorr et al. concluded that it seems possible to direct a person's scan path. The stimuli used to guide the gaze path were recognized by the participants. Unlike the red square, which was perceived as a red dot, the looming pattern was classified merely as a video artifact.

In a study, Grogorick et al. [43, 44] implemented and tested a variety of the above methods for directing attention in VR and in a custom-built dome projection environment. Besides the original SGD method of Bailey et al. [6], they used the magnification stimulus proposed by Dorr et al. [22]. They used the looming pattern in its original version with enlarged rectangles, as well as implementing a variation thereof with circles whose magnification factor decreased towards the edge based on a Gaussian filter to avoid sharp edges between the stimulus and the environment. A variation of the method with red rectangles as cue stimuli as presented by Dorr et al. [22] was also implemented. Instead of rectangles, this method was implemented with red circles. In addition to the aforementioned methods using cue stimuli in the periphery, Grogorick et al. [43, 44] also used areal spatial filtering as proposed by Hata et al. [50], i.e., they reduced the image sharpness in the non-target regions to direct the line of sight to the targets. Figure 2.11 presents an overview of the visual cues used in the experiment. The analysis of the recorded gaze data for the individual methods revealed that the target region received almost no attention without the use of gaze guidance. All

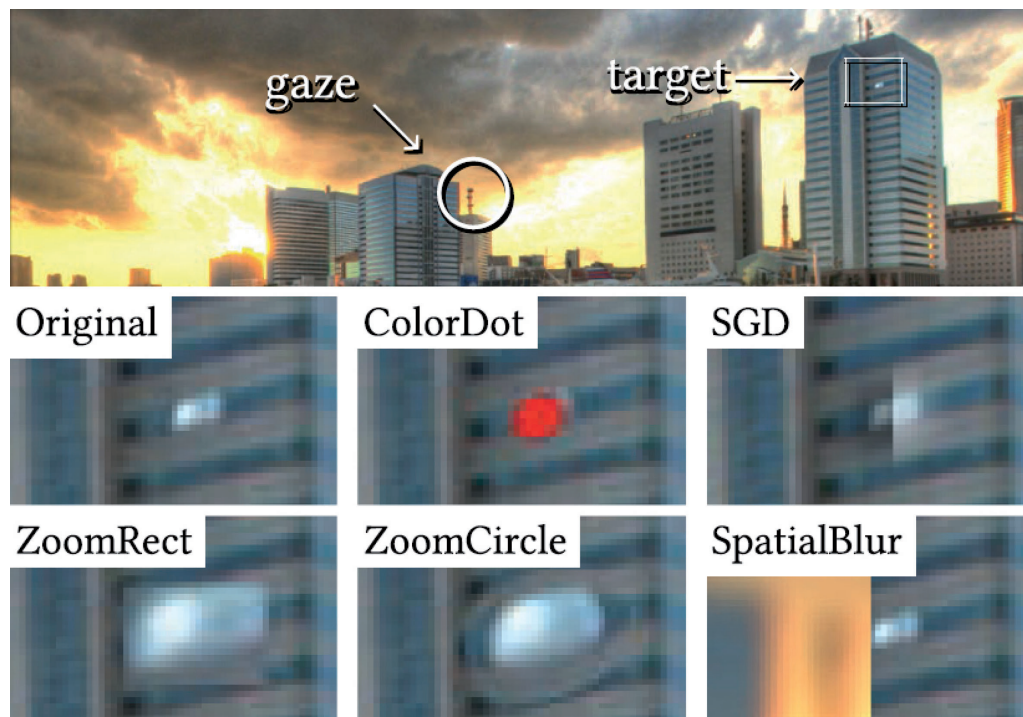


Figure 2.11: Study design of Grogorick et al. Current viewpoint marked with circle and target area marked with rectangle (top). Visual cues examined in the study (bottom). Different states of SGD and SpatialBlur are shown side by side [44].

other methods were able to significantly increase attention at the area of interest. The temporal performance was comparable for the ColorDot, ZoomRect and ZoomCircle methods and showed significant improvements. Only spatial blur performed comparatively worse in terms of time.

In contrast to the previously explained approaches, Lu et al. [99] developed an approach based on increasing the contrast between a virtual target and the surrounding region to point to the target. The virtual targets were black crosses displayed on white squares with varying transparency (see Figure 2.12). The study took place on a computer screen, but served as a simulation for AR applications. The black cross was randomly displayed at eight different predefined positions, both in images and videos, and participants had to find it as quickly as possible. The results showed a significant reduction in error rates and performance compared to the condition in which participants had to solve the task without the white squares being displayed. A follow-up study of Lu et al. [100] with video see-through head-mounted displays yielded similar results.



Figure 2.12: Different contrast levels investigated by Lu et al. from zero (left) to maximum (right). Adapted from [99].

2.2.3 Summary

This section shows that there are already many different approaches to directing visual attention, both obvious and subtle. While the obvious methods work with an explicit highlighting of the target objects, e.g. in the form of arrows or frames, the subtle methods try to ensure that the user ideally does not even notice that his attention is being directed. This has the advantage that people are not bothered by even more visual overload, since there is already enough visual information affecting us.

Most subtle methods have been studied on computer screens or in VR. Only a few experiments that are closer to reality have been conducted so far. Augmented Reality studies could only be found in the field of VST AR with smartphones or head-mounted displays, but not in the field of OST AR, where subjects see reality enriched with virtual objects. However, as AR with AR glasses becomes more and more popular, it is important to take a closer look at this.

2.3 Augmented Reality

The term Augmented Reality refers to the computer-aided extension of reality by virtual elements. By overlaying reality with, for example, texts, images, videos, or 3D objects, a connection is created between the real and virtual world [5, 11, 149]. The technology is already used a lot in areas such as medical applications [150], education [19], and architectural and urban design [137]. The three key principles of AR as defined by Azuma [5] are that it: (1) combines real and virtual, (2) is interactive in real-time and (3) is registered in 3D. The first AR application that fulfilled these key principles was developed by Sutherland [140] already in 1965.

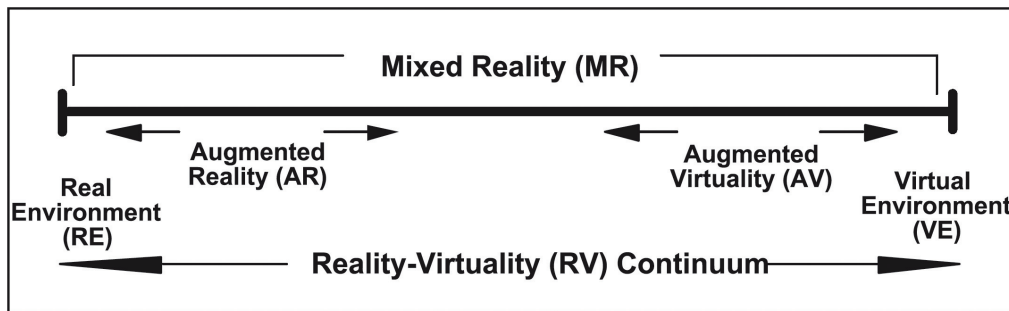


Figure 2.13: The Reality-Virtuality Continuum by Milgram and Colquhoun [113].

There are two forms of AR: VST AR and OST AR. In VST AR, the overlays are superimposed on a reproduction of reality, such as a video stream, in real time [7]. Smartphones and tablets, but also Video See-through Head Mounted Displays (VST HMDs), are suitable for this purpose. In OST AR, on the other hand, the real world is augmented with digital content [120]. This is mostly done with the help of semi-transparent displays, such as in AR glasses. In Section 2.3.2, the functionality as well as the advantages and challenges of OST AR are presented in detail.

Milgram and Colquhoun [113] published a taxonomy in 1999 into which all types of real and virtual environments can be classified. Figure 2.13 shows a diagram of the so-called Reality-Virtuality Continuum, whose idea had already been originated by Milgram and Kishino [114] in 1994. It extends from the real environment without any augmentation, through AR and through Augmented Virtuality, where the digital world is augmented with real-world objects, to the pure virtual environment. The middle range of the continuum is called Mixed Reality and encompasses both AR and Augmented Virtuality.

The systems developed in this thesis use augmentations of reality exclusively, so they are located in the left-middle range of the reality-virtuality continuum. In the studies presented in Chapter 4, TAR is used, which is described in more detail below.

2.3.1 Tangible Augmented Reality

When physical props are used to manipulate or rearrange virtual content in AR, this is called TAR. By coupling digital information with physical objects [59], manipulations of virtual objects can be faster [10] and more accurate [144]. By

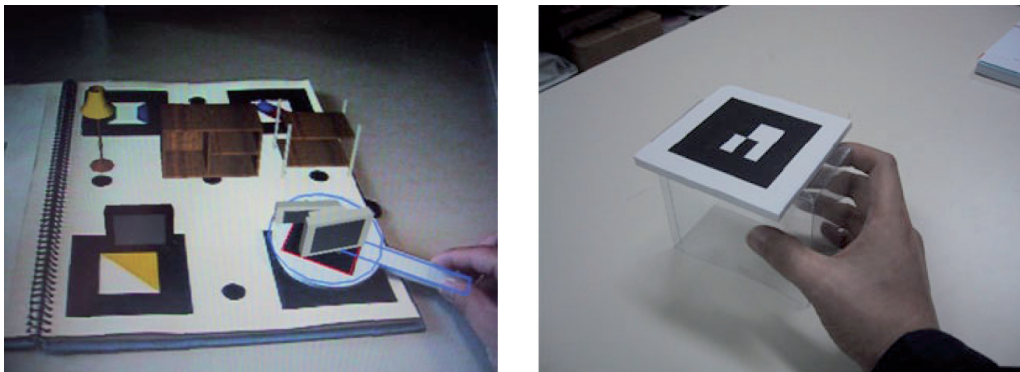


Figure 2.14: The TAR controller applications *Magic Paddle* [74] (left) and *MagicCup* [73] (right).

using physical props, TAR enables intuitive and natural interaction [12] and opens up multiple new use cases [20].

The concept of TAR, which combines the visualization capabilities of AR with the intuitive physical manipulation offered by Tangible User Interfaces, was defined by Billingham et al. [12] in 2008. Prominent works that have used tangibles in AR to interact with virtual objects include the *Magic Paddle* [74] and the *MagicCup* [73]. The *Magic Paddle*, which was developed by Kawashima et al., is a physical paddle that serves as a controller to manipulate virtual objects, such as picking them up, moving them, or deleting them (see Figure 2.14, left). The *MagicCup* is a similar controller developed by Kato et al. in the context of cityplanning. It is a transparent, upside-down cup that is placed over the virtual object to select it. Afterwards, the selected virtual object can be picked up and manipulated on the tabletop (see Figure 2.14, right).

In addition to using physical controllers for interaction, there are also applications that allow more direct interaction with virtual objects. One example is the Cubical User Interface developed by Lee et al. [94]. It consists of two tangible cubes, which have magnets on the sides and buttons on the corners. On all sides of the cubes there are additional optical markers. By tracking the markers, the cubes can be overlaid with virtual objects and serve as a two-handed interface to build models from virtual parts (see Figure 2.15, left). In Choi's research [20], TAR was used for product usability assessment. In his study, interaction with a simplified prototype of a space heater overlaid in VST AR was compared to interaction with the real device and interaction with a pure AR representation (see Figure 2.15, middle). The TAR representation outperformed the pure AR representation. Lee and Park [95] also used TAR to prototype product designs.



Figure 2.15: Example TAR applications: Cubical User Interface [94] (left), TAR for product usability assessment [20] (middle), and Augmented Foam [95] (right).

Their application *Augmented Foam* allows designers to create a mock-up prototype from foam. Using fiducial markers placed on the mock-up, a detailed visual overlay is displayed on the tangible object and designers can explore it intuitively (see Figure 2.15, right).

A drawback of the system presented above is that there are dropouts in the AR visualization when covering the markers with the hands. Other TAR applications try to minimize this drawback by combining optical marker tracking with other methods. Düwel et al. [23], for example, use multiple tangible cubes to build a composing object from them. In addition to optical markers, they use embedded computing to determine which cubes are connected to each other on which side. Through this, it is sufficient that only a single marker on a single cube is detected to display the composed virtual objects at the correct position. Bozgeyikli and Bozgeyikli [15] used a combination of the optical markers and the electromagnetic tracking of the controller of the Magic Leap 1 AR glasses³ to place the virtual object as precisely as possible at the position of the tangible prop. The tangible prop consisted of a cube with image markers on each side and the Magic Leap 1 controller inside.

If no completely precise position of the virtual object is required for the application, tracking can also be done exclusively with the help of appropriate controllers/trackers, such as the Magic Leap 1 controller [129] or the HTC VIVE tracker⁴, as in the TAR application by Englmeier et al. [25].

Especially when both hands are free for interaction, which is the case when AR headsets are used, a variety of novel use cases can be implemented. In most applications for AR headsets, interaction is still done through controllers, hand gestures or speech [18, 98, 118] and the applications are mostly used exclusively

³<https://www.magicleap.com/ml1-devices> (last accessed: 2023-08-15)

⁴<https://www.vive.com/de/accessory/tracker3/> (last accessed: 2023-08-15)

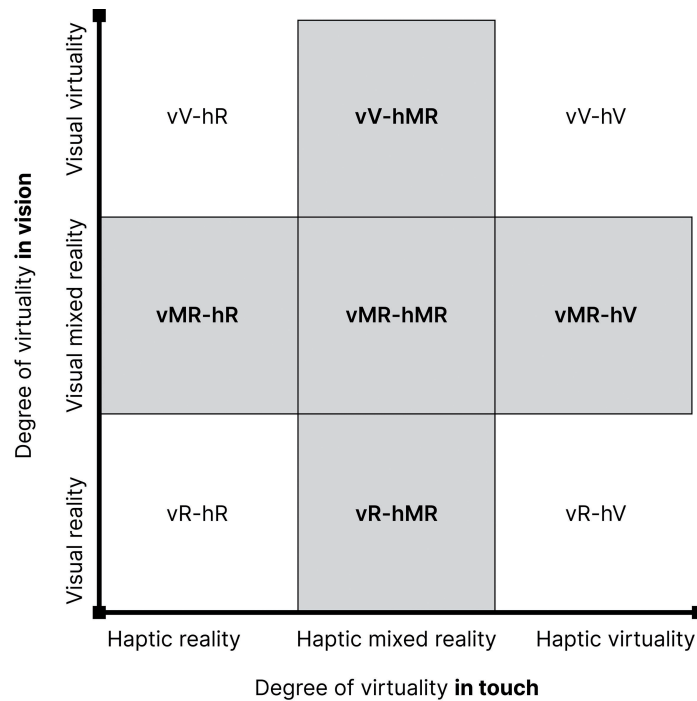


Figure 2.16: Visuo-haptic Reality-Virtuality Continuum developed by Jeon and Choi. Adapted from [61].

to display 3D objects, such as in construction [57]. By using physical proxy objects to interact with the virtual content, 3D visualizations can not only be displayed, but also, for example, new composite objects can be created together. Examples of this are the creation and modification of buildings, or jointly touchable city planning in TAR. In AR games, the gaming experience could be made even more realistic by using tangible objects, e.g. in the form of touchable game figures [56].

Jeon and Choi have developed a taxonomy in which they have added the haptic dimension to the Reality-Virtuality Continuum of Milgram et al. [113, 114]. In their so-called *Visuo-haptic Reality-Virtuality Continuum*, one axis describes virtuality in vision while the other axis represents virtuality in touch. The axes are each divided into three sections, so that there are a total of nine possible combinations (see Figure 2.16). The continuum of Jeon and Choi ranges from pure real environments without a synthetic stimulus (vR-hR) to pure virtual environments with pure virtual haptic sensations (vV-hV). Tangible Augmented Reality in this context belongs to the left center at vMR-hR, because here real objects are interacted with in the visual Mixed Reality.

2.3.2 Optical See-through Augmented Reality

The use of OST AR offers many advantages, but there are also special characteristics that must be taken into account when developing applications for OST AR HMDs. Whether VST AR or OST AR is better suited for an application always depends on the specific use case. Neither technique is perfectly suited for every imaginable use case. A major advantage of OST AR glasses is that the transparent display allows an almost unobstructed view of the real world, unlike VST AR where the virtual content is superimposed on a video stream [7, 120]. Therefore, OST AR is particularly suited for critical applications where a live view of reality is essential, or for applications where multiple people are working collaboratively, as line-of-sight communication is possible.

A special feature of current OST AR HMDs is that due to the technology used, the virtual objects are always slightly transparent and let real-world objects shine through. This is why findings from the areas of VST AR and VR cannot be transferred directly to OST AR and new investigations must be carried out specifically for this area. Additionally, ambient lighting has an effect on viewing in OST AR HMDs because in the additive lighting model used in current OST AR HMDs, the light emitted from the display is added to the existing light from the physical environment [32]. Gabbard et al. [32, 33, 34] have shown that background, text drawing style, and interaction have an impact on the readability of AR content. In their studies, they use OST AR systems to evaluate a number of different text drawing styles on different natural backgrounds.

The background also affects how much ambient light hits the OST AR system [34], which also has an impact on the legibility of the text. Erickson et al. [27] determined the impact of environmental lighting conditions on the contrast of AR content. In their study, they measured the amount of light entering the eye under 6 different ambient illuminance levels, ranging from less than 1 *lx* (lux) to over 20000 *lx*, covering both indoor and outdoor lighting conditions. The results show a significant decrease in perceived contrast with increasing illuminance, such that it is almost completely eliminated in bright outdoor environments.

2.3.3 Summary

In this section the basics of AR were described. In particular, TAR and OST AR were presented, which we used in our studies. Augmented Reality offers a good possibility to enrich reality with additional virtual information. If the

virtual content is also to be interacted with, physical tangibles are well suited for this purpose, as they enable a more intuitive interaction. By using OST AR glasses, both hands are free for interaction and the user can still interact with the outside world through the transparent display. The display of virtual content in the OST AR glasses has specific characteristics and changes depending on the environmental lighting. Accordingly, results of studies conducted in VR or VST AR cannot be transferred directly to OST AR. Instead, specific investigations are needed that also take the characteristics of this technology into account.

2.4 Haptic Perception

In order to be able to evaluate how interaction with objects in OST AR is perceived, it is necessary to understand how haptic perception works. In the following, we first illustrate the anatomy of the somatosensory system and explain the process of object identification by haptic exploration. Then we describe the intersensory intersection of haptic and visual perception. At the end we summarize the most important aspects.

2.4.1 Anatomy of the Somatosensory System

Haptic perception is considered to be a combination of cues provided by tactile (cutaneous) and kinesthetic (proprioceptive) receptors during active touch or manipulation of objects [41, 92]. Kinesthetic receptors include mechanoreceptors, which are located in muscles, tendons, and joints. They enable the perception of movements of the body and limbs. Cutaneous receptors include mechanoreceptors that respond to pressure or deformation of the skin and thermoreceptors that respond to thermal stimuli of the skin. They enable the perception of touch, vibration, tickle, and pain [41, 42]. Cutaneous receptors are located throughout the body in the skin, which is the largest sensory organ of the human body [41, 51].

The skin consists of two major layers, the epidermis and the dermis, in which the ends of the mechanoreceptor units are located. Close to the surface of the skin, near the epidermis, are the Merkel receptor and the Meissner corpuscle. Deeper in the skin are the Ruffini organ and the Pacinian corpuscle (see Figure 2.17). The Merkel receptor and the Ruffini organ adapt slowly and send a continuous response to sustained skin deformation. This enables the perception of shape and texture (Merkel receptor) as well as the perception of stretching of the skin [42, 51].

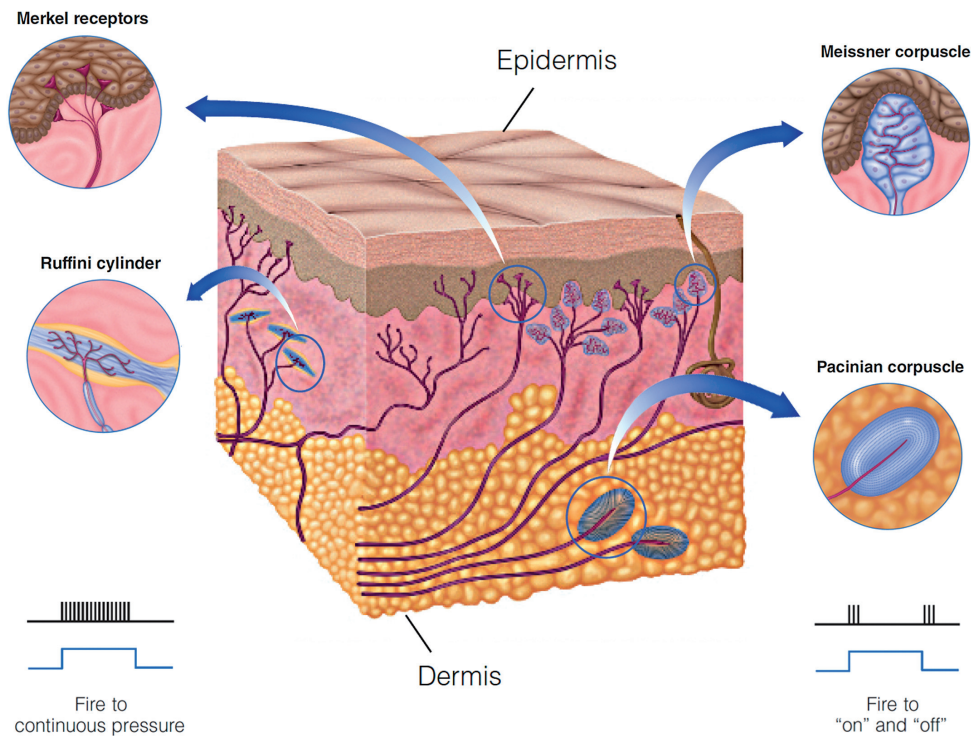


Figure 2.17: Layer of the skin with four different mechanoreceptors. Adapted from [42].

The Meissner corpuscle and the Pacinian corpuscle, on the other hand, adapt rapidly and react to the appearance and sometimes also to the end of a skin deformation. This allows the Meissner corpuscle to control e.g. handgrip or the perception of movement on the skin. The Pacinian corpuscle enables the perception of vibration or fine textures when the finger is swiped over it [42, 51]. The accuracy of tactile perception varies strongly across the body surface and is strongest at the fingertips and weakest at the back. For fingertips, spatial acuity decreases with age, researchers found [40, 152].

2.4.2 Haptic Exploration of Objects

Haptic perception is a very complex process. Various systems are needed to identify an object [42]. The sensory system is responsible for recognizing touch and temperature as well as positions and movements of fingers and hands. The motor system is needed to enable movement of the fingers and hands. And finally, the cognitive system is used to interpret the sensory information [42].

When all these processes work together to actively detect an object, this is called active touch. In contrast, passive touch is when touch stimuli are applied to the skin, e.g., when something is pressed against it [42, 51, 92].

Research has shown that people are able to recognize very familiar objects within one to two seconds [77]. For identification they mostly use a number of distinctive movements, which have been named by Lederman and Klatzky [90, 92] as Exploratory Procedures. Exploratory Procedures are various hand movements that are executed to determine a specific property of an object [41]. For example, contour following and enclosure are used to accurately determine the shape of an object [42, 92], or pressure is used to determine the hardness [41, 92].

2.4.3 Intersensory Interaction

When we interact with an object, we use our senses to gather information about the object's geometric properties, such as shape, size, orientation, and curvature as well as its material properties, such as temperature, compliance, texture, and weight [159]. Both vision and touch contribute to our perception of these properties. Studies have shown that texture information of objects can also be perceived through audition. Auditory cues, such as the sound of fingers rubbing against a surface, can provide valuable information about the roughness of a surface [89, 91].

The brain integrates information from multiple sensory modalities to come up with the most accurate estimate of an object's properties. This integration is based on the relative reliability of each sensory cue, with more reliable cues being assigned a greater weight [53]. The relative weighting of vision and touch additionally varies depending on the specific task being performed.

In laboratory situations, vision mostly dominates over touch [39, 127]. However, under certain conditions, touch showed a dominant role over vision [46, 93]. When geometric properties are assessed, vision tends to be weighted more heavily than touch. However, when material properties such as roughness are assessed, touch provides a more accurate assessment than vision [92].

An example experiment demonstrating the so-called Visual Dominance Effect, that is, the tendency of vision to dominate other senses in the perception of spatial and physical information [123], is the rubber hand illusion. In this experiment, a participant's real hand is hidden from view while a rubber hand is placed in front of them. The experimenter simultaneously strokes the participant's hidden

hand and the rubber hand with a paintbrush, creating the illusion that the rubber hand is actually a part of their own body. Through this, participants experience a change in their physical self-awareness as they begin to integrate the rubber hand into their body schema [49]. The rubber hand illusion and other studies of visual dominance illustrate the complex relationship between vision and touch in our perception of the body and the environment.

2.4.4 Summary

In this section, we have presented an overview of haptic perception, which is provided by tactile and kinesthetic cues. It is a complex process involving a variety of different receptors under our skin that respond to both continuous and on/off stimuli. By combining different hand movements during active touch, we are able to recognize objects and their properties. However, not only does haptic perception play a decisive role in determining object properties, but also visual perception, which often predominates in our overall perception. This must be taken into account when designing studies in which properties of objects need to be identified and evaluated.

2.5 Proxy Interaction

Physical proxy objects are very suitable for interacting with virtual content because they provide a very intuitive interaction with the virtual objects [12] (see Section 2.3.1). In the past, little research has been done regarding the acceptable level of difference between physical and virtual objects to ensure a satisfactory user experience. In the following section, we discuss relevant studies that have examined the extent to which a proxy object can deviate from its virtual representation in VR, VST AR, and OST AR. We also present different ways of tracking objects in space, which is necessary for conducting studies.

2.5.1 Proxies in Virtual Reality

Most research regarding proxy interactions and possible size differences between physical and virtual objects has been carried out in VR. Simeone et al. [133] investigated how large the discrepancy between a physical proxy and virtual element in VR can be without breaking the VR illusion. In their first study, participants



Figure 2.18: Virtual substitutions of the mug (left) and physical props (right) used in the study of Simeone et al. [133].

interacted with various objects in VR. The room, which was set up like a living room in reality, represented a medieval courtyard in VR. For each virtual object in VR, there was a physical counterpart in the living room that could be touched and manipulated. The participants' task was to interact with a mug, which was substituted in different ways in VR: a matching virtual replica, a virtual model with aesthetic differences, a model where a part was added or omitted, a functionally different model and a virtual model with categorical differences, where there is no longer a connection between physical and virtual object (see Figure 2.18, left). They found that differences in shape and perceived temperature make the object seem significantly less credible than an exact replica. The same applied to substitutions with smaller virtual objects, while they found no significant difference in believability for larger virtual representations. In their second study, several tangible objects were mapped onto a virtual lightsaber with which participants were asked to hit floating spheres. Participants interacted with a flashlight, an umbrella, and a toy lightsaber (see Figure 2.18, right). The flashlight was reported to be the least tiring and received the highest rating for the feeling of wielding a real lightsaber. It was also most preferred by participants because of its light weight. The authors conclude that significant differences between the physical object and the virtual substitute in the parts that are interacted with the most have the strongest negative effects on believability.

De Tinguy et al. [21] investigated how similar virtual and physical objects must be in order to feel the same. They considered variations in width, local orientation and local curvature. Their study consisted of three parts, in which a certain object had to be touched each time (see Figure 2.19, left). Only the virtual overlay changed based on one of the three features (see Figure 2.19, right). To ensure that participants touched the physical object even if it had a different size, they used a warping effect to redirect the virtual finger position in VR. The largest possible discrepancies that remained unnoticed by the users were determined in terms of

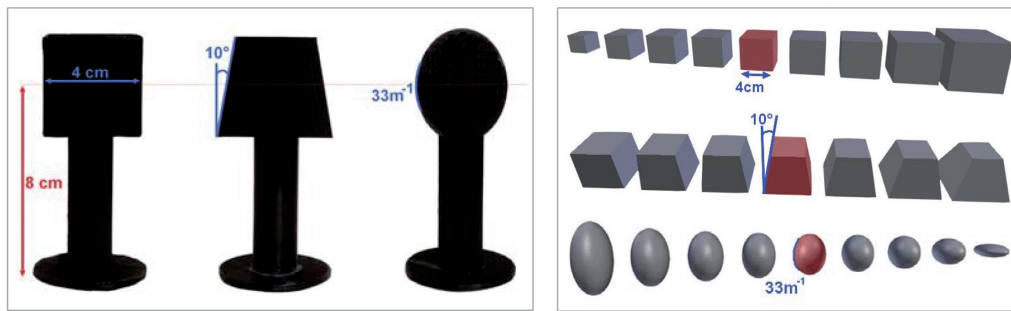


Figure 2.19: Study setup of the experiment by de Tinguy et al. [21]. Tangible objects (left). Virtual objects with varying width, varying orientations and varying curvatures (right). Adapted from [21].

local curvature. Differences were only detected at a discrepancy of 66.66%. For orientation, no differences were detected up to a discrepancy of 43.8%. According to the study's findings, it is also possible to change the width of objects by up to 5.75% in VR without the user noticing any difference.

Bergström et al. [9] also discovered that it is possible to use smaller and larger physical objects as proxies for interacting with virtual objects without the size difference being noticed in VR. Similar to de Tinguy et al. they use a method to manipulate the position of fingers, which they called *Resized Grasping*. These and similar methods are only possible in purely virtual environments and cannot be transferred to AR, where users see their own real hands.

Based on the results of previous studies, Nilsson et al. [119] established three criteria that a proxy object must meet in order to be successfully used in VR: (1) sufficient similarity, (2) complete co-location, (3) compelling contact forces. Physical objects to be used for interaction in VR should be as similar as possible to the virtual object being controlled in terms of haptic properties, such as shape, size, and weight. In addition, it is important that the proxy object is co-located with its virtual counterpart. Furthermore, appropriate stimuli should be provided when forces in VE are applied to an object that the user is holding. How important the individual criteria are always depends heavily on the VR application. There are use cases that require very high precision and for which compliance with the criteria is therefore very important. Other use cases, however, also allow small deviations, e.g. in the case of entertainment applications [119].



Figure 2.20: Physical props used in the study of Kobeisse and Holmquist [79].

2.5.2 Proxies in Video See-through Augmented Reality

In VST AR, there are also already initial investigations into possible differences between physical proxy objects and their virtual representations. Kwon et al. [85] conducted two experiments using a VST HMD to investigate the effects of shape and size differences between virtual objects and physical props in TAR. Three physical objects were used in the experiments: a flat cuboid, a cube, and a larger cuboid that were each overlaid with three different virtual objects: a floppy disk and two different-sized cosmetic boxes. The participants had to observe the objects and match a given orientation as quickly as possible. The results showed that performance was better when the virtual object matched the physical object. Measures of user perception and manipulability also indicated higher realism and better perception of physicality when shape and size matched. However, it remained unclear whether the observed results were primarily influenced by size or shape. To remove this uncertainty, an additional experiment was conducted using the same props but varying the size of the virtual objects while keeping the shape constant. However, they found no significant performance differences among the five different size conditions. Based on these findings, the researchers concluded that the observed results in the main study were primarily influenced by the shape differences rather than the variations in size.

Kobeisse and Holmquist [79] explored the design possibilities of using generic objects as substitutes for historical artifacts within the context of cultural heritage in VST AR. In the user study conducted for the paper, participants interacted with a virtual 3D model of a Bronze Age urn using four different interfaces: a touch screen, a flat AR marker, a generic wooden cylinder, and a 3D-printed replica of the digital artifact (see Figure 2.20). The findings show that the 3D-printed replica, which closely resembles the physical object, offers the most realistic method of interaction. However, the study also suggests that using a generic object like the wooden cylinder can provide a more immersive experience when compared to the touch screen and flat AR marker. Overall, the results indicate that tangible



Figure 2.21: Virtually overlaid everyday proxy objects determined by the algorithm of Hettiarachchi and Wigdor [55] (right). Scene without overlays (left).

interfaces enhance engagement and offer a more authentic experience compared to traditional observation without tactile interaction.

2.5.3 Proxies in Optical See-through Augmented Reality

In OST AR, very few studies on proxy interactions have been performed so far. Hettiarachchi and Wigdor [55] introduced a system called *Annexing Reality*, which matches ordinary objects as tangible interfaces to virtual objects. They used a Kinect camera⁵ to analyze the objects in the physical environment and the OST AR HMD Meta 1⁶ to overlay the virtual objects onto the identified physical objects. Physical objects were selected as tangible interfaces if they were similar in shape and size to the virtual object. To ensure an even better match, the size of the virtual objects was subsequently adjusted to that of the physical objects (see Figure 2.21). By identifying matching proxy objects in the environment, their approach enables users to experience TAR without requiring specialized props for each application. The authors evaluated their *Annexing Reality* graphical tool in a developer study in which experts rated it as easy to learn and use. Similarly, Szemenyei and Vajda [141, 142] developed algorithms that enable automatic matching of everyday physical objects with virtual objects. In both the system of Hettiarachchi and Wigdor [55] and the system of Szemenyei and Vajda [141, 142], the matching process primarily focuses on the shape of the objects. Hettiarachchi and Wigdor's system even automatically resizes the virtual object to fit the physical object, which is not feasible in many use cases.

⁵<https://www.vrnerds.de/kinect-v2/> (last accessed: 2023-08-15)

⁶<https://developers.shopware.com/blog/2015/05/04/unboxing-meta-augmented-reality-glasses/> (last accessed: 2023-08-15)

2.5.4 Object Tracking

In order to conduct studies on the interaction with physical proxy objects in OST AR, it is necessary to know exactly where the physical object is located in space at all times in order to be able to overlay it with the appropriate virtual content. The position of objects in space can be determined using object tracking systems, which can be either sensor-based or vision-based [97].

Sensor-based tracking methods use specific sensors to track objects in the environment. Known sensor-based tracking methods used for object tracking include GNSS (Global Navigation Satellite System), IMU (Inertial Measurement Unit), RFID (Radio Frequency Identification), BLE (Bluetooth Low Energy), Ultrasonic, LIDAR (Light Detection and Ranging) and magnetic tracking systems [82, 97, 156, 162]. Tracking using GNSS requires portable GNSS units to be installed on the objects to be tracked. By measuring the distances between the GNSS receiver and at least four satellites with known positions, the receiver is able to calculate its own position [97]. IMU-based tracking systems use a variety of sensors to measure the linear and rotational movements of an object, enabling them to estimate the position and orientation of objects in space [97]. RFID systems determine the position of objects by using specific readers to detect the RFID tags attached to the objects [97]. In BLE systems, small devices (beacons) are attached to objects or placed in the environment and regularly emit signals. If the signals are detected by a device, its position can be estimated [97]. Ultrasonic systems emit sound waves and measure the time until the reflected waves return. If there are several ultrasonic sensors that determine the distance to objects, the position of objects can be detected and tracked [82]. LIDAR-based tracking works in a similar way. Instead of sound waves, laser pulses are emitted and the time until the reflected light returns is measured. The returning laser points can be analyzed to determine the position and structure of objects in space. In magnetic tracking, magnetic field sensors measure the magnetic field around an object and can thereby detect the position and movement of objects in space [156, 162].

Vision-based object tracking systems use cameras or visual sensors to track objects. The visual data captured by the cameras is analyzed and interpreted by computer vision algorithms to recognize and track the objects. Methods used for vision-based object recognition include marker-based tracking, feature-based tracking, and object recognition and tracking [1, 124, 136, 145]. In marker-based tracking, predefined markers are applied to the objects to be tracked. These markers can be fiducial markers, such as QR codes, or colored patterns, which are tracked

by the system and used to determine the position and rotation of the respective object [1, 124, 145]. In feature-based tracking, specific visual features of objects, such as corners, edges, texture, or distinctive points, are tracked to determine the position and orientation of objects [1, 124, 136, 145]. Object recognition and tracking uses machine learning approaches to recognize and track objects [136].

Both sensor-based and vision-based tracking have their advantages and disadvantages, so there are approaches that combine the two to achieve even more accurate results. One example is SLAM (Simultaneous Localization and Mapping), which combines information from camera images with, e.g., depth information or sensor information from IMUs. The system not only determines the position of the object in space, but also simultaneously creates a map of the surroundings [4].

The choice of the appropriate system must depend on the requirements of the application, the spatial conditions and the objects to be tracked. Sensor-based tracking systems are often used when tracking with vision-based systems is difficult. This is the case, for example, when objects are occluded or when poor lighting conditions are present. When choosing the tracking method, whether the objects are to be tracked in a fixed environment or in dynamic scenarios, also plays a significant role. Stationary object tracking systems are suitable for controlled environments. These are permanently installed and focused on a specific area, and hence they usually have a high level of accuracy. If the system is to be used in different environments, for example, which requires easy movement of the system, mobile object tracking systems are more suitable. It must be considered individually for each application which requirements are relevant and the choice of the system must be adapted accordingly.

2.5.5 Summary

In this section, we have presented the relevant work investigating feasible differences between physical proxy objects and their virtual representations in VR, VST AR, and OST AR. Only a few studies on shape and size differences have been conducted in VR and VST AR so far. The studies in VR have investigated whether or to what degree the physical object can differ from the virtual one without the difference being noticed. For studies in OST AR, however, it is of greater importance to what degree the difference is tolerated, because due to the technology, the overlays are always a bit transparent, so that one always sees the physical object in the background to a certain degree. In the studies in VR,

also finger redirection techniques were used in order to improve the realism and accuracy of interactions, which is not possible in OST AR, where one sees the real hand instead of a virtual hand. Thus, not all results from VR can be directly transferred to OST AR. What can be assumed based on the results of the studies in VR and VST AR is that the best results are achieved when the physical object is a replica of the virtual object. Additionally, it can be assumed that the criterion of complete co-location established by Nilsson et al. [119] is also applicable to OST AR. For the studies in OST AR, where we want to investigate how differences in size and shape between virtual and physical object affect perception, it is important that the virtual object is displayed at the exact position where the physical object is located in space. In this section, we have presented several methods that can be used to track objects in space. In order to display the virtual object at the exact position, highly accurate tracking is required, which can be provided by a professional stationary motion tracking system.

Chapter 3

Understanding the Perception of Visual Cues for Gaze Guidance

In this chapter, we explore how people can be pointed to relevant content as subtly as possible in real-world environments. We first present a study that investigates the distance from the current line of sight at which different objects are detected and how close the objects must be to the fixation point to be able to see detailed information (see Section 3.1). We then present different studies in which we investigated how well certain visual stimuli are suited to direct attention. We conducted one study in a real-sized projected environment (see Section 3.2.1), two studies in an instrumented environment (see Section 3.2.2), and one study in Augmented Reality (see Section 3.2.3).

The two studies in the instrumented environment were conducted as part of a Software Campus project funded by the German Federal Ministry of Education and Research.

3.1 Peripheral Perception of Visual Cues

As mentioned in Section 2.1.1, numerous studies have already been performed on perception in the periphery. Using novel methods, we wanted to specifically investigate how limited perception is in the FOV with respect to different objects. For this purpose, we investigated in a study how close the different objects have

to be to the visual fixation point in order to be able to perceive them at all and how close they have to be in order to be able to identify their details perfectly.

Hypotheses

The focus of the study conducted was to show that the appearance of an object has an influence on the detection of the object in the periphery. Furthermore, we wanted to check whether an object needs to be closer to the fixation point in order to make a detailed statement about it, rather than just perceiving it. We also wanted to investigate to what extent the level of detail of the object or the difficulty of the recognition task has an influence on the angle from the fixation point from which it can be recognized in detail and the task can be solved. Therefore, we formulated the following hypotheses:

- H1: The appearance of an object affects the angle from which it can be seen or recognized in detail.
- H2: The recognition of object details is possible only at shorter distances to the fixation point, compared to the distances required for just perceiving the object.
- H3: The difficulty of the task influences the angle from which it can be solved.

Study Task

We conducted a study to test the above hypotheses. The task in the study was to notice objects displayed in the FOV and to identify their details. While the participants fixated a cross in the center of the TV screen with their eyes, objects from eight different directions were presented one after the other (see Figure 3.1). These objects gradually approached the fixation point until they were identified by the participants. A total of six different object types were tested, for which participants had to indicate when they saw them and when they could identify their details correctly. The object recognition tasks were divided into peripheral and foveal tasks. Foveal tasks included detailed objects that we assumed could only be correctly identified in the foveal area.

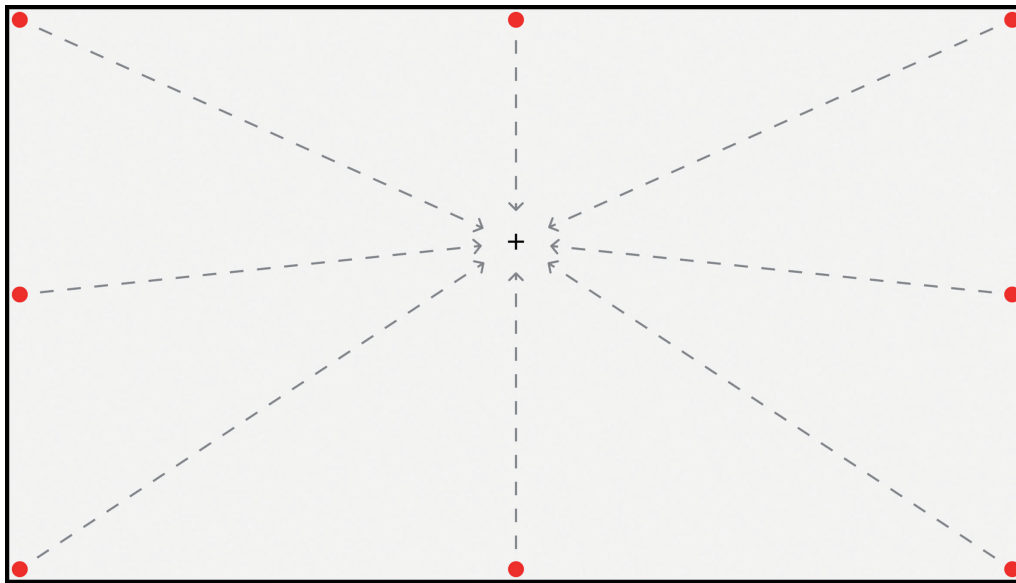


Figure 3.1: Illustration of the study task on the TV screen: Objects appearing from eight different positions while the participant fixates on the cross in the middle.

Participants

12 student volunteers participated in our study. All participants had normal vision, which we checked with Landolt ring vision tests⁷ [116] and color vision tests⁸ before each study run.

Study Setup

The study was conducted on a 55 inch TV with a screen width of 110 cm and a screen height of 61.5 cm. The subjects' heads were fixed at a distance of 25 cm centrally in front of the TV using a chin rest. The display thus covered approximately 65.6° of the horizontal visual FOV in each direction and approximately 52.4° of the vertical visual FOV in each direction.

Participants had to fixate a cross shown in the center of the display throughout the study. In order to check that the participants adhered to this, the eye positions were recorded with the help of a Pupil Pro eye tracker [72] and if the participants (accidentally) looked away, the trial was interrupted until the focus was back on

⁷<https://www.onlinesehtests.de/sehtest-kreise-landolt-ringe.php> (last accessed: 2023-08-15)

⁸<https://web.archive.org/web/20201208160704/http://www.dfisica.ubi.pt/~hgil/p.v.2/Ishihara/Ishihara.24.Plate.TEST.Book.pdf> (last accessed: 2023-08-15)

the cross. To ensure that the eye tracker provided meaningful positional data, a calibration was performed with each participant at the beginning.

Design and Procedure

The study was designed as a within-subjects experiment. The order of the tasks was counterbalanced by a Williams design Latin square of size three [161] for both the three foveal and the three peripheral tasks.

The order of the positions from which the objects approached was randomly determined. For the peripheral tasks, the initial display of the objects was as far away from the fixation point as possible and then got closer and closer to it. As soon as the participants could see the object in their periphery, they confirmed this with a left mouse click. As soon as the corresponding detail question could be answered, e.g., how many edges the presented object had, the left mouse button was pressed again. In the case of foveal tasks, the objects were initially displayed at a shorter distance from the fixation point, since the focus here was on completing the task, and it could be assumed that the objects had to be located in the foveal area in order to answer the questions correctly. The size of the objects that were superimposed corresponded to 2° of the field of vision in each case. A very light gray (#F6F6F6) was chosen as the background color.

Peripheral Tasks The peripheral tasks included the correct recognition of Landolt rings, the recognition of basic colored shapes, and the recognition of a pulsating circle (see Figure 3.2, left).

For the Landolt rings, participants had to press the mouse button as soon as they saw them and then again as soon as they could name in which direction the opening of the ring was pointing. In the case of the colored objects, the first step also involved naming the moment when the objects were seen. Next, participants were asked to click once they could name the color of the object and a third time when they could also name the shape. The choice of shape (rectangle, circle or triangle) as well as the choice of color (red, green or blue) was randomized. The flashing circles were medium gray (#CCCCCC) and thus had a relatively low contrast to the background color. Here, participants were instructed to click as soon as they could perceive the presence of a circle and indicate one of the 8 directions from which it appeared.

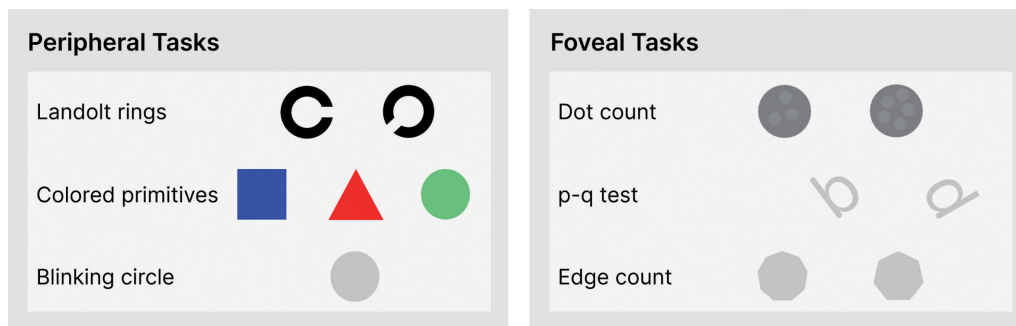


Figure 3.2: Illustration of the peripheral (left) and foveal tasks (right). The displayed objects are only examples for each type of task.

Foveal Tasks The foveal tasks included a task in which small dots had to be counted on a circle, a p-q test, and a task in which the number of edges of a polygon had to be named (see Figure 3.2, right). For each circle, one to five small circles were shown with a low color contrast to make the task as difficult as possible. The participants had to click the left mouse button when recognizing the correct number of small circles. In the p-q task, the task was to recognize whether the object displayed was a p or a q. The letters were randomly rotated and the user had to click as soon as he or she recognized whether the object was a p or a q. In the third task, medium gray polygons were displayed that were randomly rotated and had 7 to 10 edges. The number of edges was intentionally chosen so high that close looking or counting was required. Participants clicked as soon as they could name the number of edges.

Results

Figure 3.3 shows the distances at which the different object types were seen (blue), the distance from which the color could be correctly recognized (red; only colored primitive shapes) and the distances at which the associated task(s) could be answered correctly (green). The objects were mostly seen directly as soon as they appeared at the edge of the display area. Only the flashing gray dot and the p's and q's were seen a little later. As soon as the flashing dot was seen, the correct direction could be named by the participants. The openings of the Landolt rings, the basic shapes and the p's or q's could only be named at smaller distances to the fixation point. For counting the points or corners, the objects even had to be in the immediate vicinity of the fixation point.

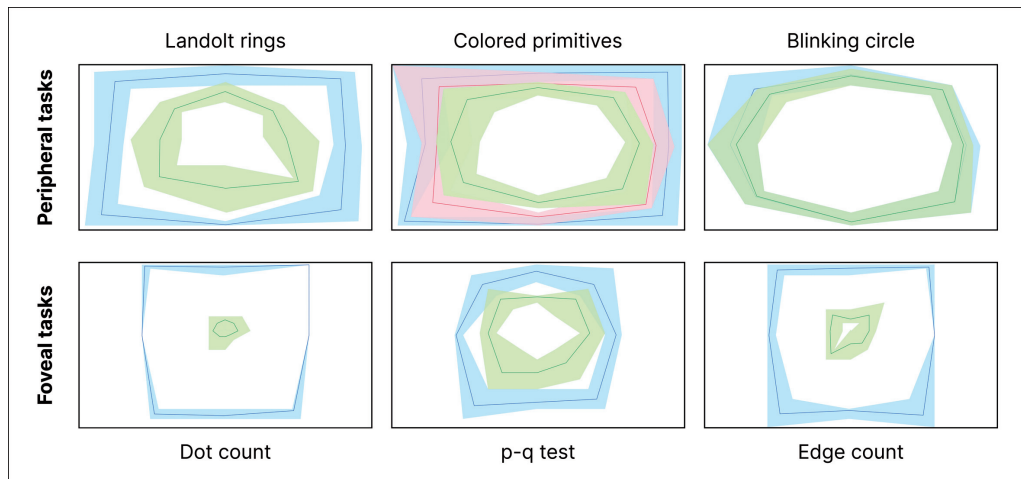


Figure 3.3: Visualization of the distances from which the objects were seen (blue), the color was correctly recognized (red; only colored primitives) or the details were correctly recognized (green). Darker lines represent average values.

On average, Landolt rings were seen from an angle of 50.4° to the fixation point, and the side of the opening was correctly detected from 28.8° . The colored primitives were seen on average from 53.3° , from about 50.2° their color was correctly recognized, and from 39.8° additionally their shape was correctly detected. The flashing circle was seen from 45° . The direction in which it was seen could be correctly named at 43.8° on average. The correct counting of points was only possible from an angle of 4.4° . The differentiation of rotated p's and q's could be done from 21.1° and the correct identification of the number of edges was possible on average at 15.7° .

Discussion

As expected, the more difficult foveal tasks could only be solved from a shorter distance to the fixation point. The Landolt ring task was also only solvable from an angle of 28.8° and could therefore also be grouped into the foveal tasks, especially since it is also quite close to the result of the p-q test with an angle of 21.1° . By far the most difficult task was counting the low-contrast dots. Apparently, the fixation of the circle was necessary to solve the task. The objects were already clearly seen before the associated task could be solved. The colored basic shapes were recognized earliest, closely followed by the Landolt rings. The flashing circles were seen last in the peripheral tasks, which is probably due to the fact that their contrast to the background is significantly lower than that of the col-

ored basic shapes or the Landolt rings. In the foveal tasks, the p's and q's were seen slightly later than the other objects, which is probably also due to their low contrast to the background as well as to the fine detail of the letters.

Hypothesis H1 was supported by the results, as different objects were detected in detail at different distances from the fixation point. The results also support our hypothesis H2 that objects are seen earlier than detail tasks can be solved. Likewise, our hypothesis H3 is supported, as the objects had to be significantly closer to the fixation point to solve more difficult tasks. The results obtained in this study fit with very early findings (see the paragraph *Perception of Colors and Visual Resolution in the Periphery* in Section 2.1.1) and serve as basic knowledge for our studies on gaze direction guidance.

Limitations

In the study, the test subjects were placed very close in front of a 50 inch monitor. It cannot be completely ruled out that with larger screens or projections where the participants have a greater distance to the display, slight deviations of the determined values can occur, since in this case more influences, e.g. by the surrounding lighting, have an effect on the test person.

In our study, colors were not distinguished as in other studies [78]. Therefore, no statement can be made about which color was recognized earlier.

Overall, we have only considered a small selection of possible objects and tasks and therefore cannot say in detail for each object from which viewing angle it will be recognizable. However, the results allow us to make a reasonable estimate of the distances from which the recognition of other objects, e.g., based on their color, shape, or texture, might be possible.

Conclusion

In this study, we investigated visual perception in the peripheral and foveal FOV. We investigated, for different objects entering the FOV from outside it, at what point these become visible and when we can begin to perceive exact details of the different objects. The results of the study show that the objects are seen very early, but the detection of object details is possible only at shorter distances. For very fine-detail objects (i.e., very difficult tasks), detection of the details becomes possible only when the objects are very close to the fixation point. The results are

in line with those of older studies (see the paragraph *Perception of Colors and Visual Resolution in the Periphery* in Section 2.1.1) and serve as a basis for the studies described in the following section.

3.2 Perception of Visual Stimuli for Guiding Gaze

In this section, we present several studies on gaze guidance using visual cue stimuli. In the first study, we examined the extent to which methods for subtle gaze direction guidance on computer screens are suitable for real-sized projected environments. We then tested various visual methods for gaze guidance in an instrumented environment and examined which methods are suitable for measuring gaze direction. Finally, we conducted studies on subtle gaze guidance in OST AR.

3.2.1 Subtle Visual Cues for Gaze Guidance at Projections

We first conducted a study to investigate to what extent the results from previous studies (see Section 2.2 and Section 3.1) can be applied to real-sized environments. In this study, particular attention was paid to the SGD method of Bailey et al. [6], which guides visual attention using subtle luminance changes in the target region. Instead of a PC screen, as in Bailey et al., our investigations took place in front of a real-sized projected shopping shelf. The aim was to investigate, on the basis of various tasks, to what extent it is possible to guide the direction of gaze by luminance adjustments in such a setup.

The implementation and execution of the study was done within the bachelor's thesis of Rutsch [130] based on a concept given by the author of this thesis, which was refined together. For the outcomes presented in this chapter, hypotheses were formulated, the recorded raw data was re-analyzed, and the new results were discussed.

Hypotheses

In our study, we wanted to show that even in a real-sized setup, it is possible to use brightness changes in the target region to direct a person's visual attention to a predefined object. We investigated both subtle brightness changes analogous to Bailey et al. and obvious brightness changes to guide gaze direction. We used

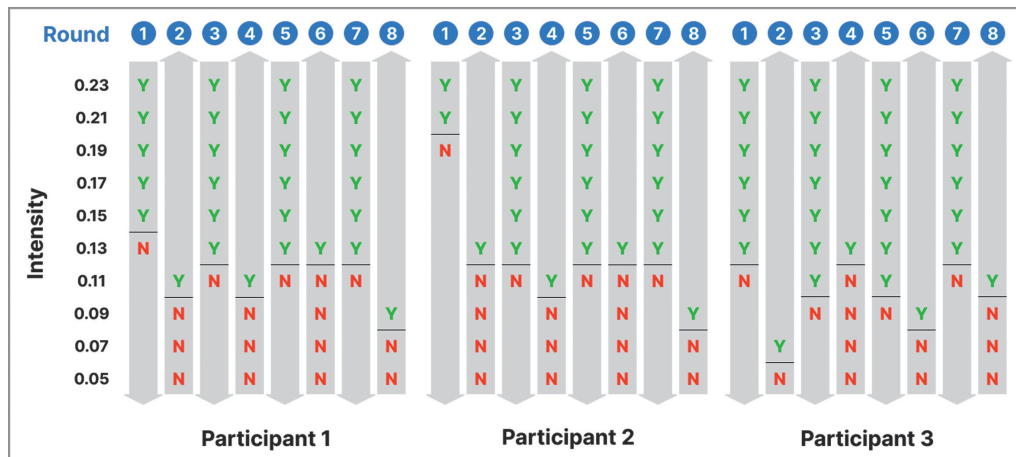


Figure 3.4: Presentation of the results of the method of limits for determining the cue intensity based on [42]. Y = cue has been recognized; N = cue has not been recognized. Adapted from [130].

visual search tasks [143] to compare the different methods. We assumed that the use of visual stimuli would affect performance in visual search tasks and that obvious visual cues would improve performance more than subtle visual cues. We therefore hypothesized that:

- H1: Visual stimuli influence the performance of visual search tasks in projected real-sized environments.
- H2: Obvious visual cues displayed at the target improve performance in visual search tasks more than subtle cues.

Pre-Study

First, we conducted a small preliminary study to determine the necessary intensity of the subtle cue stimuli. These should not be too strong: just strong enough to be perceived by the participants, but not obviously seen by them. To determine the absolute threshold, we used the method of limits, which provides a good average of accuracy and speed and therefore has a good effort-benefit ratio. In several runs, participants were presented with flashing dots with stepwise changes in brightness intensity, alternating in ascending or descending order. At each step, the participants reported whether or not they could see the stimulus. With descending intensity, a run stopped as soon as the stimulus could no longer be seen; with ascending intensity, a run stopped as soon as the stimulus was

detected. After performing all runs, the transition value for each participant was determined from all runs. This value indicates the threshold between perceptible and imperceptible stimulus for the respective participant. At the end, the mean value of all participants was determined, which was used as stimulus intensity in the main study.

A total of three participants took part in the described preliminary study. Figure 3.4 shows the results of the runs of the individual participants. The absolute threshold determined was a stimulus intensity of 0.11. In the main study, which is described in detail below, the visual stimuli were displayed at the determined stimulus intensity.

Study Tasks

We tested the above hypotheses in our main study using two different visual search tasks. The first task was to find a specified object within the real-sized environment. The objects to be searched were described either by a single property or by two properties, which required a conjunction search accordingly [42].

In the second task, the participants had to detect image changes that we had made. To do this, they were first shown the unchanged environment and shortly afterwards the changed environment.

In order to find out what influence, if any, the chosen methods for gaze guidance have in the different tasks, the visual stimuli were superimposed on the target objects. Both tasks were investigated with the support of obvious stimuli (condition *Obvious Cue*), with the support of subtle stimuli (condition *Subtle Cue*) and without any support (condition *No Cue*).

Participants

30 volunteers (22 male, 8 female) participated in our study. The participants were evenly distributed among one of the three test conditions. 10 participants (8 male, 2 female) had to complete the search task without assistance, another ten (7 male, 3 female) were shown obvious cues and the last ten (7 male, 3 female) were provided with subtle stimuli that should help fulfill the task. No participant reported any form of visual impairments.

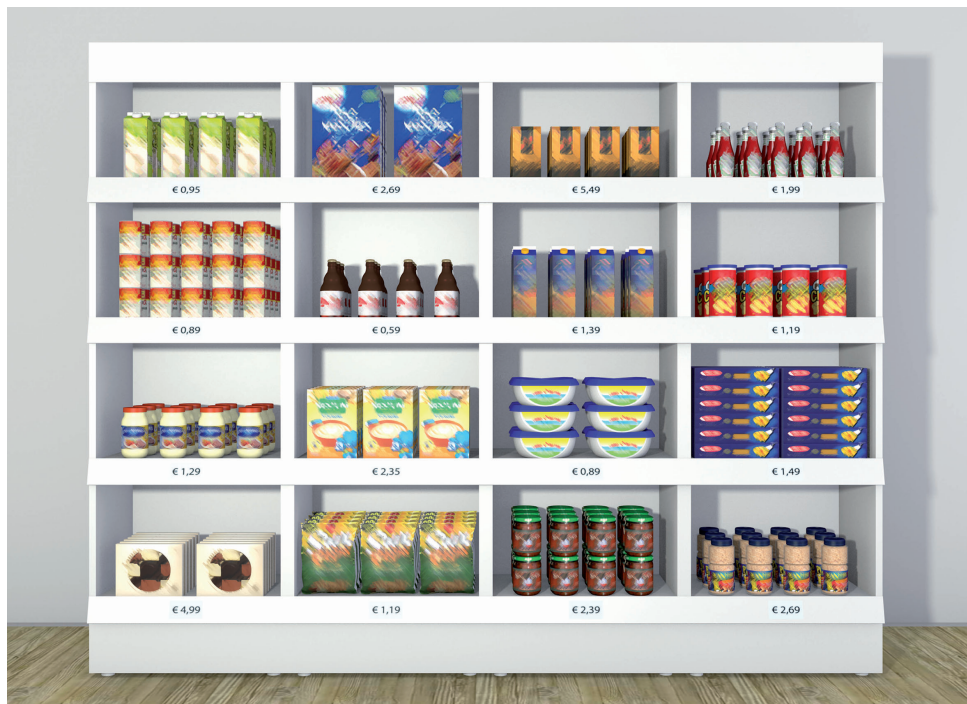


Figure 3.5: Example of one rendered shelf used in the study. The products have been made unrecognizable for publication. Adapted from [130].

Study Setup

In the study, the test subjects stood centrally at a distance of two meters in front of a projection screen on which images with a width of 271 cm (987 px) and a height of 213 cm (758 px) were projected, enabling the simulation of real-sized environments. The projection thereby covered approximately 34° of the horizontal FOV in each direction and approximately 38.1° of the vertical visual FOV in each direction. Depending on the study condition, visual cues were displayed on specified objects within the images as a support to direct participants' attention to the searched object. Eye data were recorded using a Pupil Pro eye tracker and the time to solve the task was measured. At the start, each participant underwent an eye tracker calibration and had to complete a short demographic questionnaire.

Design and Procedure

The study was designed as a between-subjects experiment, so every participant perceived only one of the three test conditions. We followed the method of Bailey

et al. [6] to display the subtle cue stimuli. The brightness changes occurred at a frequency of 10 Hz in the peripheral FOV of the participants. The subtle stimuli had a radius of 5 px in our setup. For the obvious stimuli, we used a stimulus size of 15 px so that they would be easily visible. Within this circular area, black and white color values were alternately blended under the image. Towards the edge, the change was reduced to avoid creating a sharp edge between the visual stimulus and the original image. This transition was implemented using a Gaussian function [165]. To prevent the subtle stimuli from being overtly perceived, it was ensured that they were only displayed when the target position was not in the subject's foveal FOV, corresponding to about 4° around the fixation point [139].

A shopping shelf with 16 product compartments was used as the simulated real environment (see Figure 3.5). The shelf images used were created using the 3D program SketchUp⁹. A total of 25 different products were created to ensure that participants were shown a newly constructed shelf each time to avoid a learning effect. Each shelf model contained one product that did not appear in any other shelf model. The other products were randomly sorted into the shelf. In total, ten different shelf constellations were created in this way. For the second part of the study, shelf images also had to be created in which a change was recognizable, such as a new lid color. A total of 20 different shelf images were therefore created. To make the shelf images appear as real as possible, a realistic shadow cast was created by a lighting source.

Part 1 The first part of the study consisted of finding a given product on the shelf image. A total of 10 different products had to be searched for one after the other on ten different virtual shelves. For each of the shelves, there was a predefined question that described the product being searched for either on the basis of one property or on the basis of two properties. Only the combination of the characteristics clearly defined the product to be searched for. A question with one characteristic could be, for example, "How much does the product with the green lid cost?" and a question with two characteristics could be, for example, "Which product costs €0.89 and has a blue label?" The respective shelf task combinations were presented to the participants in random order.

Before each subtask, the participants were shown a black picture with a white cross in the middle, which they had to fixate. Before the shelf image was dis-

⁹<https://www.sketchup.com> (last accessed: 2023-08-15)

played, the associated question was read out to the participants. Only when the participants confirmed that they had understood and internalized the question the corresponding shelf image was displayed and the time measurement started. The participants now had to find the matching product as quickly as possible. As soon as they thought they had found the product, they could end the task by saying "STOP". At this moment, the timing also ended. They then told the study leader the answer to the question and it was noted whether the answer was correct or incorrect. In order to keep the length of the study controllable, a subtask was automatically ended after 10 seconds by fading out the corresponding shelf image. However, the participants still had the opportunity to answer the question afterwards.

Part 2 The second part of the study involved finding a changed area on the shelf. Changes were, for example, that a product was rotated or missing or that the font color or the price tag were different.

Ten different subtasks were presented to the participants one after the other in random order. For each subtask, the unchanged shelf was displayed for 20 seconds at the beginning. Then a black intermediate image with a white cross in the middle appeared, which the participants had to fixate. This was displayed for 2 seconds and was intended to ensure that participants did not perceive the change directly, as this exploited change blindness [125, 126]. The participant was then shown the corresponding changed shelf image. This image was displayed for a maximum of 20 seconds, during which the participants had the opportunity to detect the image change. If the change was detected earlier, the participants could end the task and thus also the timing with the word "STOP". This also faded out the shelf image. The participant then verbally communicated his answer and it was noted whether it was correct or incorrect.

Results

We investigated the effects of peripheral cues for gaze direction guidance on the measured dependent variables *task completion time* and *correctly solved tasks* reported below with a uniform procedure. For evaluation of the effect of the different peripheral cues on the time to solve the task, we first applied a Kruskal-Wallis test with two degrees of freedom and a significance level of $\alpha = 0.05$ to compare the samples collected in the three test conditions with each other, separately for part 1 and part 2 of the study. We report the *p*-value along with

	No Cue	Subtle Cue	Obvious Cue
Part 1 (all)			
Sample Size	100	100	100
Mean	6000.2	6013.8	5318.6
Standard Deviation	2265.5	2217.1	2258.5
Part 2 (all)			
Sample Size	100	100	100
Mean	15284.4	14162.8	8517.1
Standard Deviation	5811.3	5744.6	6622.4
Part 1 (correct answers)			
Sample Size	76	76	89
Mean	5733.5	5752.7	5393.4
Standard Deviation	2064.7	1987.7	2166.3
Part 2 (correct answers)			
Sample Size	18	32	73
Mean	7555.8	10624.1	6378.5
Standard Deviation	3647.3	4982.8	5088.4

Table 3.1: Measured task completion times in part 1 and part 2 of the study. Presentation of sample sizes, means, and standard deviations in the three test conditions for all records and only for the records with correct answers.

the test statistic χ^2 . If a significant influence of the condition on the dependent variable was detected, we used Mann-Whitney U tests ($\alpha = 0.05$) to perform pairwise comparisons among all test conditions.

Task Completion Time In evaluating the effect of test condition on task completion time, we examined all records as well as only those records in which participants successfully completed the visual search task, i.e., provided the correct answer in the specified time, resulting in different sample sizes in the conditions. Table 3.1 shows the sample sizes of the different conditions as well as the corresponding means and standard deviations. In part 2 of the study, the sample size was especially small in the condition *No Cue* ($N = 18$) and in the condition *Subtle Cue* ($N = 32$) when only records with correct answers were taken into account.

When all records were considered, the Kruskal-Wallis test detected significant effects of the visual cue used on the task completion time in both part 1 ($\chi^2 = 6.446, p = 0.040$) and part 2 ($\chi^2 = 62.038, p < 0.001$) of the study. The Mann-Whitney U tests showed for part 1 of the study that condition *No Cue* differs

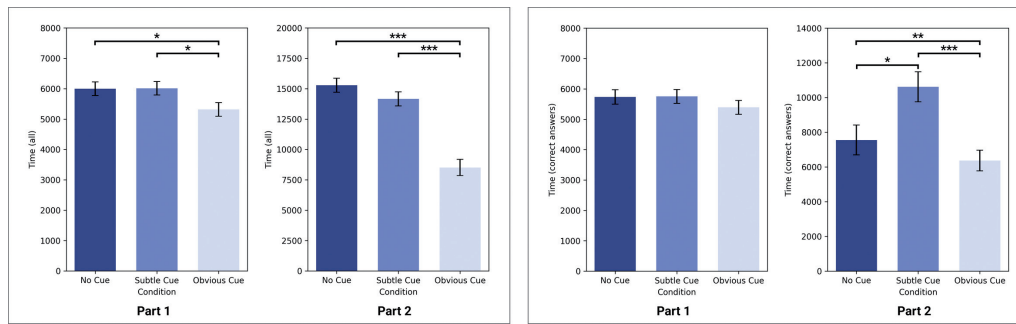


Figure 3.6: Task completion times for all records (left) and only for records with correct answers (right) in study part 1 and study part 2. Significant differences marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

significantly from condition *Obvious Cue* ($U = 5863.5, p = 0.035$). Furthermore, the condition *Obvious Cue* differs significantly from condition *Subtle Cue* ($U = 4069.0, p = 0.023$) in part 1 when all records are analyzed. For part 2 of the study, the Mann-Whitney U tests revealed that the condition *No Cue* differs significantly from condition *Obvious Cue* ($U = 7844.0, p < 0.001$). Significant differences were also detected between condition *Obvious Cue* and condition *Subtle Cue* ($U = 2471.0, p < 0.001$). Figure 3.6 (left) visualizes the results.

Considering only the records with correct answers, the Kruskal-Wallis test confirmed that there is an effect of the visual cue used on the time to solve tasks in part 2 ($\chi^2 = 20.938, p < 0.001$) of the study. For part 1 of the study no significant effect on the task completion time was detected ($\chi^2 = 2.034, p = 0.362$). The Mann-Whitney U tests revealed that in part 2 of the study, the condition *Subtle Cue* differs significantly from the condition *No Cue* ($U = 195.0, p = 0.045$) in terms of task completion time. A significant difference was also detected between condition *No Cue* and condition *Obvious Cue* ($U = 887.0, p = 0.022$) and between condition *Subtle Cue* and condition *Obvious Cue* ($U = 578.0, p < 0.001$) (see Figure 3.6, right).

Correctly Solved Tasks Table 3.2 shows the means and standard deviations regarding the correctly solved tasks for the different conditions. The Kruskal-Wallis test revealed that the test condition has a significant effect on the number of correctly solved tasks both in part 1 ($\chi^2 = 7.605, p = 0.022$) and part 2 ($\chi^2 = 14.611, p < 0.001$) of the study.

The Mann-Whitney U tests for part 1 of the study showed no significant difference of condition *Subtle Cue* vs. condition *No Cue* ($U = 48.500, p = 0.937$) in terms of

	No Cue	Subtle Cue	Obvious Cue
Part 1			
Sample Size	100	100	100
Mean	7.6	7.6	8.9
Standard Deviation	1.27	0.97	0.99
Part 2			
Sample Size	100	100	100
Mean	1.8	3.2	7.3
Standard Deviation	2.39	1.81	2.75

Table 3.2: Correctly solved tasks in part 1 and part 2 of the study. Presentation of sample sizes, means, and standard deviations in the three test conditions.

correctly solved tasks, while a significant difference was detected for condition *No Cue* vs. condition *Obvious Cue* ($U = 21.000, p = 0.027$) and for condition *Subtle Cue* vs. condition *Obvious Cue* ($U = 82.0, p = 0.014$). Also in part 2 of the study, the condition *Subtle Cue* did not differ significantly from the condition *No Cue* ($U = 27.0, p = 0.084$) regarding correctly solved tasks. A significant difference was only detected for condition *No Cue* vs. condition *Obvious Cue* ($U = 8.0, p = 0.002$) and for condition *Subtle Cue* vs. condition *Obvious Cue* ($U = 89.5, p = 0.003$) (see Figure 3.7).

Discussion

We hypothesized that visual cues have an influence on the performance of visual search tasks in real-sized projected environments (H1). Hypothesis H1 was supported by the results. Especially the condition with obvious cues performed significantly better than the condition without cues in almost all tests. Only the investigations on task completion times, in which records were taken into account only if correct answers were given, did not yield any significant differences. This can possibly be attributed to the small sample sizes in this investigation. The results also show a tendency that the subtle cue stimuli were able to draw visual attention to the target region. This is especially true for part 2 of the study. However, significant improvements compared to the condition without cues could not be found. We suspect that the subtle cue stimuli did not lead to significant improvements because some of the participants may not have been able to perceive the cue stimuli at all. In the preliminary study, we determined the intensity value at which the cues were recognized from a few participants

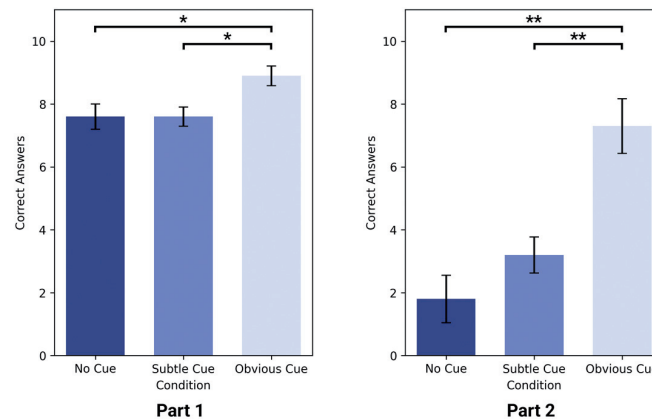


Figure 3.7: Correctly solved tasks in study part 1 and study part 2. Significant differences marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

and used the calculated mean value for the main study. It would probably have made more sense to use a value for the main study that ensured that the stimuli could actually be perceived by everyone. Ideally, for assistance systems that use subtle visual cues, the threshold should be determined individually for the respective user. In this study, however, this was not possible because the participants would have already known that visual stimuli were displayed in the environment and might have explicitly looked out for them, which would have significantly influenced the results.

Furthermore, we hypothesized that overt visual cues would improve performance more than subtle cues (H2). Our results support hypothesis H2. In all investigations, except for the investigations on task completion times in which only records with correct answers were considered, the condition with obvious cues performed significantly better than the condition with subtle cues. It was to be expected that the obvious cues would perform better than the subtle cues, since obvious cues are perceived faster and can therefore be responded to more quickly. Some participants may not have perceived the subtle cues at all, as described above, which is why no significant differences could possibly be determined in this test condition.

The results of the test conditions in the two task parts (see Figures 3.6 and 3.7) suggest that the effect of visual cue stimuli differs depending on the task. It can be seen that the effect is considerably larger in the second part of the study than in the first part, which can be explained by the difficulty of the different parts of the study. The first part of the study was relatively easy and quick to solve even

without cue stimuli. The differences that had to be found in the second part of the study, however, were mostly very difficult to recognize, so that assistance in the search was probably a great support here.

Regarding the task completion time in the second part of the study, the results are quite unusual. Here we expected the task completion time to decrease with increasing visibility of the stimuli. These unusual results for the conditions without and with subtle cues can probably be explained by the small sample size. Many participants were not able to give the correct answer in these conditions, which is why only a few measurements were included in the calculations (see Tables 3.1 and 3.2). In addition, when supported with subtle stimuli, compared to when no cue stimuli were shown, many more correct responses were given, but they were given late. It is possible that it took participants a little longer to respond to these stimuli, whereas without help from cue stimuli, participants may have found most objects quickly by chance.

Limitations

We investigated the extent to which visual cues displayed in the user's peripheral FOV can guide gaze to predefined target locations. The results can only give an initial indication of how well the different visual cue stimuli assist. How strong the effect of the individual visual stimuli actually is could only be determined in a much larger study.

In addition, it can be assumed that due to the choice of an average stimulus intensity, some participants could not perceive the stimuli and that the gaze guidance could therefore not work at all for these participants. Unfortunately, this cannot be determined afterwards, since no query was made as to whether the participants perceived anything unusual. In future studies it should be ensured that the participants can perceive the stimuli, or it should at least be recorded whether the perception of the stimulus worked or not.

Conclusion

In this study, we investigated the extent to which it is possible to direct people's gaze to fixed objects using subtle and obvious visual cues. For this purpose, we generated two different visual search tasks and, depending on the test condition, blended in subtle or obvious visual cues on the objects to be searched when they were in the participant's peripheral FOV. In addition to the groups of participants

that were shown subtle or obvious cues, a third group performed the task without any assistance.

The results show that the use of obvious cues yields significantly better results in terms of correctly solved visual search tasks than subtle cues, or when the task is solved without assistance. Since a correct answer to the questions was only possible by looking at the searched-for object, it can be concluded that the direction of gaze can be guided with the help of obvious visual cues.

The results for assistance with subtle visual cues do not show any improvement compared to the condition without cues. It is assumed that this may be due to the chosen study setup and that with a larger number of participants and a more wisely chosen stimulus intensity, clearer results could have been obtained.

3.2.2 Gaze Guidance in Instrumented Environments

Our next studies took place in a real instrumented environment. We implemented it based on the framework presented in Section 5.1. For this, we built a shopping shelf that could trigger cues at individual product compartments using actuators (see Section 5.1.2). During the studies, the participants' gaze direction was measured and taken into account accordingly. In the first study, we compared different cues and different sensors for gaze direction detection. In the second study, we investigated the extent to which adaptive cue stimuli can be used to guide gaze.

Study 1

In the first study, the goal was to find out how well different cue stimuli are suited for directing attention. In addition, the objective was to determine whether precise eye tracking achieves better results in directing attention than if only a coarse gaze direction is used as the basis for controlling attention.

Hypotheses The goal of the study was to show that the visual cues influence the speed of solving visual search tasks. It was also meant to show that the visual cue stimuli result in different evaluations. In our study, we investigated three different cue stimuli that were intended to direct the participants' visual attention to searched-for products. We assumed that more obvious cue stimuli would lead to faster response times. It was not expected that there would be

an influence on the number of correct answers, since the search tasks were easy enough to be solved without the cue stimuli. The study also aimed to find out whether the choice of the sensor for gaze direction detection has an influence on the study results. Here, we did not expect a strong influence, but that even a rough gaze direction would be sufficient to successfully direct attention to a predefined object. We therefore hypothesized that:

H1: The use of different stimuli has an effect on task completion time in visual search tasks.

H2: Performance is noticeably better for more obvious visual cues than for less obvious ones.

H3: Different visual stimuli lead to different evaluations regarding *Assistance*, *Disturbance*, *Favor* and *Usability*

H4: The use of different stimuli does not affect the correct solving of simple search tasks.

H5: Different sensors used for gaze direction detection do not affect the performance.

Study Task To test the above hypotheses, we used visual search tasks in which participants had to search for products in the instrumented shopping shelf. To determine how well different visual cues performed in the instrumented environment, the search tasks were processed with the assistance of different visual cues. The participants were given tasks that could only be solved by looking at the respective product. We compared light-dark flashing price tags (condition *Blinking*), illuminated product compartments (condition *Lighting*) as well as blurred price tag displays for products not being searched for (condition *Blurring*). The different cue stimuli were tested using both the Pupil Pro eye tracker as a sensor for gaze direction detection (condition *Eye Tracking*) and the OptiTrack motion capturing system¹⁰ (condition *Motion Capturing*).

Participants 12 student volunteers participated in our study. All had normal or corrected-to-normal vision, which we checked by a Landolt ring and a color vision test at the beginning of each study session.

¹⁰<https://optitrack.com> (last accessed: 2023-08-15)

Study Setup During the study, participants were centered 1 meter in front of the instrumented shopping shelf, which had a width of 240 cm and a height of 192 cm (see Figure 5.4). This covered approximately 50.2° of the horizontal FOV in each direction. Different cues were displayed at individual product compartments. The direction of gaze was determined using both a Pupil Pro eye tracker and the OptiTrack motion capturing system. The times needed to solve each task, and whether the tasks were solved correctly, were recorded. The participants had to fill out a questionnaire after each performed condition, with which they evaluated the respective condition. At the end of the study, the participants also had to fill out a final questionnaire in which they were asked to rank the conditions.

Design and Procedure The study was designed as a within-subjects experiment. Half of the participants started with the eye tracker as sensor for gaze detection; the other half started with the motion capturing system. Participants had to perform search tasks with both sensors with the aid of the three cue stimuli. The order in which the different cue stimuli were displayed was counterbalanced by a Williams design Latin square of size three [161]. Participants were asked five questions per cue stimulus. In total, a participant thus had to complete 30 different search tasks during the study. While the task was read out, the participants looked at a red dot in the middle of the shelf. Only when the question was finished the participants were allowed to start searching. At this point, the cues were also displayed and the timing started. The participants now had to search for the answer as quickly as possible. As soon as they had found the answer, they pressed a button to stop the time. They then told the experimenter their solution and it was noted whether it was correct or incorrect.

Results We investigated the effects of the visual cues for gaze guidance as well as the effects of the sensor types on the measured dependent variables *task completion time* and *correctly solved tasks* reported below with a uniform procedure. For evaluation of the effect of the different visual cues, we first applied a Kruskal-Wallis test with two degrees of freedom and a significance level of $\alpha = 0.05$. When determining the effect of the cue stimuli, we evaluated both the overall effect, and the effect separately for the two sensor types. We report the p -value along with the test statistic χ^2 . If a significant influence of the condition on the dependent variable was detected, we used Mann-Whitney U tests ($\alpha = 0.05$) to compare all test conditions with each other.

	Blinking	Lighting	Blurring
Overall			
Sample Size	114	117	107
Mean	6036.5	6260.4	7165.8
Standard Deviation	2847.7	4498.2	4478.5
Eye Tracking			
Sample Size	56	58	53
Mean	6097.4	5857.3	7142.2
Standard Deviation	2851.6	3092.4	5273.3
Motion Capturing			
Sample Size	58	59	54
Mean	5978.7	6656.7	7188.6
Standard Deviation	2867.8	5545.0	3600.5

Table 3.3: Measured task processing times overall and separately for the sensor types *Eye Tracking* and *Motion Capturing*. Sample sizes, means and standard deviations are shown for the three visual cue conditions.

When determining the influence of the visual cues used on the task completion time, we only considered the times of the tasks in which the participants provided a correct answer. This results in different sample sizes in the individual conditions. Table 3.3 shows the sample sizes of the different visual cue conditions. During the study, 120 questions were asked per cue condition. In total, only 21 of the 360 questions asked were answered incorrectly.

The Kruskal-Wallis test showed that overall there was a significant effect of the visual cue used on task completion time in the study ($\chi^2 = 8.7736, p = 0.012$).

The subsequent Mann-Whitney U tests showed that the condition *Blinking* differs significantly from the condition *Blurring* in terms of task completion time ($U = 5038.000, p = 0.025$). A significant difference was also detected for condition *Lighting* vs. condition *Blurring* ($U = 4956.500, p = 0.005$). The condition *Blurring* provided significantly worse task completion times than the conditions *Blinking* and *Lighting* (see Figure 3.8).

When investigating the influence of the chosen sensor on the task completion time, the Kruskal-Wallis test did not show any significant difference between the two sensors used. Table 3.4 shows the sample sizes, means, and standard deviations of the modes *Eye Tracking* and *Motion Capturing*.

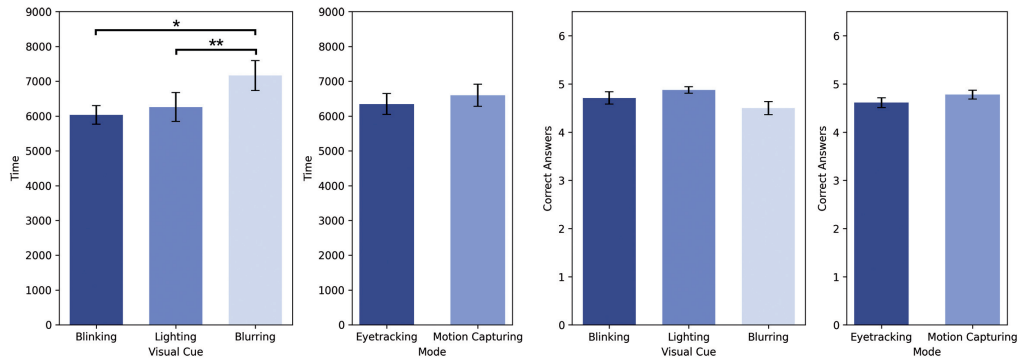


Figure 3.8: Task completion times and correctly solved tasks grouped by visual cue stimuli and by sensor mode. Significant differences marked with * ($p < 0.05$) and ** ($p < 0.01$).

	Eye Tracking	Motion Capturing
Sample Size	167	171
Mean	6347.1	6598.2
Standard Deviation	3877.7	4184.2

Table 3.4: Measured task completion times. Sample sizes, means, and standard deviations for the two sensor types *Eye Tracking* and *Motion Capturing*.

Kruskal-Wallis tests were also used to determine the effect of visual cues and the effect of the selected sensors on the number of correct responses. In our study, neither a correlation between visual cue and correctly solved tasks nor between sensor type and correctly solved tasks could be determined.

The evaluation of the individual conditions was based on the questionnaire that had to be filled out by the participants after the completion of each condition. This questionnaire was used to assess the extent to which the presented cue was helpful in locating the searched-for product (*Assistance*), how much the cue disturbed the participants (*Disturbance*), how much the participants liked the cue (*Favor*) and how well they could imagine using this cue in everyday life (*Usability*). The evaluation was based on 7-point Likert scales. Figure 3.9 visualizes the results of the questionnaire.

We examined the effect of the selected cue stimulus on the four items described above. For this, we first performed a Friedman test with 2 degrees of freedom and a significance level of $\alpha = 0.05$ for each of the questions. We report the p value along with the test statistic χ^2 . If a significant effect of visual cues on responses to the question was found, we used Wilcoxon's signed-rank tests

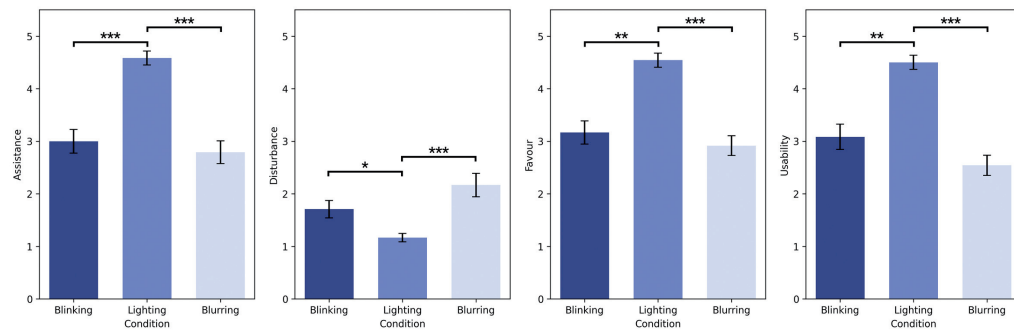


Figure 3.9: Ratings regarding helpful assistance, disturbance, favor and everyday usability of the visual cues. Significant differences marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

($dof = 11$, $\alpha = 0.05$) to compare all conditions. We report Bonferroni-corrected p values, test statistic W , and pairwise rank-biserial correlation r as effect size.

Assistance: The Friedman test showed that the visual cue significantly influences the assistance rating ($\chi^2 = 21.432$, $p < 0.001$). Wilcoxon's signed-rank test revealed that condition **Lighting** has significantly higher assistance ratings than condition **Blinking** ($W = 4$, $p < 0.0001$, $r = 0.962$) and condition **Blurring** ($W = 5$, $p < 0.001$, $r = 0.952$).

Disturbance: The Friedman test showed that the visual cue also significantly influences the disturbance ($\chi^2 = 15.672$, $p < 0.001$). Wilcoxon's signed-rank test revealed significantly lower disturbance ratings for condition **Lighting** than for condition **Blinking** ($W = 10$, $p = 0.032$, $r = -0.780$) and condition **Blurring** ($W = 0$, $p < 0.001$, $r = -1$).

Favor: The Friedman test revealed a significant influence of the visual cue on favor ($\chi^2 = 20.354$, $p < 0.001$). Wilcoxon's signed-rank test showed that condition **Lighting** has significantly higher favor ratings than condition **Blinking** ($W = 3.5$, $p = 0.002$, $r = 0.954$) and condition **Blurring** ($W = 9$, $p < 0.001$, $r = 0.922$).

Usability: The Friedman test also showed that the visual cue significantly influences the usability ($\chi^2 = 24.105$, $p < 0.001$). The results of Wilcoxon's signed-rank test demonstrated that condition **Lighting** has significantly higher usability ratings than condition **Blinking** ($W = 0$, $p = 0.001$, $r = 1$) and condition **Blurring** ($W = 5$, $p < 0.001$, $r = 0.960$).

In the final questionnaire of the study, the individual conditions were ranked by the participants. The best rating was given to the condition **Lighting**, which

was rated as the best by 9 of the 12 participants. The worst rated condition was *Blurring*, which was rated last by 8 participants.

Discussion We hypothesized that the use of different cue stimuli in visual search tasks has an influence on task completion time (H1). Our results support hypothesis H1. Depending on the visual cue used, there are significant differences in task completion time.

Furthermore, we hypothesized that performance would be significantly better for more obvious cue stimuli than for less obvious cue stimuli (H2). In our study, we showed that the less obvious cue stimuli of condition *Blurring*, in which the price tags of the products not searched for were blurred, performed significantly worse than the two other visual cue stimuli used, in which the shelf was brightly lit (*Lighting*) or the associated price tag was flashing (*Blinking*). The two conditions *Lighting* and *Blinking* did not differ much from each other, which is probably explained by the fact that they were both easy to recognize and fast to interpret.

Additionally, we hypothesized that the conditions would be evaluated differently (H3). Our results support hypothesis H3. Condition *Lighting* was rated significantly better than the other conditions on all four points investigated (Assistance, Disturbance, Favor, and Usability) although temporal performance was not better than for condition *Blinking*. This could be due to the fact that the cue in condition *Lighting* was clearly visible, but still not distracting because it did not flash permanently.

We hypothesized that the cue stimulus used does not affect the correct solution of simple search tasks (H4). Hypothesis H4 was also shown by our results. The number of correct responses did not differ significantly for the visual cue stimuli used. The tasks were simple tasks that could be solved after a certain time even without a cue stimulus. Since there was no time limit in the study conducted, it could be assumed that the results would be similar for all three conditions.

Finally, we hypothesized that the sensor used to detect gaze direction would not have a strong influence on the performance of visual search tasks assisted by visual cue stimuli (H5). Hypothesis H5 was also supported by our results. Both task completion time and the number of correct responses were almost identical in the *Eye Tracking* and *Motion Capturing* modes. This shows that even a system that can only roughly determine the user's gaze direction can be used as a sensor in a framework for gaze direction control.

In the final rating of the cue stimuli used, *Lighting* was evaluated as the best cue stimulus. This matches the results of the evaluated questionnaires, in which Lighting scored the best in all examined points.

Limitations In this study, we investigated the extent to which visual attention can be guided by different visual cue stimuli using different methods of gaze direction detection and the extent to which different cue stimuli lead to different outcomes. In our study, we only considered the mode of action of three different stimuli and were thus only able to make a comparison between these three stimuli. A study with a larger number of stimuli used, in which various subtle stimuli are also considered, would provide even more detailed results.

Another limitation of the study is that the conditions were not examined in relation to a baseline. Thus, only a difference between the three visual cue stimuli used can be determined, but not to what extent these provide better results than if the same search task had been performed without visual cue stimuli.

In the study, only two different sensors were investigated to determine gaze direction. There are other technical possibilities, e.g. to determine the head orientation, which are considerably less precise than a professional motion capturing system like the one we used in our study. Here it would be interesting to investigate to what extent the results change with even less precise gaze direction.

Conclusion The aim of the study conducted was to determine how well different visual cue stimuli are suited for directing visual attention. In addition, it was investigated to what extent the results are influenced by different mechanisms for determining the direction of gaze. In particular, it should be examined whether an exact gaze direction is required or whether it is also possible to work with an approximate gaze direction in order to direct a user's attention in a targeted manner. To investigate this, we had the participants perform various visual search tasks. They received help in the form of three different visual cues that were supposed to direct their gaze to the searched-for object. The determination of the current gaze position was either done precisely by eye tracking, or a rough gaze direction was obtained by head tracking.

The results show that different visual cue stimuli are evaluated differently. Likewise, the visual cues influence the task completion time to different extents. Thus, depending on the cue used, attention can be directed to predefined objects with different degrees of success. In this context, highly visible cue stimuli perform

better than less obvious cue stimuli. In our study, the condition *Lighting*, in which the entire product compartment was brightly illuminated, performed best. The results of directing the gaze with the help of the eye tracker hardly differ from those with the motion capturing system. Hence, no exact eye position is necessary to direct the attention to objects; the alignment of the head is already sufficient for this.

Study 2

In the second study, we wanted to find out to what extent a stimulus intensity adaptively adjusted to the user (in terms of contrast and frequency) influences task completion time. In addition, we wanted to investigate to what extent such a form of gaze direction guidance is evaluated in comparison to static cue stimuli with fixed blink frequency and contrast.

Hypotheses The aim of the study conducted was to show that adaptive cue stimuli direct attention at least as well as visual cues with a fixed stimulus intensity. Furthermore, we wanted to show that adaptive stimuli are not perceived worse than non-adaptive ones. For this purpose, we had the study participants perform visual search tasks with the help of three different conditions. In one condition, the participants received no cue at all and had to solve the task on their own. In another condition, cues were displayed on the price tags with a fixed contrast and frequency. In the third condition, the stimulus intensity was adjusted depending on the user's reaction and the distance of the line of sight to the target object. Here, the stimulus intensity was increased as long as the participant did not respond to the stimulus or, if necessary, the stimulus was even changed if there was no response. In addition, the closer the gaze moved towards the target object, the more the stimulus was weakened.

We assumed that the conditions in which participants were supported with visual stimuli would be rated better than the condition without assistance and would also have a significantly better task completion time. Furthermore, we assumed that the task completion time with the help of the adaptive subtle stimuli is about as good as that with the help of the static stimuli. This is because, although the stimulus is intensified when it is not reacted to, at the same time, the stimulus is reduced the closer one's gaze approaches the target object. Because of this adaptation, the stimulus intensity was always only as high as necessary, so we

expected a better evaluation of the condition with adaptive stimuli compared to the condition with static stimuli. We therefore hypothesized the following:

- H1: The task completion time for visual search tasks is significantly improved by the use of visual cues.
- H2: The conditions that support the search with visual cues are rated significantly better than the condition without assistance.
- H3: Task completion time is not worse for visual stimuli whose intensity is adaptively adjusted to the user than when a static stimulus intensity is used.
- H4: Adaptive visual stimuli for assistance are evaluated better than static visual cue stimuli.

Study Task To test the above hypotheses, we again used visual search tasks in this study, in which participants had to find previously specified products on the instrumented shopping shelf we built (see Section 5.1.2). Flashing price tags were used as visual cues.

During the study, participants had to solve tasks with adaptive cues (condition *Adaptive Cue*), tasks with fixed cues (condition *Static Cue*), and tasks without the help of cues (condition *No Cue*).

Participants 6 student volunteers participated in our study. All had normal or corrected-to-normal vision, which we checked at the beginning of each study session by a Landolt ring and a color vision test.

Study Setup The study setup was the same as in the first study. Participants were centered 1 meter in front of the instrumented shopping shelf. With a width of 240 cm and a height of 192 cm, this covered a horizontal FOV of about 50.2° in each direction.

Depending on the condition, a visual cue stimulus was displayed on the price tag of the searched product. The Pupil Pro eye tracker was used to determine the gaze direction needed to adjust the stimulus intensity. The time the participants needed to complete the tasks was recorded. In addition, after each completed condition, participants had to rate the condition using a NASA-TLX questionnaire [48].

Design and Procedure The study was designed as a within-subjects experiment. Participants had to solve five visual search tasks in each of the three different conditions (*No Cue*, *Static Cue*, *Adaptive Cue*). The order in which the conditions were presented to the participants was counterbalanced by a Williams design Latin square of size three [161]. Participants were asked five questions per cue stimulus. In total, a participant thus had to complete 15 different search tasks during the study.

The participants looked at a red dot in the middle of the shelf until the task was read out to them. After that, the cue stimulus was displayed and they were allowed to start searching. At the same time, the time measurement started. The participants now had to find the searched-for product as quickly as possible.

While in the condition *No Cue* no assistance was provided, in the other two conditions flashing visual cues were displayed. In condition *Static Cue*, black and white flashing price tags were used as the cue stimulus with a fixed flashing frequency and contrast. In the condition *Adaptive Cue*, the intensity, i.e. the frequency and contrast, was determined based on the distance of the current viewpoint to the target object. The smaller the distance to the target object was, the smaller the blinking frequency and the lower the contrast became. The further the viewpoint moved away from the target object, the higher the blink frequency and also the contrast became. When the distance and thus the stimulus intensity was maximal for at least 500 ms, the stimulus was changed. In total, there were two different visual stimuli that were adaptively adjusted: black and white flashing of the price tag and red and white flashing of the price tag. Half of all the search tasks performed started with the black-and-white flashing price tags, while the other half started with the red-and-white flashing price tags. The allocation to the individual tasks was random.

As soon as the participants had found the answer to the question posed, they pressed a button and the time was stopped. Once they had completed a condition, they rated it using the NASA-TLX questionnaire.

Results We investigated the influence of the visual cue used on task completion time as well as NASA-TLX ratings by first performing a Friedman test with 2 degrees of freedom and a significance level of $\alpha = 0.05$. We report the p -value along with the test statistic χ^2 . If a significant influence of the condition on the dependent variable was detected, we used Wilcoxon's signed-rank tests ($\alpha = 0.05$) to compare all test conditions with each other.

	No Cue	Static Cue	Adaptive Cue
Mean	6075.3	2476.5	2967.5
Standard Deviation	2946.7	865.1	2367.7

Table 3.5: Means and standard deviations of the task completion times in the three test conditions.

The Friedman test showed that overall there was a significant effect of the visual cue used on task completion time in the study ($\chi^2 = 34.067, p < 0.001$). The subsequent Wilcoxon signed-rank tests showed that the condition *No Cue* significantly differed from the conditions *Static Cue* ($W = 1, p < 0.001$) and *Adaptive Cue* ($W = 29, p < 0.001$) with respect to task completion time. No significant difference in task completion time was found for condition *Static Cue* vs. condition *Adaptive Cue* in this study. Table 3.5 provides an overview of the means and standard deviations with respect to the task completion time. In the condition *Static Cue* the task completion time was better than in the condition *Adaptive Cue*, in which there were also high deviations.

The result of the evaluation of the NASA-TLX questionnaires can be seen in Figure 3.10. The Friedman tests performed first provided a very weak significance only for *Effort* ($\chi^2 = 6.0, p = 0.0498$). The subsequent Wilcoxon signed-rank tests, however, could not detect any significant differences between the conditions with respect to *Effort*.

Overall, it can be seen that condition *No Cue* performed worse in every evaluation than the other two conditions in which participants were supported in their search by visual cues. The condition *Adaptive Cue* performed worse than the condition *Static Cue* in many cases, but the scores were very close.

Discussion We hypothesized that the use of visual cues in visual search tasks would significantly improve task completion time (H1). Our results confirm hypothesis H1.

In addition, we hypothesized that the conditions with visual cue stimuli would be rated better by participants than the condition without cue stimuli (H2). We could not prove hypothesis H2. However, the analysis of the NASA-TLX questionnaires shows a clear trend in favor of the conditions with visual cue stimuli. Nevertheless, no significant difference could be found in our study, which is probably due to the very small number of participants.

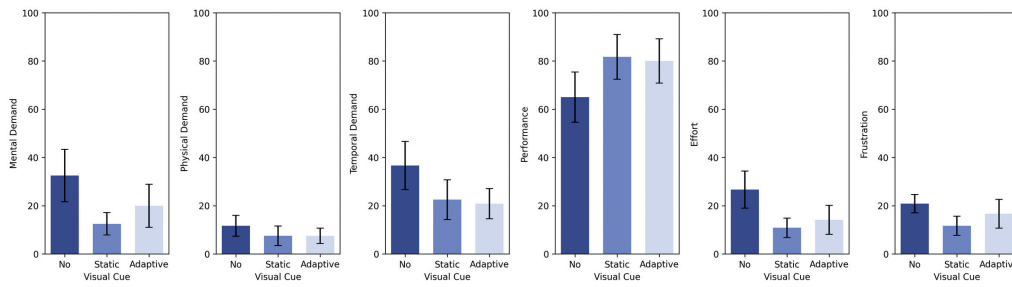


Figure 3.10: NASA-TLX ratings of the tasks without visual cues, with static visual cues and with adaptive visual cues. No significant differences were detected.

We hypothesized that task completion time would not be worse when using adaptive cues than when using static cues (H3). In the study we conducted, the adaptive cue stimuli did not perform significantly worse in terms of task completion time, but they performed slightly worse than the static cue stimuli. It can be seen that there is a high standard deviation in the adaptive cue stimuli, which may lead to the temporal differences in the conditions. In order to determine to what extent static and adaptive visual cues differ with respect to task completion time, a further study with a larger number of participants needs to be conducted.

Additionally, we hypothesized that adaptive cue stimuli would be evaluated better than static cue stimuli (H4). The results of the evaluation of the NASA-TLX questionnaires reject hypothesis H4. No significant difference was found between the condition *Static Cue* and the condition *Adaptive Cue*. The evaluation showed about the same results. It can be assumed that the participants may not have noticed the difference between the conditions or hardly noticed it at all, since the visual stimuli differed only slightly from each other and the adaptive adjustment of the stimuli was ideally not noticed.

The adaptive stimuli were not rated worse than the static stimuli in the study conducted, but are expected to offer some advantages compared to the latter. Since the stimulus is individually adapted to the respective person and is only set as intense as necessary, the user can be guided to the target much more subtly. Since we did not ask how annoying the stimulus was, or to what extent it was perceived at all, we cannot make any precise statements about this. This would have to be investigated in detail in further studies.

Limitations In this study, we investigated the extent to which cue stimuli that adaptively adjust to the user are suitable for guiding gaze direction and

compared them for this purpose to a condition with static cue stimuli and a condition without cue stimuli. Differences could be identified with respect to the condition without a cue stimulus, some of which could also be determined to be significant. However, the study conducted only allows for a rough assessment of the investigated conditions, as the number of participants was very small and the variation in task completion time measurements was very high. In order to obtain a comprehensive comparison between static and adaptive stimuli and to determine their advantages and disadvantages, a much larger study must be carried out, in which specific questions are also asked about the evaluation of the presented stimulus.

Conclusion The aim of the study conducted was to investigate the extent to which adaptive visual cue stimuli are suitable for directing visual attention. In the study, we had participants perform visual search tasks under three different conditions. In each condition, the task completion time was determined and the conditions were evaluated with the help of NASA-TLX questionnaires. The results show that the visual cue stimuli, both adaptive and static, are suitable for directing attention, as they produced significantly better results than when participants received no assistance. However, the results for the condition with adaptive cues did not differ significantly from those for the condition with static cues, which may also be due to the small number of participants.

To investigate in more detail whether there are significant differences between the two conditions, a much larger study must be conducted. In this study, it would also be interesting to examine not only the performance of the two conditions, but also how the cue stimuli are perceived in terms of subtlety and which form of stimulus is preferred by the participants.

3.2.3 Subtle Gaze Guidance in Augmented Reality

In order to direct people's gaze in highly frequented public areas, visual stimuli cannot be placed in the environment for all to see. On the one hand, this would be distracting, and on the other hand, personal information would be visible to others. One possible solution is to show the visual stimuli only to the person for whom they are intended, for example by superimposing them in AR glasses. To our knowledge, no one has yet implemented subtle gaze direction methods in OST AR, so we investigated this in a study.

The study was implemented and conducted as part of the master's thesis by Jonczyk [62]. The concept was developed jointly based on a basic concept given by the author of this thesis. For this dissertation, the hypotheses were reformulated, the analysis based on the raw data was redone, and new analyses were added and discussed.

Hypotheses

The aim of the study was to examine whether subtle stimuli that proved to direct attention on PC screens could also be used in OST AR to direct gaze. Additionally, we wanted to investigate how subtle the displayed stimuli are.

We assumed that displaying subtle visual stimuli in AR would direct gaze to specific target objects better than if no stimulus was displayed. We also expected that different methods of directing gaze would lead to different results. Finally, we hypothesized that the different methods would vary in subtlety and that directing attention would even be possible without noticing the visual stimulus. Therefore, the following hypotheses were made:

- H1: Subtle visual stimuli displayed in OST AR are suitable for guiding gaze direction.
- H2: Different methods direct visual attention to different degrees.
- H3: The different methods vary in subtlety.
- H4: Directing visual attention in OST AR is possible without the user perceiving the visual cue stimulus.

Study Tasks

We tested the above hypotheses in a study in which we investigated the effect of different modulations when viewing images. The participants' task was to freely view images that were shown to them. Depending on the condition, a predefined target area was modulated using different subtle visual stimuli. We investigated the four different conditions *Luminance*, *LuminanceDiDe*, *ColorDot* and *Baseline* (see the paragraph *Design and Procedure* in this Section).

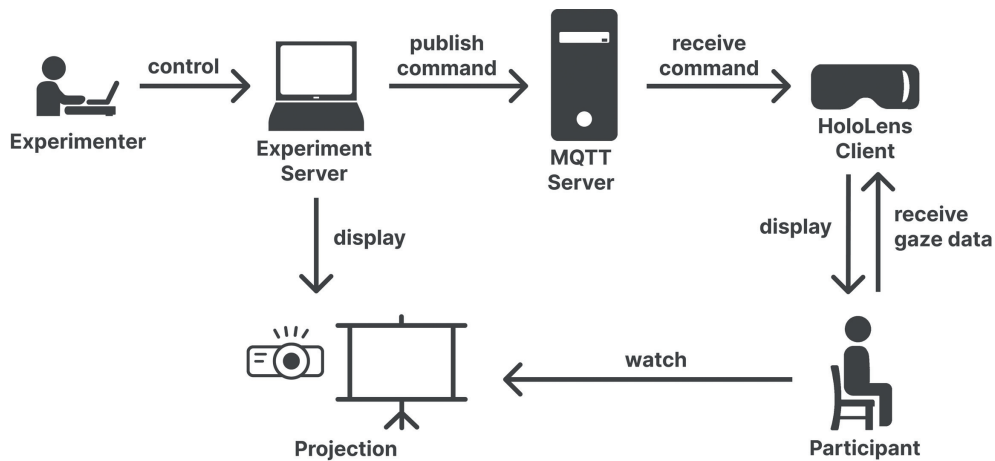


Figure 3.11: Illustration of the study setup.

Participants

20 volunteers (16 male, 4 female) aged between 19 and 64 ($M = 28.25$, $SD = 10.977$) were recruited to participate in the study. All had normal or corrected-to-normal vision, which was a prerequisite to participate in the study.

Study Setup

The study took place in a laboratory environment. This was darkened and only indirectly illuminated with a softbox studio lamp to ensure that the same lighting conditions prevailed for each participant. We used a projector to simulate different real-world environments and situations on a screen.

Participants were seated at a fixed distance of 2.5 m centered in front of the projection. They wore a HoloLens 2 whose brightness was set to 100% and in which the visual stimuli were superimposed in OST AR. The integrated eye tracker was used to determine gaze direction. The projected images were 1.34 m × 2.0 m in size, covering approximately 21.8° of the horizontal FOV in each direction, which is approximately the same as the HoloLens 2 maximum viewing angle of 43°. To be able to place the various artifacts in the AR view at the correct location in the real world, we used Vuforia¹¹, which recognized the images projected onto the canvas with the HoloLens 2 webcam and superimposed the AR overlays at the specified positions in the participant's view. Communication with the

¹¹<https://www.ptc.com/de/products/vuforia> (last accessed: 2023-08-15)

HoloLens was established using MQTT (Message Queuing Telemetry Transport). Figure 3.11 shows a visualization of the study setup.

The participants were presented different pictures in different conditions one after the other. After each picture, the participants were asked whether they noticed anything abnormal. If so, they were also asked what was unusual. At the end of the study, the participants had to fill out a concluding questionnaire. In addition to demographic questions, this also asked if the participants had seen a white flickering dot or a red dot at any time during the study. If so, they were asked to rate on a Likert scale from 1 (= not disturbing at all) to 5 (= very disturbing) how disturbing this dot was and whether they liked the Luminance modulation or the ColorDot modulation better. This rating was only done by participants who had actually experienced both visual stimuli.

Design and Procedure

The study was designed as a within-subjects experiment in which a total of four different conditions were tested. Depending on the condition, different visual stimuli were used to direct attention to the predefined target position. In the following, the four conditions which were implemented by Jonczyk [62] based on a jointly developed concept are presented in more detail:

Luminance For the condition *Luminance*, we used the subtle gaze direction method of Bailey et al. [6]. Instead of light-dark modulations, we used only white artifacts when using the HoloLens 2 as AR glasses, since the HoloLens cannot display black due to its additive color model. As with Bailey et al., the displayed artifact had a size of 0.76° of the human visual field and was switched on and off with a frequency of 10 Hz. The white intensity of the displayed circle also decreased towards the edge, which was implemented using a Gaussian falloff function. The deactivation behavior was also implemented analogous to Bailey et al. for the condition *Luminance*. The stimulus was switched off as soon as the gaze moved in the direction of the target. This was the case when the angle between the current gaze direction vector and the direction vector to the target was less than or equal to 10° [6].

LuminanceDiDe The condition *LuminanceDiDe*, which stands for **Luminance** with **D**istance **D**etection, was developed by Jonczyk [62]. For this condition, the

same subtle cue was used as for the condition *Luminance*, only the deactivation behavior was adapted. While for condition *Luminance* the stimulus was always deactivated when the gaze moved towards the target, regardless of the distance to the target, for *LuminanceDiDe* a region around the target was defined where the starting point of the saccade had to be located. Only when the gaze was already within a radius of about 25.16° (corresponding to about four times the mean saccade amplitude of 6.29° [153]) around the target point and moved towards the target, the stimulus was deactivated. In this way, we wanted to prevent the stimulus from being faded out too early and to ensure that the target was reached more often.

ColorDot For the condition *ColorDot*, we used an adapted variant of the subtle gaze direction method *ColorSquare* by Dorr et al. [22], which was introduced and used in studies by Grogorick et al. [43, 44]. Instead of a red square, we used a red circle with a size of 1° of the human visual field, as did Grogorick et al. Analogous to Dorr et al. the intensity of our red artifact was adapted to the local spatial contrast. The visual stimulus was hidden as soon as a saccade was detected or, if no saccade occurred, after 120ms at the latest.

Baseline For the condition *Baseline*, no visual stimuli were displayed in the AR glasses.

A total of 36 landscape images were selected and shown to the participants one after the other during the study (see Appendix A). Each condition was implemented for each of the images. However, each participant was shown each image in only one condition to avoid a learning effect. In total, each participant experienced each condition nine times, so a total of 180 datasets were generated per condition. The four conditions were evenly distributed among the 36 images, so that each condition was shown the same number of times on each image (five times). In total, 720 datasets were recorded during the entire study.

The images selected for the study were a variety of landscapes ranging from urban photography to beach images. All images had salient regions, but not only one area with extreme high saliency. Miniatures of the images are shown in Appendix A. The modulation spots on each image were pre-determined and were the same for each of the conditions. The selection of modulation spots was similar to that of Grogorick et al. [44]. For each image, a spot was selected



Figure 3.12: Visualization of the red modulation spots and the grey areas where no modulations should be displayed (left). Adapted from [62]. Example image with the modulation spot placed at the left door (right).¹²

that was not located in a very salient image region, but also not in a region that was completely featureless. To ensure this, a fine-grained saliency map based on the system of Montabone and Soto [115] was prepared in advance for each image. We also ensured that the selected spots on the images were spatially uniformly distributed throughout the viewer's visual FOV. In addition, the spots had to be neither too close to the initial viewing direction nor at the edge of the HoloLens FOV, since the color representation of the overlays in the HoloLens 2 changes slightly towards the edges. Figure 3.12 (left) shows the distribution of the modulation spots throughout the different pictures.

To ensure that the overlays were visible to the participants in the correct positions, we calibrated the HoloLens 2 to the eyes of the respective participant at the beginning of each study run. We then had participants view a white cross on a black background that was located in the center of the projection. Participants were allowed to freely view the image for 20 seconds. Afterwards, they were asked if they noticed anything unusual and if so, what it was. This was to test how subtle the presented stimulus was to the participant. This procedure was repeated for each of the 36 images, which were presented to the participants in random order. At the end of the study session, the participants filled out a concluding questionnaire, in which they could also rate the different conditions. In addition, the user's gaze direction was recorded throughout the study, which was used to determine the total fixation time of the target, the number of fixations of the target, and the time to the first fixation of the target.

¹²Adapted from <https://unsplash.com/de> (last accessed: 2023-08-15)

Results

We investigated the effect of the different visual stimuli on the user's gaze orientation when viewing the images. We compared the number of fixations on the target, the total fixation time of the target, and the time until the first gaze at the target. In addition, we conducted investigations on the subtlety of the individual modulations and performed heatmap investigations.

We first checked the effect of the conditions on the *Number of Fixations* and on the *Total Fixation Time* using Friedman tests with a fixed significance level of $\alpha = 0.05$ and three degrees of freedom. We report the p -value along with the test statistic χ^2 . When significant effects were revealed, we conducted post-hoc tests using Wilcoxon's signed-rank test with a fixed significance level of $\alpha = 0.05$ and 179 degrees of freedom to compare all conditions with each other. We report Bonferroni-corrected p -values, the test statistic W and the matched pairs rank-biserial correlation r as an effect size. To investigate the effect of the conditions on the *Time to First Fixation* we had to exclude all datasets where no fixation on the target was detected. Therefore, we applied a Kruskal-Wallis test with three degrees of freedom and a significance level of $\alpha = 0.05$. We report the p -value along with the test statistic χ^2 . If a significant influence of the modulation condition on the *Time to First Fixation* was detected, we used Mann-Whitney U tests ($\alpha = 0.05$) to compare all test conditions with each other. We also investigated the subtlety of the individual modulations and examined heat maps of gaze directions.

Number of Fixations The Friedman test indicated a significant influence of the modulation condition on the *Number of Fixations* of the target ($\chi^2 = 143.405, p < 0.001$). The post-hoc Wilcoxon signed rank tests revealed significant differences for condition *Baseline* compared to condition *Luminance* ($W = 1164, p < 0.001, r = -0.817$), compared to condition *LuminanceDiDe* ($W = 1090, p < 0.001, r = -0.815$) and compared to condition *ColorDot* ($W = 1403, p < 0.001, r = -0.620$). Significant differences were also detected for the luminance conditions *Luminance* ($W = 3215.5, p < 0.001, r = 0.425$) and *LuminanceDiDe* ($W = 1164, p < 0.001, r = 0.375$) compared to condition *ColorDot*. The comparison of the conditions *Luminance* and *LuminanceDiDe* showed no significant differences (see Figure 3.13, left).

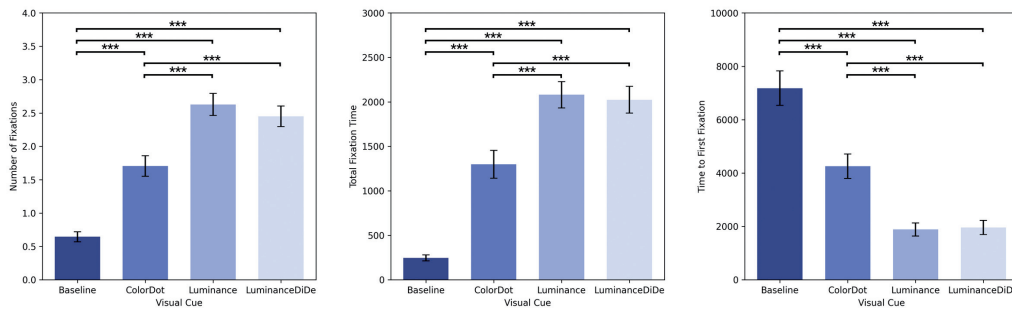


Figure 3.13: Number of fixations of the target, total fixation time and time to first fixation grouped by visual cue stimuli. Significant differences marked with *** ($p < 0.001$).

Total Fixation Time The Friedman test showed a significant influence of the modulation condition on the *Total Fixation Time* of the target ($\chi^2 = 162.069, p < 0.001$). The post-hoc Wilcoxon signed rank tests again revealed significant differences for condition *Baseline* compared to condition *Luminance* ($W = 669, p < 0.001, r = -0.909$), compared to condition *LuminanceDiDe* ($W = 772, p < 0.001, r = -0.887$) and compared to condition *ColorDot* ($W = 1496, p < 0.001, r = -0.674$). Significant differences were also detected for the luminance conditions *Luminance* ($W = 4268.5, p < 0.001, r = 0.433$) and *LuminanceDiDe* ($W = 4178, p < 0.001, r = 0.432$) compared to condition *ColorDot*. The comparison of the conditions *Luminance* and *LuminanceDiDe* again showed no significant differences (see Figure 3.13, middle).

Time to First Fixation The Kruskal-Wallis test indicated a significant influence of the modulation condition on the *Time to First Fixation* of the target ($\chi^2 = 104.270, p < 0.001$). The post-hoc Mann-Whitney U tests revealed significant differences for condition *Baseline* compared to condition *Luminance* ($U = 9372.0, p < 0.001, r = -0.653$), compared to condition *LuminanceDiDe* ($U = 8816.5, p < 0.001, r = -0.646$) and compared to condition *ColorDot* ($U = 5424, p < 0.001, r = -0.348$). Significant differences were also detected for condition *ColorDot* compared to the luminance conditions *Luminance* ($U = 13616.5, p < 0.001, r = -0.461$) and *LuminanceDiDe* ($U = 12542.5, p < 0.001, r = -0.426$). The comparison of the conditions *Luminance* and *LuminanceDiDe* showed no significant differences (see Figure 3.13, right).

	Successful Target Viewing	Unrecognized Stimuli in Successful Trials	Undetected Stimuli
Baseline	38.89%	100%	100%
ColorDot	63.89%	40%	54.44%
Luminance	90%	17.28%	20.56%
LuminanceDiDe	85%	17.65%	22.78%

Table 3.6: Percentages of successful target viewing, unrecognized cue stimuli in the successful trials and overall undetected stimuli. Adapted from [62].

Successful Gaze Guidance We investigated in how many cases participants looked at the defined target region in the individual conditions. Gaze guidance was considered successful if at least one fixation of the target occurred. Table 3.6 shows the results of successful gaze guidance to the target for the individual conditions. The highest value was achieved by the Luminance modulations, where the target was viewed successfully in 90% (*Luminance*) and 85% (*LuminanceDiDe*) of the cases. For the condition *ColorDot* the participants looked at the target in about 64% of the cases. In the condition *Baseline*, the target region was viewed at least briefly in about 39% of the cases.

Subtlety We evaluated participants' responses to the question of whether they noticed anything unusual when looking at the last image. The results of this analysis can be found in Table 3.6. We first examined how many stimuli went undetected. In the condition *Baseline*, in which no stimuli were shown, no stimuli were detected by the participants. The *ColorDot* stimuli remained undetected in over 54% of the cases, while in the condition *Luminance* and *LuminanceDiDe* it was only around 22%.

In addition, we examined the rate of unrecognized stimuli for the runs in which visual attention was successfully directed to the target. For condition *Baseline*, in which no visual stimulus was presented, no one perceived anything unusual, so the rate here is 100%. The second best value is achieved by condition *ColorDot*, in which in 40% of the cases the gaze was directed to the target region without being noticed. For the conditions *Luminance* and *LuminanceDiDe* the gaze was correctly directed in slightly more than 17% of the cases without being perceived by the participants.

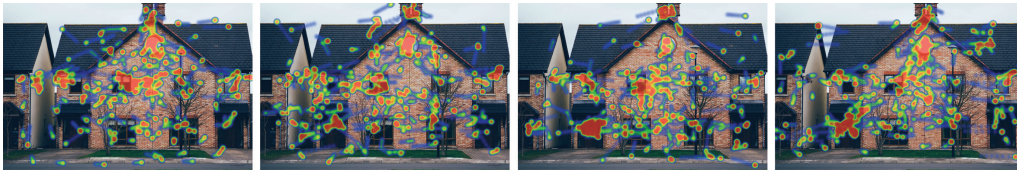


Figure 3.14: Heatmap visualization for the different modulation variants. Each image shows the distribution of eye data from five different participants. Conditions from left to right: Baseline, ColorDot, Luminance, LuminanceDiDe. The target was located at the left door (see Figure 3.12, right).

Preference Besides the dependent measures reported above, each participant was asked in the concluding questionnaire to rank the modulations. Here, the participants were asked to specify whether they would prefer the ColorDot or the Luminance modulation. A distinction between the two Luminance modulations was not possible at this point, since the difference was not apparent to the participants during the study. Moreover, only those participants who had perceived both Luminance and ColorDot modulations were able to make the evaluation. A total of 13 ratings were given. 10 of the participants preferred the ColorDot modulation while only 3 participants rated the Luminance modulation better.

Heatmap Observations We additionally evaluated the eye tracking data by creating heatmaps on the images that reflect the participants' gaze when viewing the images. Figure 3.14 shows examples of heatmaps of the same image visualized for the different modulation variants. Each image was viewed by 5 different participants in each condition; the heat maps visualize the eye data of all 5 participants. No participant was shown an image twice to rule out a learning effect, so the eye data from all 20 participants is included in the figure. In this example image, it is very clear how differently the images were viewed in each condition. In the condition *Baseline*, the main focus was on the large round window, which has a very high saliency, and there was no fixation of the target which was located on the left door (see Figure 3.12, right). In the condition *ColorDot*, the gaze was already temporarily directed to the target. With the Luminance modulations, especially with the condition *Luminance*, the successful steering of the gaze to the target is even more obvious.

Discussion

We hypothesized that subtle visual cue stimuli displayed in OST AR are suitable for guiding gaze direction (H1). Our results confirm hypothesis H1. In all conditions where subtle cues were used, we found a significantly higher number of fixations of the target as well as a significantly higher total fixation time of the target. At the same time, without visual cues it took significantly longer to fixate the target than in the other conditions. The results show that even in the condition without a cue, the participants' gaze reached the target region in some cases. The long time needed to fixate the target for the first time, as well as the low duration and the low number of fixations of the target, lead to the assumption that these were mainly accidental hits when scanning the image with the eyes. When visual cues were used, the gaze was directed to the target significantly faster and for longer.

Furthermore, we hypothesized that different methods direct visual attention to varying degrees (H2). Hypothesis H2 is also supported by our results. We found significant differences between the condition *ColorDot* and the luminance modulations *Luminance* or *LuminanceDiDe* in fixation duration, fixation time and time to first fixation of the target. The differences between the individual conditions can also be seen in the number of successful trials. If the modulations *Luminance* or *LuminanceDiDe* were used to direct attention, it was possible to direct the gaze to the target in significantly more cases than with the condition *ColorDot*, which itself, however, performed significantly better than the condition *Baseline*.

We hypothesized that different methods differ in subtlety (H3) and that directing visual attention is possible even without perceiving visual cues (H4). The results of the conducted study support both hypothesis H3 and hypothesis H4. More than half of the visual stimuli presented in condition *ColorDot* were not perceived by the participants. In conditions *Luminance* and *LuminanceDiDe*, slightly less than a quarter of the stimuli were not seen. Thus, a clear difference between the conditions is noticeable. In more than 17% of the cases where gaze was successfully guided with the Luminance modulations, the visual stimulus remained undetected. For the *ColorDot* modulation, it was as high as 40% of the successful trials in which participants did not perceive anything unusual. From the results, we can conclude that subtle cue stimuli superimposed in OST AR are able to successfully guide gaze.

With regard to the evaluation of preference, the ColorDot modulations performed significantly better than the Luminance modulations. It is assumed that this is the case because, firstly, they were noticed less often and, secondly, they were less annoying due to the constant representation (no flickering).

Limitations

In our study, we investigated whether subtle visual cues displayed as overlays in AR glasses are suitable for directing visual attention. In our investigations, we mainly restricted ourselves to well-known methods from related work, which have already proven to be useful in tests on computer screens or in VR. However, many other visual modulations are imaginable that could be suitable for guiding gaze in AR and would need to be investigated accordingly. Likewise, an adaptation of the already investigated modulations is conceivable. For example, it could be investigated to what extent the success rate of the ColorDot modulation improves when the stimulus is repeatedly displayed until a response to it is detected, and what influence this has on the subtlety of this method. Overall, it would make sense to investigate to what extent stimulus intensities adapted to the user can lead to a better subtlety.

We calibrated the eye tracker of the HoloLens to the eyes of the participant using the corresponding app. We cannot guarantee that this calibration always worked properly and that the overlays were displayed correctly in depth. However, the eye gaze logs do not indicate any irregularities in this regard, and none of the participants reported anything of this nature. Similarly, there could theoretically have been misplacement or flickering of the overlays due to instability in image recognition with Vuforia. Again, this effect was not reported by participants and recognition was also extremely stable and error-free in our pre-tests.

In our study, we used projected images instead of the real world so that we could test a wider variation of environments. An investigation of how the methods behave in real environments would be useful as a supplement.

To ensure that each participant had the same testing conditions, we conducted the study in the laboratory under fixed lighting conditions. Changing the lighting conditions is expected to change the results, as the overlays will be more or less obvious depending on the brightness. This needs to be investigated.

Conclusion

We conducted a study to examine whether subtle visual stimuli superimposed in OST AR glasses are suitable for directing visual attention. We investigated methods known from related work, using a red dot (ColorDot) and a black-and-white modulation (Luminance), which we adapted to the optical see-through conditions. We additionally created an adaptation of the Luminance modulation and compared all three modulations with the baseline, in which the tasks had to be solved without visual cues.

The results of our study show that directing the gaze works with the help of the subtle visual stimuli we have studied. The methods vary in their success rate and subtlety. The Luminance modulations provided significantly better results, but are clearly less subtle than the ColorDot modulation. Accordingly, it cannot be clearly determined which method is the better one. Depending on the use case, it must be decided whether success rate or subtlety is more important. In addition, many other subtle visual cues are imaginable that might be suitable for directing gaze, but would need testing to determine their suitability.

3.3 Summary

In this chapter, we first determined the basics of when visual stimuli are visible in the periphery and when details can be detected. In doing so, we have found that objects must be very close to the fixation point if small details are to be detected. The mere presence of objects is perceived at a much greater distance from the fixation point. Depending on the stimulus used, the angle here is between 45° and 55°. In the remainder of the chapter, we investigated how visual stimuli can be used to guide gaze direction. In our studies, we found that it is possible to use visual cue stimuli to direct attention in real and real-sized environments. The results of the studies show that visual cue stimuli are a great asset especially in difficult search tasks. The more obvious the visual stimulus, the faster attention is directed to the target object. However, constant stimuli are preferred over obvious flickering cues.

Besides obvious cues, subtle visual cues are also able to guide gaze direction in real and real-sized environments. Likewise in AR, gaze guidance is possible using subtle visual cue stimuli. Here, the constant stimulus (red circle) is also preferred

over the flickering luminance modulation, which is, however, significantly more effective in AR.

There will probably be no visual cue that is equally well suited for every situation. Rather, depending on the use case, it must be decided whether the stimulus should be as subtle as possible or whether efficiency is more important. Based on this, a suitable visual stimulus must be selected.

Chapter 4

Understanding Visual-haptic Perception of AR Proxy Interaction

The results of the last chapter have shown that visual augmentations of reality significantly influence our perception of the environment. Therefore, we need to carefully consider which augmentations are used and how they influence our perception. Especially when we want to use physical proxies to interact with virtual objects, it matters how the visual and haptic perception of the objects is affected by the augmentation, because in this case a direct link between reality and the virtual object exists.

In AR, tangible proxy objects are used to interact with virtual content in order to achieve a more seamless interaction between the real and virtual world. Since it is not possible to create and use a single matching physical replica for interaction with each virtual object, abstracted proxy objects are needed that can be used for interaction with a variety of virtual objects. By using abstracted proxy objects there is inevitably a difference in size, shape, texture and material between the physical and the virtual object, which can have an influence on the perception of these objects during the interaction. Therefore, we wanted to investigate to what extent abstracted proxy objects can be used without significantly degrading e.g. usability or performance.

We investigated in several studies to what extent a variation in size or shape between the virtual object and its physical representation is feasible and what

influence the environmental illumination has (see Section 4.2). The results of these studies have already been published in papers [65, 68, 69]. The studies were implemented based on the framework presented in Section 5.2. In addition to measuring performance in terms of task completion time, we evaluated usability and presence using questionnaires (see Appendix B). The creation of the presence questionnaires was based on well-known questionnaires used in VR studies (see Section 4.1).

4.1 Measuring Presence in Tangible AR

In Virtual Reality the feeling of reality is called presence, which is high when a person feels that he or she is really in the represented virtual world [52]. This can be determined with a variety of different VR presence questionnaires [131]. However, the VR presence questionnaires cannot be used for the evaluation of AR applications. The reason is that one is in reality, enriched with virtual information, but not in a virtual environment. Evaluating the presence in VR, i.e. how much one feels immersed in virtual reality [134], is therefore not applicable in AR. A search for comparable questionnaires for evaluating the feelings of reality in AR applications revealed only the ARI (Augmented Reality Immersion) questionnaire by Georgiou and Kyza [38]. This was developed specifically for location-based AR, assessing a much wider definition of immersion from game-based research containing factors like attraction or usability of applications, while we focus on a single factor, presence, defined by Heeter as the feeling of “being there” [52]. Therefore, new questionnaires had to be created.

To generate initial ideas for an AR questionnaire to measure presence, we first analyzed the most widely used VR presence questionnaires – the IPQ [58] and the VR SUS questionnaire [148] – to see if they could be used in their entirety, or at least in part, for an AR questionnaire, or if they could be converted into an AR questionnaire through minor adaptations. We found that the IPQ is very VR-related and mainly deals with the sense of being present in VR. Only a very small number of questions concerning sense of reality could be transferred to an AR questionnaire. Regarding the VR SUS, however, we identified the option to transform 4 of the 6 questions into an AR setting. Only questions number 3 and 6 were too VR-related.

Our goal was to measure the tangible qualities of TAR applications. However, there are many AR applications that do not use tangible interaction and which

would therefore be excluded when generating a TAR-specific questionnaire. Therefore, we decided to generate two versions from the 4 identified VR SUS questions: one questionnaire for pure AR experiences with a focus on the visual perception, and one questionnaire incorporating tangible interaction. This allows us to evaluate pure AR applications with the AR presence questionnaire and, in the case of TAR applications, to additionally test the interaction experience with the TAR presence questionnaire.

Since the questions in VR SUS were thoroughly determined by studies, an attempt was made to transfer them 1:1 from VR to AR. The question “To what extent were there times during the experience when the virtual environment was the reality for you?”, e.g., was transformed into “To what extent were there times during the experience when the visual overlays were reality for you?” in the AR presence questionnaire (see Appendix B.1.1), and to “To what extent were there times during the experience when the visual overlays felt real during the interaction?” in the TAR presence questionnaire (see Appendix B.1.2).

For the studies on lighting variations (see Section 4.2.3) and on shape variations (see Section 4.2.4) we additionally created shortened versions of the questions to be able to display them in the limited AR view (see the paragraph *Tangible AR Questionnaire* in Section 5.2.2).

The questionnaires are evaluated in a different way than the VR SUS presence questionnaire. The AR and VR scores would not be comparable due to the different number of questions. In addition, it is to be expected that presence in OST AR is not rated as high as presence in VR, since one is in a mixed reality environment with slightly translucent overlays, which is naturally more confusing than, for example, being in a purely virtual world. Therefore, the evaluation of the AR presence questionnaire and the TAR presence questionnaire is carried out over the total mean value of all four questions.

Using two questionnaires with four items each can be a first step to measure the feeling of presence but does not serve as a definitive measuring tool for the overall experience of users in the augmented environment.

Therefore, in addition to the presence questionnaires, we created specific questionnaires for each study, e.g., to find out how disturbing something is perceived to be (see Appendix B.2).

4.2 Perception upon Interaction with a Divergent Proxy Object

All conducted studies followed the same basic approach (see Section 4.2.1). We considered size differences between the virtual and physical object (see Section 4.2.2), the influence of illuminance on the perception of size differences (see Section 4.2.3), and differences in object shape (see Section 4.2.4).

4.2.1 Basic Approach for Evaluating AR Proxy Interaction

All studies took place in the same setting following the same procedure. The studies were conducted in a quiet laboratory environment. The room was darkened and illuminated only with artificial light to ensure that all participants had the same lighting conditions. Tracking was performed using OptiTrack cameras that were mounted at a height of about 2.6 m on a truss and aimed at the center of the tracking area.

The participants sat on a chair at a table in the middle of the tracking area. The position of the chair in relation to the table was defined, as was the distance between the subject's head and the top of the table (see the paragraph *Head-Desk Distance Check* in Section 5.2.2). This ensured that all participants had a similar perspective on the interaction elements. Outside the tracking area was a table for the experimenter, which was used for the secret arrangement of the physical props and the operation of the Experiment Server (see Section 5.2). For visualizing the virtual overlays and the targets, a HoloLens 2 was used.

The tasks were each performed on a single-color background that was larger than the plate on which the objects were placed to ensure that the same background was visible behind the objects at all times, even when they were held in the air, for example. This ensured that the objects would stand out equally well against the background in every situation.

Furthermore, the automatic eye calibration was performed using the application available for this purpose on the HoloLens 2. Subsequently, this calibration was checked with the help of a calibration triangle and if a fine calibration to the eyes of the respective participant was necessary, this was carried out (see the paragraph *Manual Eye Calibration* in Section 5.2.2).

Prior to each individual task, we ensured that the user's FOV was set correctly so that the overlays were completely visible at all times and covered the physical objects. For this purpose, subjects were asked to adjust their FOV to markers on the table (see the paragraph *Field-of-View Check* in Section 5.2.2).

Before each run, the props were secretly arranged on a plate by the experimenter and the plate was covered with a box. The box was then placed in front of the participant and, when the FOV was correctly set, quickly removed by the experimenter. This also automatically started the measurement of the processing time of the respective task (see the paragraph *Automatic Task Start and Timing* in Section 5.2.2).

After each individual task, questionnaires had to be filled out by the participants. In the study on size variations, these questionnaires were still filled out on paper, while in the studies investigating lighting variations and shape variations, the questionnaires were completed in AR with the aid of a proxy pen object (see the paragraph *Tangible AR Questionnaire* in Section 5.2.2).

At the end of each study trial, a concluding questionnaire had to be completed, which included general questions related to the respective study as well as demographic questions.

4.2.2 Investigation of Size Variations

In a first step, the factor of size differences between the tangible object the user is interacting with and the overlying virtual representation in terms of performance, usability and immersion in OST AR was investigated. We wanted to find out whether it is feasible to use smaller or larger props as interactive elements for a virtual object. In a study, we therefore investigated to what extent the size of the physical object can vary from the size of the virtual overlay without a significant worsening of execution time, feeling of disturbance and feeling of presence.

The study was implemented and executed by Ruble [128] as part of his bachelor's thesis. It is based on a concept given by the author of this thesis, which was refined together. The results of the study have been processed and already published in a joint paper [68].

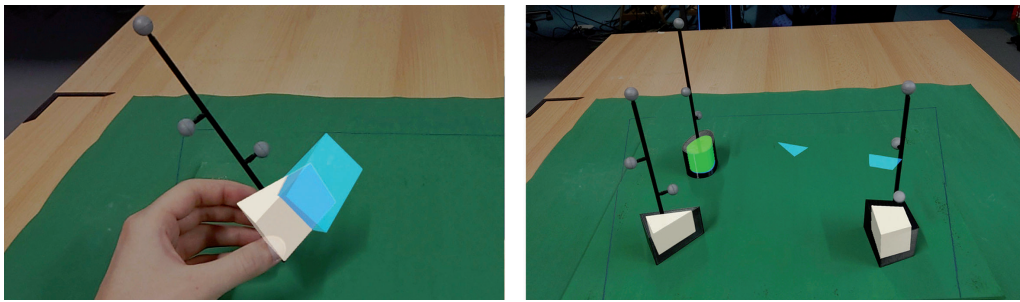


Figure 4.1: Participant's perspective for part 1 of the study with size condition M (left): Fitting virtual overlay (white) to 3D target (blue). HoloLens view of study part 2 (right): Egg-shaped cylinder placed on 2D target (size condition XS).

Hypotheses

The main focus of the study was to find out if it is possible to use a larger or smaller tangible prop compared to the virtual object without extreme losses in usability, and if there is a range within which presence is felt to be almost the same. Furthermore, we wished to test whether differences in size, as in VST AR [85], have no effect on performance and whether, as assumed, the size conditions would be correctly assessed by the participants. Therefore, the following hypotheses were formulated:

- H1: The size of the virtual and physical object can differ within a certain range without significant loss in usability.
- H2: The size of the virtual and physical object can differ within a certain range without significant worsening of "AR presence" and "TAR presence".
- H3: Differences in size between virtual and physical objects have no influence on performance.
- H4: Differences in size between virtual and physical objects can be estimated correctly by the participants.

Study Tasks

Two tasks were defined to test the above hypotheses. The first part of the study was exploratory, so that the participants could observe and feel differences without time pressure and could become familiar with the interaction in TAR. During this exploration phase, the participants had the task of successively fitting virtual



Figure 4.2: Study setup: The participant’s interaction area positioned at the center of the tracking zone.

objects represented as overlays on the physical prop to virtual 3D targets (see Figure 4.1, left).

In the second part of the study, we additionally wanted to find out if a difference in size between a virtual object and a physical object has an impact on performance. Therefore, we had the participants solve puzzle tasks under time pressure. For this task, three different objects had to be placed on corresponding visualized 2D targets on a plate. We decided to have participants interact with multiple objects, so that the influence of disturbances during grasping would be increased [85] (see Figure 4.1, right). In both parts of the study, docking tasks were chosen that required grasping, rotating, and arranging of objects. These tasks – even if they seem simple – represent basic elements in complex goal-oriented activities [104]. Regardless of the use case a physical prop is used for, this tangible object is always grasped, lifted, turned and placed, whether it is e.g. a game piece on a virtual board or a piece to configure a composite object.

Participants

14 volunteers (9 male, 5 female) aged between 21 and 28 ($M = 24.5$, $SD = 2.279$) were recruited to participate in the study. All had normal or corrected-to-normal vision and 12 were right-handed. The participants were asked about their prior

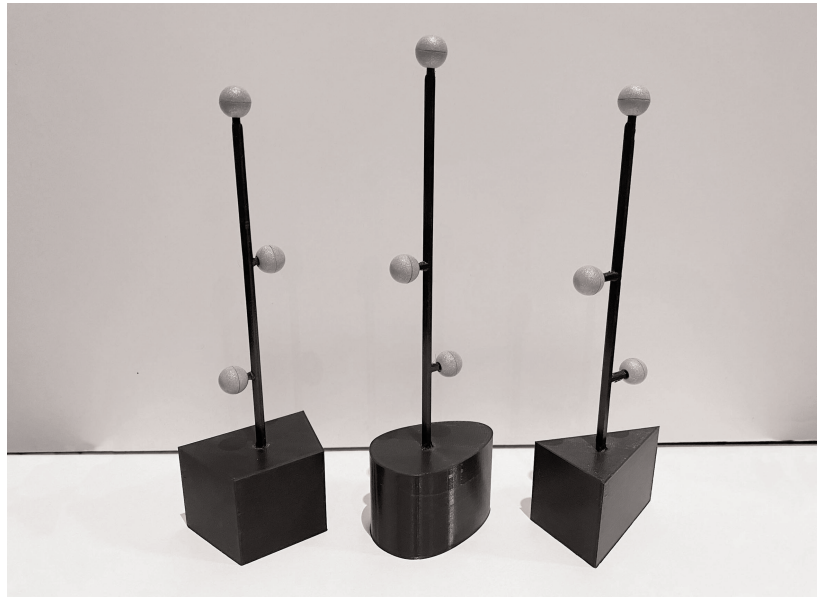


Figure 4.3: Physical props with attached marker trees on top.

experience with AR in general, as well as AR glasses, on a 7-point Likert scale from 1 (= never) to 7 (= regular). They reported mostly low experience with AR ($M = 2.214, SD = 1.762$) and minimal experience with AR glasses ($M = 1.571, SD = 1.089$).

Study Setup

For the study, the laboratory environment was darkened and only indirectly illuminated by two softbox studio lamps (see Figure 4.2) to avoid the influence of different lighting conditions. Participants' heads and physical props were tracked through a combination of 11 OptiTrack Flex 3 cameras, which were mounted on a truss of about 4 m x 4 m.

The tasks were performed on a monochrome green background. The physical props were black and equipped with black marker trees on top (see Figure 4.3). The color of the overlays was set to white, which is the least translucent color on the HoloLens. Additionally, the opacity of the overlays was set to 100% and the brightness of the HoloLens 2 to maximum to achieve the lowest possible translucency of the overlays. The distance between the chair and the desk was constant, and the distance between the markers on the HoloLens 2 and the desk was adjusted to 45 cm.

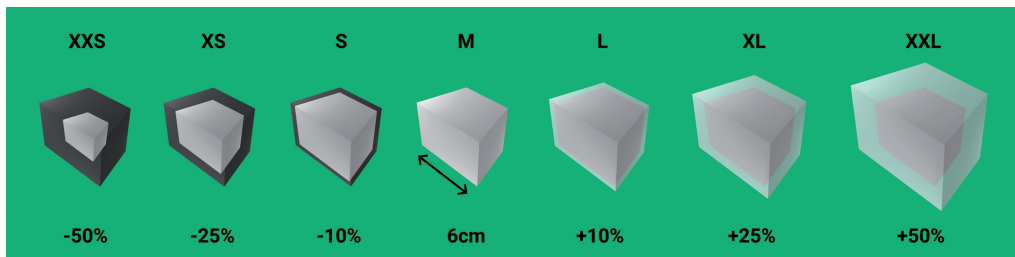


Figure 4.4: Size variations of the virtual overlays (white) compared to the physical proxy objects (black). Condition M is the base condition with matching size of virtual and physical object.

After completion of each condition, three questionnaires had to be completed: an AR and a TAR presence questionnaire (see Section 4.1 and Appendix B.1) and a size perception questionnaire (see Appendix B.2.1). The size perception questionnaire contains questions that focus on the perception of size differences and perceived disturbance.

All questionnaires were rated using 7-point Likert scales. For example, size was assessed by asking participants to rate the size of the virtual object compared to the physical object from 1 (= much smaller) to 7 (= much larger). By using a proprietary questionnaire instead of a standard usability questionnaire such as NASA-TLX [48], it was possible to specifically examine how the interference was perceived when grasping and interacting with the object. We deliberately refrained from additionally measuring usability with NASA-TLX in order to keep the amount of work demanded of participants as low as possible, since the questionnaires had to be filled out 14 times by each participant.

Lastly, the participants answered a final questionnaire (see Appendix B.3.1). Here, demographic information was requested in addition to a classification of the size ratios based on performance and usability.

Design and Procedure

The study was designed as a within-subjects experiment. In total, seven different size conditions were tested. The order of the size conditions in both parts of the study was counterbalanced by a Williams design Latin square (LS) of size seven [161]. Figure 4.4 shows the size variations of the virtual overlay. Condition M represents the baseline where virtual and physical object have an equal length of 6 cm. Sizes S and L portray a small size variation with 10% difference in length;

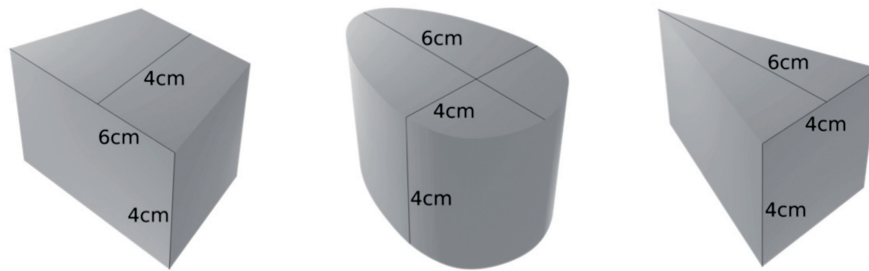


Figure 4.5: Shapes of the tangible objects. Left to right: trapezoidal prism, egg-based cylinder and triangular prism.

width and height are always scaled by the same factor as length. Following are conditions XS and XL with a size variation of 25% as well as XXS and XXL with a 50% size difference from the physical object. The small size difference (S and L) was chosen to find out if only small size variations are possible without serious losses in the measured values. We used a minimal size difference of 10% instead of 5%, in contrast to de Tinguy et al. [21], because our research interest was different. Instead of investigating at what point the user notices the change, we wanted to find out how much we can vary size without causing a significant degradation in usability and performance. Based on a pre-test with four persons, 5% was expected not to cause such effects. The maximum size difference (XXS and XXL) was chosen to be accordingly large (+/-50%), in order to find the limit of possible size variations.

In the study, three different 3D printed shapes were used, with which participants had to interact (see Figure 4.5). These shapes were intended to be different basic shapes, which create a distinct feeling when touching and interacting with them. Instead of common bases like an equilateral triangle, square and circle, we purposely modified them to guarantee that there is only one possibility to match a given target. Participants thereby have to perform a maximal rotation (up to 180°) of the physical prop. Our aim was to provoke more interaction with the objects and give participants the opportunity to perceive the influence of size variations. We chose a length of 6 cm and a width and height of 4 cm because this size can be easily grasped [36]. Furthermore, we oriented our design to existing investigations in VR and VST AR to produce comparable results [21, 85].

Part 1 In part 1 of the study, participants interacted with the different shapes in sequence. The order of the interaction with the three different shapes was

counterbalanced by a 3 x 3 LS for each condition. There were six possible positions where the 3D targets could be placed, all with equal distance to the initial position of the physical prop. The selection of the position was balanced by a 6 x 3 LS for each condition and prop shape.

The orientation of the physical objects on the plate as well as the rotation of the 3D targets were determined randomly. However, for the targets, only rotations which obey the following rules were considered: The upward normal vector of the 3D target must not point downwards or be too close to pointing sideways, and the upward normal vector must not form an angle with the vector to the viewer's eyes which is too close to 90°. These rules ensure that all targets are solvable without head movements, as this would require more time to solve the tasks. Furthermore, they ensure that the physical props do not have to be flipped, as this is not possible due to the marker trees on top.

To prevent ambiguity, the undersides of the virtual props and the virtual targets were colored orange to make them distinguishable from the top side, which was communicated to the participants at the beginning.

For each of the seven size conditions, all three shapes were interacted with successively. The task was to match the displayed virtual object to a 3D target object of the same shape and size in position and orientation (see Figure 4.1, left). Once this was achieved accurately enough, the overlay temporarily turned green and the next target was displayed immediately. When all six targets were solved, the overlay stayed green and no new targets appeared. A matching was determined "solved" exactly when errors below a threshold of 1 cm in distance and 30° in angle between prop and target were detected consistently for 0.5 seconds. In a pre-test we found that these values provide the best mix of feasibility and complexity.

Part 2 In part 2 of the study, participants could interact with all objects at once. In this task, there were three different positions at which virtual targets were placed. Thus there was a total of six different arrangements for the three props. These arrangements as well as the orientations of the individual targets were randomized. Likewise, the initial arrangement of the physical props and their initial orientation on the plate was random.

For each of the seven size conditions, two puzzle tasks had to be solved. The task was to place the virtual objects as quickly as possible onto the displayed virtual 2D targets on the table (see Figure 4.1, right). Before each task, the props

were arranged on the plate out of the participant's view and covered with the box. Measurement of task completion time started automatically once the box was removed and thus simultaneously with the display of virtual objects and targets. As soon as an object was placed and oriented correctly, its overlay color changed to green. Once all objects were placed and oriented correctly, the task was considered solved and time measurement stopped automatically. A placement was determined as "solved" exactly when errors below a threshold of 0.5 cm in flat distance, 1 cm in height and 7.5° in angle between prop and target were detected consistently for 0.5 seconds. These values were also determined with test participants beforehand.

We decided to use well-defined deviations in distances and rotations as a stopping criterion for the task instead of performing an evaluation with regard to the error distance and error rotation because the evaluation of the task completion time was important to us. Letting participants self-assess whether a task was solved would have greatly affected the evaluation of performance and led to uncertain study durations, as some individuals are inherently more accurate than others.

Results

We investigated the effect of size variations between a physical object and a corresponding virtual overlay on the usability (by disturbance ratings when grasping and interacting with objects), on the feeling of presence (by AR and TAR presence ratings), on the size perception (by estimates of the virtual object size compared to the physical object) and on performance (by task completion time). We evaluated these four types of results for both parts of the study individually (except for task completion time, which was only measured in part 2) using the following procedure: First we checked for the overall effect of size condition on the measured result using a Friedman test with a fixed significance level of $\alpha = 0.05$ and 6 degrees of freedom. When significant effects were revealed, we conducted post-hoc tests using Wilcoxon's signed-rank test again with a fixed significance level of $\alpha = 0.05$ and 13 degrees of freedom to find which size conditions differed from the size matching condition M, which we set as our baseline condition. In addition to the resulting p -value, the matched pairs rank-biserial correlation r is given as an effect size. Figure 4.6 summarizes our results and highlights which conditions were found not to differ significantly from the size matching condition M. However, this does not imply equality of such conditions.

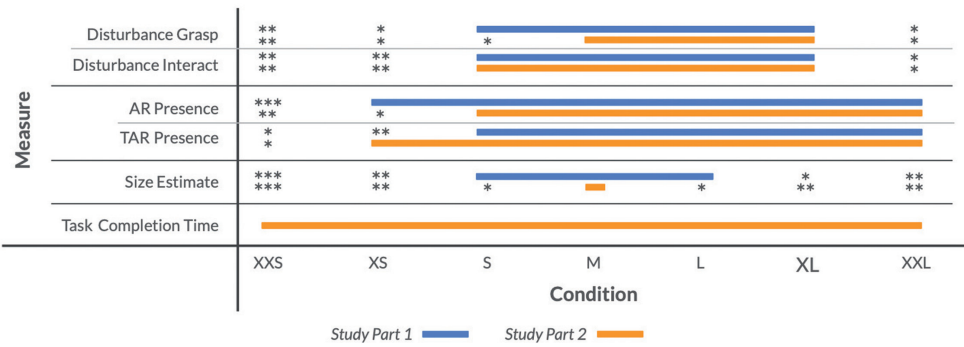


Figure 4.6: Ranges of size conditions without significant difference from baseline condition M for each measure, divided into the two study parts. Significant differences are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

Disturbance In part 1 of the study, the Friedman test indicated a significant influence of size condition on the scores of disturbance for grasping ($\chi^2 = 42.405, p < 0.001$) and for interaction with the objects ($\chi^2 = 48.129, p < 0.001$). For grasping, Wilcoxon's signed rank test revealed significant differences for XXS compared to M ($W = 0, p = 0.002, r = 1$), XS compared to M ($W = 8, p = 0.014, r = 0.795$) and XXL compared to M ($W = 7.5, p = 0.02213, r = 0.773$). Conditions S, L and XL did not differ significantly from M in their grasping disturbance scores. Similarly for interaction, the post-hoc tests showed significant differences for XXS compared to M ($W = 0, p = 0.002, r = 1$), XS compared to M ($W = 0, p = 0.009, r = 1$) and XXL compared to M ($W = 7, p = 0.036, r = 0.745$). Again, for conditions S, L and XL, no significant difference in interaction disturbance was detected.

In part 2 of the study, the Friedman test also indicated a significant influence of size condition on the scores of disturbance for grasping ($\chi^2 = 41.196, p < 0.001$) and for interaction with the objects ($\chi^2 = 31.676, p < 0.001$). For grasping, Wilcoxon's signed rank test revealed significant differences for XXS compared to M ($W = 0, p = 0.002, r = 1$), XS compared to M ($W = 0, p = 0.013, r = 1$), S compared to M ($W = 4, p = 0.025, r = 0.822$) and XXL compared to M ($W = 0, p = 0.021, r = 1$). For conditions L and XL, no negative influence could be identified. Similarly for interaction, the post-hoc tests only showed significant differences for XXS compared to M ($W = 1.5, p = 0.002, r = 0.967$), XS compared to M ($W = 2, p = 0.006, r = 0.939$) and XXL compared to M ($W = 2.5, p = 0.033, r = 0.861$).

Therefore we can conclude a significant effect of size variation on disturbance during grasping and interaction. For grasping, conditions XXS and XS with large and medium size reduction result in significantly higher disturbance scores, followed by condition XXL with large size increase and smaller effect. In part 2 of the study, even a small size reduction (condition S) led to such an effect. For interaction, conditions XXS and XS with large or medium size reduction also show significantly increased disturbance, and again the only condition with increased virtual object size having this effect was XXL.

Presence In part 1 of the study, the Friedman test indicated a significant influence of size condition on the scores of AR ($\chi^2 = 30.296, p < 0.001$) and TAR presence ($\chi^2 = 22.266, p = 0.001$). For AR presence, Wilcoxon's signed-rank test revealed significant differences only for XXS compared to M ($W = 3, p < 0.001, r = -0.943$) while for TAR presence significant differences could be found for XXS compared to M ($W = 13.5, p = 0.011, r = -0.743$) as well as XS compared to M ($W = 8, p = 0.003, r = -0.848$).

In part 2 of the study, the Friedman test also indicated a significant influence of size condition on the scores of AR ($\chi^2 = 33.468, p < 0.001$) and TAR presence ($\chi^2 = 24.752, p < 0.001$). For AR presence, Wilcoxon's signed-rank test revealed significant differences for XXS compared to M ($W = 5.5, p = 0.001, r = -0.895$) and XS compared to M ($W = 10, p = 0.014, r = -0.78$). However, for TAR presence significant differences could only be found for XXS compared to M ($W = 8.5, p = 0.011, r = -0.813$).

For condition XXS with large size reduction, a significant worsening was found in both parts of the study and for both types of presence assessed, while for condition XS with medium size reduction, a significant worsening could only be found for TAR presence in part 1 and AR presence in part 2. Enlargements of the virtual objects (conditions L, XL and XXL) or only a slight size reduction (condition S) did not lead to significantly lower presence scores.

Size Estimate The participants estimated the size of the virtual object compared to the size of the physical object on a 7-point Likert scale. Therefore, we can analyze the effect of the actual size condition on the participants' size perception. Friedman tests indicated a significant influence of size condition on the perceived size in part 1 ($\chi^2 = 69.361, p < 0.001$) and 2 ($\chi^2 = 75.783, p < 0.001$) of the study. Wilcoxon's signed-rank test as post-hoc revealed that in part 1, only conditions

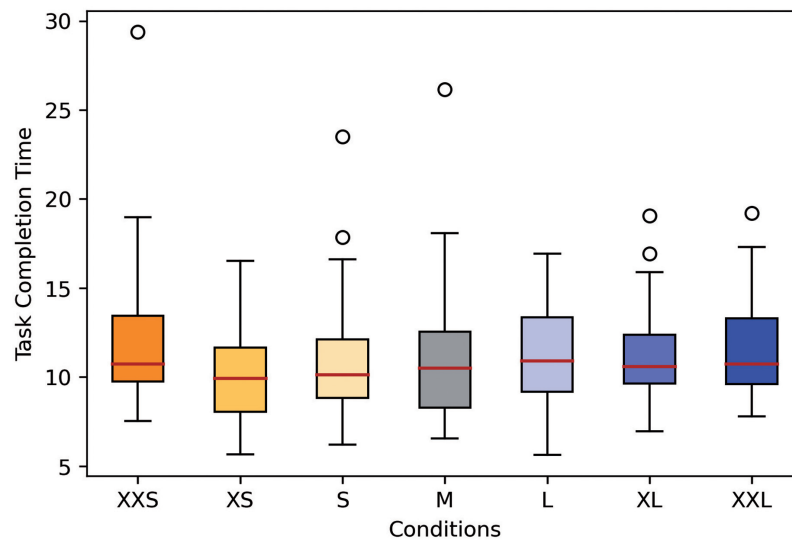


Figure 4.7: Task completion times in seconds for each size condition in part 2 of the study.

S ($W = 3, p = 0.233, r = -0.6$) and L ($W = 2, p = 0.773, r = 0.333$) had no significant differences in the size estimate compared to M as a baseline, whereas in part 2 of the study, all size conditions differed significantly in their estimate from the baseline.

These values show that for small size differences (conditions S and L), the size of the virtual overlays could not always be correctly estimated.

Task Completion Time The results of the time measurements in part 2 of the study are displayed in Figure 4.7. The Friedman test indicated a significant influence of size condition on task completion time overall ($\chi^2 = 14.082, p = 0.029$). However, Wilcoxon's signed-rank test revealed significant differences only between the conditions with size variation, and none compared to the size matching condition M.

Final Questionnaire Besides the dependent measures reported above, each participant was asked in the concluding questionnaire to rank the seven size conditions with respect to perceived realism and perceived easiness. Table 4.1 shows the cumulative sum of the scores of all participants for the seven conditions. The highest valued condition is given 7 points, the second 6 points and finally the lowest valued condition 1 point each in the sum.

	XXS	XS	S	M	L	XL	XXL
Realism	20	35	59	87	83	65	43
Easiness	29	39	59	78	76	64	47

Table 4.1: Ranking scores of the size conditions for realism and easiness.

	XXS	XS	S	M	L	XL	XXL
pleasant	2	2	10	13	12	8	6
efficient	3	4	9	11	12	9	5

Table 4.2: Number of classifications for each condition as pleasant and efficient (out of 13).

Consistent with the evaluation of the AR and TAR presence scores, conditions M, L, XL and S were ranked highest in descending order in perceived realism. Regarding easiness, the order is identical to that of realism.

In addition, the participants had to indicate in their rankings up to which state the conditions feel pleasant and when they change to unpleasant (realism ranking) and up to which state the conditions feel efficient and when they start feeling inefficient (easiness ranking). Due to an error in filling out the questionnaire, one participant had to be excluded. Table 4.2 shows that conditions M, L, S and XL were rated mostly pleasant in descending order. This matches the evaluation of the disturbance scores, which showed higher disturbance with all other conditions compared to M. Condition M is the only condition which everyone agreed to be pleasant. Regarding efficiency there is a tendency towards conditions with larger virtual objects rather than smaller ones. Here L, M, XL and S were rated most efficient in descending order.

The participants also had the opportunity to submit comments on the study in a free text field. Three of them mentioned that the interaction feels more real, is easier, or is less disturbing when larger virtual objects are used, as these cover the physical objects. Another three participants pointed out that there is a certain delay between the actual hand movements and the movement of the overlay, especially in fast movements.

Sickness after the experiment was rated on the scale from 1 (= not at all) to 7 (= very sick) as low ($M = 1.357$, $SD = 0.633$) with a maximum of 3 by one person.

Discussion

We hypothesized that the size of the virtual and physical object can differ within a certain range without significant loss in usability (H1). Furthermore, we hypothesized that the size of the virtual and physical object can differ within a certain range without significant worsening of “AR presence” and “TAR presence” (H2). Our results show that in our study setup the size differences between the tangible prop and the virtual object it represents are accepted within a certain size range without worsening both the feeling of disturbance and the feeling of presence.

Figure 4.6 shows for which conditions no significant difference from the baseline M could be detected. For our setup, the results for both parts of the study are very similar. The strongest difference can be seen in the area of size estimation of the virtual object compared to the physical object. Contrary to our expectations, participants had difficulties estimating the size of the virtual overlay compared to the size of the physical object in part 1 of the study. This contradicts our hypothesis H4 that size differences can be correctly estimated if the physical objects are visible to a certain degree due to technical conditions. A reason for this might be that the overlays had a strong covering effect and the physical objects were therefore almost not perceived by the participants. While in part 1 of the study the subjects were unable to detect the difference between condition S and M or L and M, they were able to do so in part 2 of the study. This can be explained by the fact that each participant first worked on the exploration task (part 1) and thus already knew what the different sizes were when performing part 2. This made it possible to estimate the difference better than in part 1, where subjects were sometimes first shown condition S or L before condition M and they may have initially incorrectly considered it to be the matching condition. From this it can be seen that small size differences cannot be reliably detected if there is no knowledge about other better matching objects.

The learning effect regarding sizes could also explain the different levels of disturbance during grasping. In part 1 of the study, no significant differences from baseline M were found in the range from condition S (-10%) to XL (+25%), whereas the range in part 2 was only from M (baseline) to XL (+25%). If an object is not perceived as larger or smaller than the baseline, it is more likely that the sensation of grasping for these conditions will not be judged as differently either. However, once one is aware of the size differences, this will possibly affect the further evaluation.

Disturbance during interaction was not significantly distinguishable from baseline M for both parts of the study in a range from S (-10%) to XL (+25%). The assumption is that the knowledge about the sizes did not have such a strong effect here, because the difficulty is mainly in grasping. As soon as one holds the object in one's hand, the difference in size is less of an issue.

The result concerning AR presence in our study setup shows that it was possible to increase the size by at least 50% (condition XXL) and to decrease it by up to 25% (condition XS, part 1) or 10% (condition S, part 2).

Regarding TAR presence, deviations in size were also feasible. The range is from -10% (condition S, part 1) or -25% (condition XS, part 2) up to +50% (condition XXL).

Concerning the time for the completion of the tasks, no negative influence of the size conditions could be detected compared to the baseline condition M. Therefore, as hypothesized (H3), the differences in the size of the virtual and physical object do not affect performance, with respect to the conditions investigated in the study.

Overall, it can be observed that the results are similar to comparable studies in VR and VST AR. For example, de Tinguy et al. [21], who measured in VR how much a virtual object can be resized without the user noticing the change in size, found that size changes can be made in a small range without being noticed. Regarding usability (disturbance in grasping/interacting) and presence (AR/TAR presence) it can be seen that the virtual object can be considerably larger than smaller, compared to the physical one, without having a negative impact on usability and presence. This fits with the results of Simeone et al. [133], who compared only three different virtual sizes in VR (replica, -50%, +50%), but showed that a significant deterioration of believability and ease of use was only found for the smaller virtual representation. The result regarding task completion time agrees with that of Kwon et al. [85], who tested the impact of size differences between virtual and physical objects in VST AR on performance. The similarity of the results of the studies in VR and VST AR to the results of this study can be explained by the decision to use overlays that are as opaque as possible, which is why the underlying physical objects could hardly be perceived.

The delay between movement of the physical object and the virtual overlay during faster interactions is due to the technical design (see Section 5.2) and therefore cannot be completely prevented. Since this delay was the same for all

participants for all conditions, it can be assumed that it did not negatively affect the results.

Simulator sickness after the experiment was rated very low, which was to be expected, since it occurs less frequently in AR than in VR. This result is in line with the result of the study on simulator sickness in AR of Vovk et al. [154].

Limitations

In this study, we investigated the effect of size differences between the physical object and its virtual representation on usability, presence and performance. Since an abstract proxy object used for interaction differs not only in size but also in shape, texture, and material, the next step is to find out to what extent a deviation between virtual and physical object is possible with regard to these features.

The purpose of the study was to show that instead of using an exact replica, it is possible to use a physical prop that can differ in size to a certain degree from its virtual counterpart without too much negative impact on usability, presence and performance. However, no exact limits were determined as to how much one can increase or decrease the size. To reliably determine these limits, a larger sample size and appropriate methods, such as the up/down staircase procedure [35], will be needed in further studies.

The results of the study show that a limit exists to which an overlay can be smaller than the physical prop being interacted with. For virtual overlays larger than the physical prop, the limit in terms of AR/TAR presence is not foreseeable. However, it is expected that for virtual object sizes greater than 50% larger than the physical object, significant worsening with respect to presence will also occur.

We did not correct for multiple comparisons, as this would have biased the results by increasing the ranges where no significant difference from the baseline condition M was detected due to the number of conditions to compare.

The study was performed under a fixed lighting condition chosen to make the overlays appear as opaque as possible. In reality, however, the lighting conditions are usually not as constant as in the laboratory. Since it can be assumed that the overlays are perceived differently under different lighting conditions, a next step was to investigate to what extent the selected lighting conditions have an influence on the results.

Conclusion

In this study we investigated if a physical proxy a user is interacting with can vary in size from its virtual representation in OST AR without strong negative effects. We examined the effect of size differences on the feeling of disturbance, the feeling of presence, size estimation and task completion time.

The results of the study show a clear tendency that it is possible to vary the size within certain ranges without too much worsening of disturbance and presence. It is therefore most likely possible to use one single physical object as a tangible prop to interact with several virtual objects of different sizes. The size variation range is wider for virtual objects larger than the physical object than it is for smaller virtual overlays. If no prior knowledge about better fitting objects exists, slightly smaller and larger (+/-10%) objects are often perceived as having the same size as the physical object. Moreover, the size variations investigated are unlikely to negatively affect performance compared to the baseline condition M.

The results obtained are similar to the results of VR and VST AR studies, which can be explained by the fact that the overlays were so opaque that the physical objects were almost blocked from view, since the study was performed in a very dimly lit room. Therefore, we decided to investigate what effect different, more natural, lighting conditions would have.

4.2.3 Influence of Environmental Lighting on Size Variations

The study on size variations (see Section 4.2.2) was conducted in a darkened room with very high opacity overlays. However, standard AR use cases take place under brighter ambient lighting. Therefore, an investigation of how the perception of differently sized virtual and physical objects changes under more realistic brighter illumination conditions was necessary.

As the brightness increases, the contrast of the virtual overlay decreases [27] and so the visibility of the physical object behind the virtual overlay changes. Due to the different visual perception under different lighting conditions, different results are to be expected here.

In a follow-up study, we therefore investigated the influence of illuminance on possible size variations between the virtual object and its physical representation. Under controlled lighting conditions, the effect of three different indoor illuminances on interaction with differently sized virtual objects was investigated by

evaluating how much the physical objects can deviate from their virtual representations in the respective lighting conditions without having a strong negative impact on presence, usability and performance.

This study and its results have already been published in [69]. The co-authors assisted in implementing and conducting the study, and in writing the paper.

Hypotheses

In the study conducted, we investigated to what extent environmental lighting has an impact on how much a physical object can deviate from its virtual counterpart without a strong negative impact.

As illuminance increases, the contrast of the overlays displayed in HoloLens 2 decreases [27]. We therefore assumed that the virtual objects displayed over the physical ones will appear to have varying degrees of transparency, and the perceived transparency will increase as the illuminance increases. As the increased transparency of the overlays makes the physical props behind them more visible, we expected the size estimation to be more accurate in brighter light. Similarly, we expected that the changed perception of the virtual object will lead to a difference in terms of possible size ranges in which the virtual and physical object can differ from each other without any significant degradation in the perception of presence, usability and performance. We therefore stated the following hypotheses:

- H1: With increasing environmental illumination, the perceived transparency of the overlays increases.
- H2: Size estimation is more accurate in brighter lighting conditions.
- H3: The ranges in which the size of the virtual and physical object can differ without worsening of presence, usability and performance vary for different lighting conditions.

Study Task

We defined a task to test the hypotheses stated above. In order to determine to what extent the selected lighting condition has an influence on the perception of OST AR proxy interactions when using differently sized virtual representations,

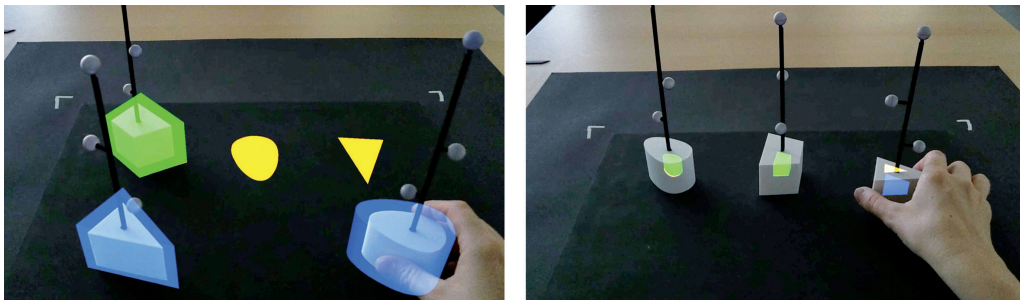


Figure 4.8: Screenshots of the HoloLens 2 during the execution of the study in size conditions XL (left) and XXS (right): matching the blue 3D overlays to the orange 2D targets. Visually the images correspond most closely to the perception in the medium lighting condition.

participants had to solve a puzzle task as quickly and as precisely as possible. They had to place and align three different virtual objects on associated virtual 2D target shapes (see Figure 4.8).

For each task, all three objects simultaneously had to be arranged two times each to generate the highest possible number of interactions, which is crucial for the evaluation of disturbance during grasping [85]. For this, the physical objects had to be lifted, rotated and moved to place them exactly, all of which are basic subtasks, but which have to be combined to solve more complex tasks [104].

Participants

24 participants (15 male, 9 female) aged between 21 and 55 ($M = 25.625$, $SD = 6.983$) took part in our study. Participants who were not associated with our institution received 15 Euro as compensation for participating in the experiment, which lasted about 90 minutes. All had normal or corrected-to-normal vision and happened to be right-handed. Prior experience with AR and AR glasses was rated on a 7-point Likert scale from 1 (= never) to 7 (= regular). Participants had low experience with AR ($M = 2.042$, $SD = 0.908$) and even less experience with AR glasses ($M = 1.5$, $SD = 0.722$).

Study Setup

The study took place in a darkened room. This ensured that the lighting conditions were the same for all participants at all times and were not influenced by



Figure 4.9: Illustrations of the laboratory setup in our three lighting conditions Dark, Medium and Bright (left to right). Studio lamps and lamps on the ceiling were used to adjust the environmental illuminance.

	Dark	Medium	Bright
Tabletop	10.75	49.5	257
HoloLens	4.4	14	45

Table 4.3: Lighting intensity in lx , measured on the tabletop pointing upwards and at the HoloLens camera pointing towards the interaction area in the different lighting conditions.

external factors. Depending on the lighting condition, the room was illuminated by 2 to 3 softbox studio lamps and the fluorescent tubes of the ceiling lighting.

Figure 4.9 shows the setup of the lamps for the respective lighting conditions. In condition Dark, only two studio lamps were used, pointing diagonally upwards away from the participant. In condition Medium, these lamps were turned towards the participant and a third studio lamp was installed, which was directed upwards to provide additional ambient light. In addition, the fluorescent tubes on the ceiling were switched on in condition Bright.

We measured the illumination on the tabletop facing upwards and from the HoloLens camera looking diagonally downwards onto the interaction surface. The measured luminance values are listed in Table 4.3.

The tracking of the HoloLens and the physical props was done with the help of six OptiTrack Prime^X 13 cameras, which were installed on a truss of about 4 m x 6 m and pointed towards the center of the tracking area. Participants sat at a table located at the center of the OptiTrack cameras.

For interaction, white physical props were used, which were equipped with a black marker tree for tracking (see Figure 4.10). Interaction was performed on a black plate on a black background, which was chosen to be large enough that the virtual objects were always visible against a black background.

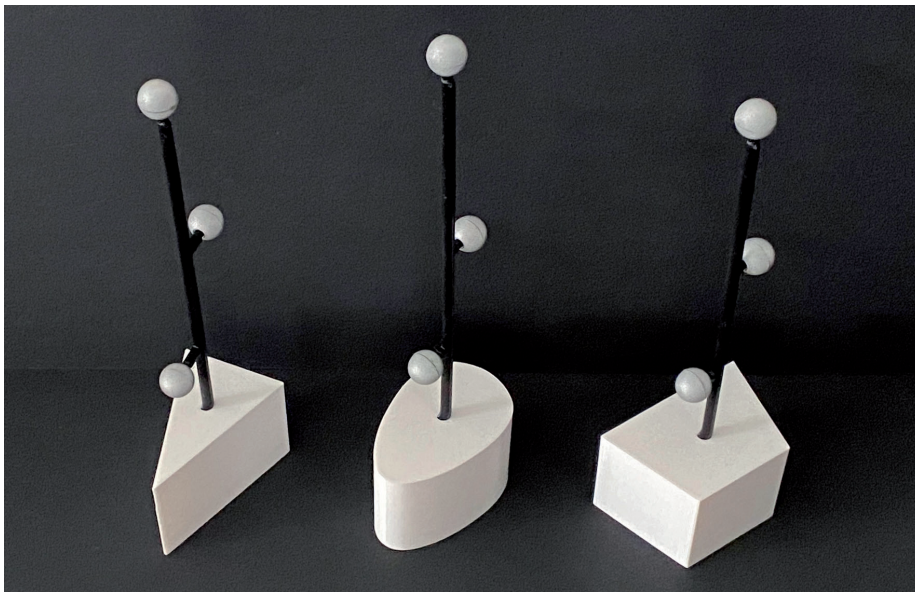


Figure 4.10: Physical props with attached marker trees on top. Triangular prism, egg-shaped cylinder and trapezoidal prism (left to right).

We used a HoloLens 2 for the visualization of the overlays, whose brightness we set to 100%. For the overlays, we chose an opacity of 100% and a medium-dark blue (#2300D1). In preliminary tests, we found that the white prop in combination with this blue overlay leads to different perceptions in each lighting condition, which would not have been possible with, e.g., a white overlay. We wanted the overlays to not be too bright in the dark, but still almost hide the physical objects. In the medium condition there should be a balance between the intensity of the overlay and the physical object, and in the bright condition the physical object should be in the focus. Figure 4.11 provides an indication of how the objects might have been perceived in the different lighting conditions.

We ensured that all participants had approximately the same viewing angle (approx. 45 degrees) on the props by placing the chair at a designated location and adjusting its height so that the distance between the HoloLens and the table was approximately 52 cm for each participant.

After finishing each task, participants had to complete three questionnaires in AR (see the paragraph *Tangible AR Questionnaire* in Section 5.2.2): an AR presence questionnaire, a TAR presence questionnaire, and a size perception questionnaire. The AR presence questionnaire consisted of four questions and evaluated how realistic the overlays in the respective task appeared and how strongly the participants felt that they were in an unaltered reality (see Section 4.1

and Appendix B.1.1). The TAR presence questionnaire, also consisting of four questions, evaluated how realistic the interaction with the virtual objects felt and how strong the feeling of interaction with the virtual overlays was while touching the differently sized physical props (see Section 4.1 and Appendix B.1.2).

In the size perception questionnaire, consisting of three questions, the size differences between the virtual and the physical object were evaluated, e.g., with respect to confusion during grasping (see Appendix B.2.1).

After completion of each lighting condition, a lighting questionnaire consisting of five questions had to be completed in addition to the other three questionnaires (see Appendix B.2.2). In this questionnaire, besides rating how transparent the overlays felt, the participants had to evaluate the naturalness of the environmental lighting in the just-experienced lighting condition. All four questionnaires were rated using 7-point Likert scales. When all tasks were completed, the participants received a final paper questionnaire asking for demographic information and additionally for a ranking of the lighting conditions (see Appendix B.3.2).

Design and Procedure

The study was designed as a within-subjects experiment. For three different lighting conditions we tested seven different size conditions each. The order of the lighting conditions was counterbalanced by a Williams design Latin square of size three [161]. We also balanced the order of the size conditions, which were presented as a block in each lighting condition.

For lighting conditions, we chose low illumination (condition Dark) similar to that in the size variations study (see Section 4.2.2), medium illumination (condition Medium), and high illumination (condition Bright).

We investigated the influence of each lighting condition on seven different size variations between the virtual and the physical object. Our baseline is represented by condition M, where the physical object is the same size as the virtual object.

Figure 4.11 visualizes the size differences between the virtual and physical objects in the individual lighting conditions. The views shown may differ from the perception in HoloLens 2, but provide an indication of how the objects might have been perceived. Since the participants' perception of the overlays is strongly affected by the environmental lighting, it is not useful to take screenshots with the HoloLens because the overlays would look the same in screenshots under all lighting conditions.

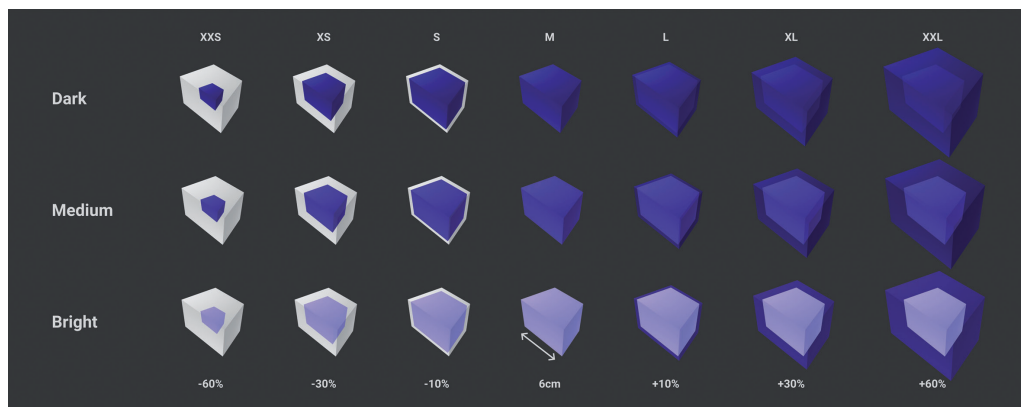


Figure 4.11: Size variations of the virtual overlays (blue) compared to the physical proxy objects (white). Condition M is the base condition with matching size of virtual and physical object. Condition XL, e.g., is obtained by scaling the virtual object by a factor of 1.3 along all three axes. The overlay opacities approximately reflect the participants' perception of the overlays in the three lighting conditions.

In addition to condition M, three smaller virtual overlays and three larger virtual overlays each were examined. Size conditions S and L represent a small size variance with a 10% difference, followed by XS and XL with a medium size variance of 30% and size conditions XXS and XXL with a large size variance of 60%. The XXL condition represents the largest possible size variance in which three objects can be interacted with simultaneously in the HoloLens 2 FOV without the overlays overlapping. Hence, we examined a larger size variance than in our previous study on size variations, in which a maximum size variance of 50% was not sufficient to determine upper limits for each measure.

We wanted to find out how much a physical object used for interaction can differ from its virtual counterpart without significantly degrading presence, usability, and performance. In addition, we wanted to determine whether these ranges change under different environmental illuminances.

For this purpose, we had participants interact simultaneously with three different physical objects. The triangular prism, the egg-shaped cylinder, and the trapezoidal prism (see Figure 4.10) had a width of 6 cm and a height and depth of 4 cm each. These shapes are taken from the study on size variations (see Section 4.2.2) and are based on existing work in VR and VST AR [21, 85] and chosen so that the objects are easy to grasp by hand [36].

Due to the design of the objects, there was only one way to correctly place objects on given targets, which at the same time required a maximum rotation of the

objects by up to 180°, so that the participants had to perform a maximum amount of interaction.

The task was to place the three virtual objects on their corresponding 2D targets. Each object had to be assigned twice to complete the task. At first, three targets were displayed in the upper part of the plate, on which the objects had to be placed. Once an object was correctly aligned, the color of its overlay changed to green (see Figure 4.8). After all objects in the top row were correctly placed, their overlay colors changed back to blue and the targets disappeared. Then new targets appeared in the lower area of the plate and the objects had to be assigned again. As soon as all objects in the bottom row were green, the task was considered completed and the timing stopped.

We placed the virtual objects so that their centers matched those of the physical objects. In order to be able to place the objects on the 2D targets, we adjusted the height of the targets in 3D space so that it matched the bottom of the virtual objects. Visually, however, from the participants' point of view, regardless of the size of the overlays, it always appeared as if the targets were on the tabletop.

There were six different possible arrangements for the targets in the upper area of the plate as well as in the lower area of the plate. We randomized at which position each target was displayed and randomly generated the rotations of each target. Likewise, a random initial arrangement of physical objects on the plate was performed.

We preliminarily determined suitable deviations in space that had to be achieved for an assignment to be considered fulfilled. As soon as the virtual object was, continuously for 0.5 seconds, less than 0.4 cm in flat distance and less than 0.5 cm in height away from the target position, and the angle between target and object was less than 3°, the object was considered correctly placed. We explicitly decided to use a stopping criterion in order to be able to perform meaningful time measurements. Since everyone evaluates accuracy differently, there would have been a strong impact on task completion time if everyone could have decided for themselves when the task was completed. Besides performance (task completion time), we measured usability (disturbance in grasping and interaction) and presence (AR presence and TAR presence) as well as the perception of lighting.

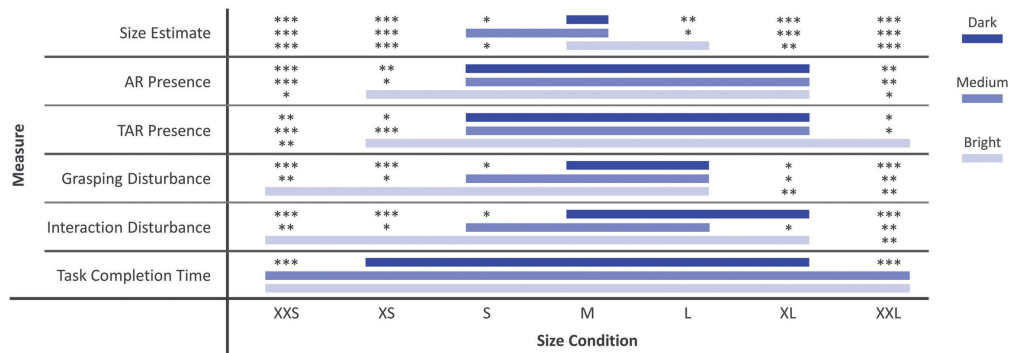


Figure 4.12: Summary of significant differences of the size conditions compared to the size-matching condition M as a baseline marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$) for all three lighting conditions. The blue bars indicate the ranges without significant difference in the respective lighting condition.

Results

We present the effects of different lighting conditions in the environment and size variations between corresponding virtual and physical objects on the aspects presented below. For each kind of measure, we first compared all samples collected in the three lighting conditions with each other by applying a Friedman test with two degrees of freedom and a significance level of $\alpha = 0.05$. Additionally, we report its test statistic χ^2 . As a post-hoc test, we used Wilcoxon’s signed-rank tests with 167 degrees of freedom and a significance level of $\alpha = 0.05$. Furthermore, we state test statistic W and the matched pairs rank-biserial correlation r as an effect size. Subsequently, we inspected each of the three lighting conditions individually to investigate the effect of the seven size variations in each specific lighting environment. For this, we used Friedman tests ($dof = 6, \alpha = 0.05$) as well as Wilcoxon’s signed-rank tests ($dof = 23, \alpha = 0.05$) to compare each size with condition M as a baseline. All of the Friedman tests showed a significant influence of the size condition in each lighting condition. For the applications of Wilcoxon’s signed rank test, we report the Bonferroni-corrected p -values. Figure 4.12 gives an overview of our results obtained using the post-hoc tests.

Overlay Transparency After each lighting condition, participants were asked to rate how transparent they perceived the overlays to be in the lighting questionnaire. The resulting mean transparency ratings for each lighting condition are displayed in Figure 4.13. The Friedman test ($dof = 2, \alpha = 0.05$) showed a significant influence of the lighting condition on the perceived transparency of

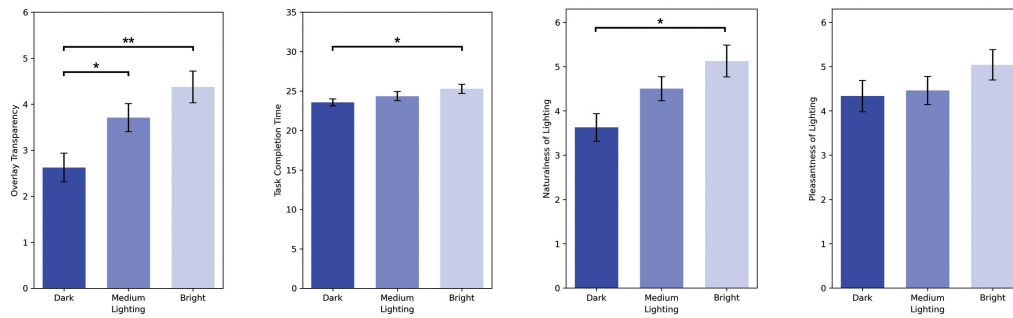


Figure 4.13: Mean transparency ratings regarding the overlays, mean task completion times in seconds and naturalness and pleasantness of the environmental lighting with marked standard errors for each lighting condition (left to right). Significant differences between conditions are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

the virtual overlays ($\chi^2 = 23.089, p < 0.001$) rated on a scale from 1 (not transparent at all) to 7 (completely transparent). The post-hoc test ($dof = 23, \alpha = 0.05$) confirmed this influence by revealing significantly lower transparency ratings in condition Dark compared to Medium ($W = 14, p = 0.015, r = -0.794$) and Bright ($W = 21, p = 0.002, r = -0.834$).

Results show that with brighter lighting the perceived overlay transparency increases, although the difference between medium and bright lighting is not statistically significant.

Size Estimate The Friedman test did not detect a statistically significant effect of the lighting condition on the overall size estimates participants gave when rating the size difference between the virtual and physical object on a range from 1 (= virtual much smaller), over 4 (= equal-sized) to 7 (= virtual much larger).

In the dark lighting condition, all size variations were rated significantly differently from the size-matching condition M. However, the estimates for condition S in medium lighting ($W = 19.5, p = 0.094, r = -0.675$) as well as condition L in bright lighting ($W = 30, p = 0.382, r = 0.5$) did not differ significantly from the baseline size M in the respective lighting conditions.

Therefore, only small size variations could not be differentiated significantly from the baseline M, a small size reduction S in medium lighting and a small size addition L in bright lighting.

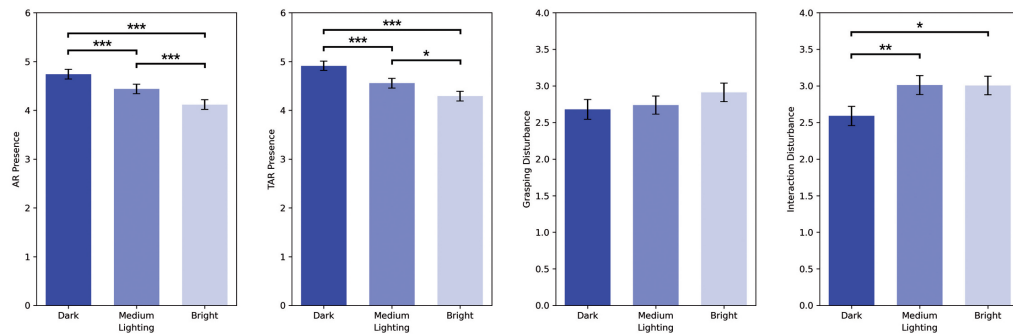


Figure 4.14: Mean presence and disturbance ratings from 1 to 7 with marked standard errors for each lighting condition. Significant differences between conditions are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

Presence The resulting mean presence ratings for each lighting condition are displayed in Figure 4.14. Starting with AR presence, the Friedman test shows a significant influence of the lighting condition on the ratings ($\chi^2 = 39.173, p < 0.001$). The post-hoc tests found that ratings in condition Dark were higher than in condition Medium ($W = 3333, p < 0.001, r = 0.387$), in Medium higher than in Bright ($W = 3759, p < 0.001, r = 0.345$), and therefore also in Dark higher than in Bright ($W = 2456.5, p < 0.001, r = 0.583$).

In the dark lighting condition, size conditions XXS ($W = 10, p < 0.001, r = -0.928$), XS ($W = 15, p = 0.005, r = -0.857$), and XXL ($W = 39, p = 0.005, r = -0.74$) led to significantly lower AR presence scores. Similarly, in medium lighting, sizes XXS ($W = 23, p < 0.001, r = -0.847$), XS ($W = 41.5, p = 0.036, r = -0.672$), and XXL ($W = 28.5, p = 0.001, r = -0.81$), and in bright lighting only sizes XXS ($W = 35.5, p = 0.019, r = -0.719$) and XXL ($W = 39, p = 0.049, r = -0.662$) showed negative effects.

Regarding TAR presence, the Friedman test shows a significant influence of the lighting in the environment on the ratings ($\chi^2 = 32.818, p < 0.001$). The post-hoc comparisons determined that ratings in condition Dark were higher than in condition Medium ($W = 3083, p < 0.001, r = 0.409$), in Medium higher than in Bright ($W = 4418.5, p = 0.015, r = 0.26$), and therefore also in Dark higher than in Bright ($W = 2836, p < 0.001, r = 0.512$).

In the dark lighting condition, size conditions XXS ($W = 19.5, p = 0.002, r = -0.859$), XS ($W = 36.5, p = 0.012, r = -0.736$) and XXL ($W = 46.5, p = 0.033, r = -0.663$) led to significantly lower TAR presence scores. Similarly, in the medium lighting condition, sizes XXS ($W = 19, p < 0.001, r = -0.873$),

XS ($W = 25, p < 0.001, r = -0.833$), and XXL ($W = 34, p = 0.016, r = -0.731$), and in bright lighting only size XXS ($W = 29, p = 0.01, r = -0.771$), significantly worsened presence.

AR and TAR presence behave very similarly to the extent that the darker the lighting condition, the higher the rated presence scores were. For dark and medium lighting conditions, we could observe significantly worse presence for very large (XXS) or large (XS) size reductions as well as very large size additions (XXL). However in the bright environment, size reductions would have to be very large (XXS) to cause such an effect on AR and TAR presence, while only the largest size addition (XXL) led to a significant decrease in AR presence. For TAR presence, we could not find an upper size deviation limit in the condition with bright lighting.

Usability The resulting mean disturbance ratings for each lighting condition are displayed in Figure 4.14. Starting with disturbance while grasping the objects, the Friedman test shows a significant influence of the lighting condition on the ratings ($\chi^2 = 7.281, p = 0.026$). However, the post-hoc tests could not find a significant difference between the scores in any of the conditions compared.

In the dark lighting condition, size conditions XXS ($W = 0, p < 0.001, r = 1$), XS ($W = 0, p < 0.001, r = 1$), S ($W = 0, p = 0.029, r = 1$), XL ($W = 11.5, p = 0.021, r = 0.831$), and XXL ($W = 0, p < 0.001, r = 1$) led to significantly higher grasping disturbance scores. Similarly, in medium lighting, sizes XXS ($W = 13.5, p = 0.006, r = 0.858$), XS ($W = 21, p = 0.029, r = 0.754$), XL ($W = 22, p = 0.018, r = 0.768$), and XXL ($W = 6, p = 0.001, r = 0.943$), and in bright lighting only sizes XL ($W = 8, p = 0.007, r = 0.895$) and XXL ($W = 21, p = 0.004, r = 0.834$), showed significantly increased ratings for disturbance during grasping.

Regarding disturbance during interaction with the objects, the Friedman test shows a significant influence of the lighting condition on the ratings ($\chi^2 = 15.825, p < 0.001$). The post-hoc tests determined that ratings in condition Dark were significantly lower than in condition Medium ($W = 2343, p = 0.002, r = -0.344$) and condition Bright ($W = 2438.5, p = 0.015, r = -0.294$).

In the dark lighting condition, size conditions XXS ($W = 0, p < 0.001, r = 1$), XS ($W = 0, p < 0.001, r = 1$), S ($W = 4, p = 0.021, r = 0.912$), and XXL ($W = 0, p < 0.001, r = 1$) led to significantly higher interaction disturbance scores. Furthermore, in medium lighting, size conditions XXS ($W = 3, p = 0.002, r = 0.965$), XS ($W = 23, p = 0.022, r = 0.758$), XL ($W = 16, p = 0.043, r = 0.765$), and

XXL ($W = 3, p = 0.001, r = 0.968$), and in bright lighting only size condition XXL ($W = 3, p = 0.003, r = 0.961$), showed significantly increased ratings.

Although we could not find a significant difference when comparing the lighting conditions to each other regarding disturbance during grasping, the results show clearly that the dark lighting condition leads to overall lower disturbance during interaction with the objects. But at the same time, in this dark lighting condition, introducing size differences between physical and virtual objects has a larger negative effect. Every size reduction (S, XS, and XXS) shows increased disturbances for both grasping and interacting, while sizes XL and XXL indicate this effect for grasping and only XXL for interacting, respectively.

For light condition Medium, both types of disturbance increase significantly for larger size deviations XXS and XS as well as XL and XXL. However, for the light condition Bright we could not find a significant worsening for smaller overlays; still, strong enlargements XL and XXL are significantly more distracting during grasping, while only XXL is more disturbing during interaction.

Performance The resulting mean time measurements for each lighting condition are displayed in Figure 4.13. The Friedman test showed a significant influence of the lighting condition on the measured task completion times ($\chi^2 = 7.429, p = 0.024$). Wilcoxon's signed-rank test revealed significantly smaller task completion times overall in Dark compared to Bright ($W = 5245, p = 0.01, r = -0.261$).

In the dark lighting condition, only size conditions XXS ($W = 23, p < 0.001, r = 0.847$) and XXL ($W = 17, p < 0.001, r = 0.887$) required significantly more time than the size-matching condition M, whereas in medium or bright lighting, none of the size conditions showed a significant deviation from baseline M.

So while the dark environment led to overall faster performance by the participants, very large size differences between the virtual and physical objects in conditions XXS and XXL have a significant negative effect compared to a matching object. These effects do not appear in the brighter light conditions.

Lighting In the lighting questionnaire, participants were also asked about how natural and how pleasant they rate the just-experienced environmental lighting to be. The Friedman test ($dof = 2, \alpha = 0.05$) showed a significant influence of the lighting condition on how natural it was rated ($\chi^2 = 11.195, p = 0.004$) to be. The post-hoc test ($dof = 23, \alpha = 0.05$) revealed that the environmental lighting

	Dark	Medium	Bright
Realism	63	50	31
Easiness	52	52	40
Preference	56	48	40

Table 4.4: Scores for each size condition in realism, easiness and preference according to participants' rankings.

was rated as significantly less natural in condition Dark than in Bright. There was no significant effect of the lighting conditions on the perceived pleasantness of the environmental lighting in the Friedman test, and likewise also not in the post-hoc tests.

Final Questionnaire After the study, participants finished with a concluding questionnaire where they were asked to rank the three lighting conditions based on the perceived realism and easiness while interacting with the objects and how much they liked the experience in the lighting condition. Table 4.4 shows the cumulative sum of the scores of all participants for the three conditions. The highest valued condition is given 3 points, the second 2 points, and finally the lowest valued condition 1 point each in the sum.

In this ranking, light condition Dark scored the most points regarding realism and preference, and equally many with condition Medium regarding easiness. Condition Bright scored lowest in every category. Sickness after the experiment was rated on a scale from 1 (= not at all) to 7 (= very sick) as quite low ($M = 2.0$, $SD = 1.383$).

Discussion

We hypothesized that environmental illuminance has an influence on the perception of virtual overlay transparency and that the overlays appear more translucent with increasing illuminance (H1). Our results support hypothesis H1. Figure 4.13 clearly shows the increase in perceived transparency with increasing brightness. A significant difference in transparency perception was found between dark lighting and medium lighting, and thus also between dark and bright lighting.

Since the physical objects are more visible with higher transparency of the virtual overlay, we hypothesized that this would also allow more accurate size estimation under brighter lighting conditions (H2). However, our findings do not support

hypothesis H2. The results of our study show that size estimation was best in low lighting. In medium lighting, however, condition S was often perceived as matching condition M in size, so that there was no significant difference. In bright lighting, on the other hand, there was no significant difference between size conditions L and M.

Furthermore, we hypothesized that ranges in which size variations are possible without significant degradation in the perception of presence, usability, and performance would differ for each lighting condition (H3). Hypothesis H3 was supported by our study results. Figure 4.12 shows the ranges in which size variations were possible without significant degradation in each lighting condition. With regard to the perceived presence, a slight increase of the ranges with increasing brightness can be seen. While a size variation from condition S to condition XL was already possible in low lighting, this range increases in bright lighting to XS to XL for AR presence and even XS to XXL for TAR presence. Therefore, with brighter environmental lighting, stronger size differences between the virtual and physical object seem possible without significant degradation of presence.

With regard to possible size variations without significant deterioration of usability (disturbance in grasping and interaction), there are also differences between the individual lighting conditions. While in the dark lighting condition only condition L was not perceived as significantly worse in terms of disturbance in grasping, in medium lighting this was already the case for conditions S and L. In bright lighting, the range without significant deterioration even increased, from condition XXS to condition L. As the physical objects were more visible as brightness increased, it is likely that this enabled participants to adjust grasping accordingly. For disturbance in interacting, the greatest difference is seen in bright lighting, where a variance between XXS and XL is possible, while in medium lighting only S to L and in dark lighting only M to XL is possible. Once the objects were grasped, the participants were probably aware of the size difference and could adjust to it. The results show that under brighter light conditions larger size variations are possible without significantly worsening the disturbance ratings.

In terms of performance, there were significant differences in low lighting for conditions XXS and XXL compared to baseline condition M. In contrast, under the two brighter lighting conditions, no differences in performance were detected between the individual size conditions and condition M. Overall, very large size differences between the virtual and physical object are therefore possible without having a strong negative impact on performance.

Comparing the overall ratings of the individual lighting conditions with each other, we see that the performance worsens with increasing environmental illumination (see Figure 4.13). Furthermore, with brighter illumination, AR presence and TAR presence deteriorate, and disturbance in grasping and interaction increases (see Figure 4.14).

For these reasons, the priority regarding the dark condition in terms of the ratings for realism, easiness and preference can likely be explained. The dark lighting condition is rated so well probably because the virtual overlays looked quite intense in the dark and there was little visual distraction from the physical objects or the participants' hands.

Even though darker lighting conditions were rated better, the brighter lighting conditions were more in line with indoor reality, as the environmental illuminance was evaluated as more natural in these lighting conditions. Therefore, when implementing real-life applications for indoor environments, minor losses in terms of presence and usability must be accepted in OST AR.

Overall, it can be seen that the range in which size differences between the virtual and physical object are acceptable increases with increasing brightness. The main surprise is that the acceptable size ranges are smallest in condition Dark. We had expected the size differences to have the greatest impact in the medium lighting condition, where they are most noticeable as the physical prop and the virtual object have roughly equally good visibility. However, size differences have the greatest influence in low lighting. Therefore, especially under brighter, more natural, indoor illumination levels, it is possible to use physical objects with a smaller or larger size compared to their virtual counterpart as tangible prop.

Limitations

In this study, we investigated the effect of environmental illuminance on the perception of size variations between a virtual overlay and its corresponding physical proxy object during interaction.

We selected three lighting conditions that were to influence the perception of the overlays to different degrees. However, we only considered artificial lighting conditions, without the influence of daylight. To ensure constant lighting conditions during the entire study, only a maximum illuminance of 257 lx (measured on the table upwards, or respectively 45 lx at the HoloLens pointing at the interaction area on the table) was possible. At higher illuminance levels, the results

would probably have been even more extreme, since the contrast of the HoloLens decreases significantly in this range according to [27].

Additionally, we only considered size variations between -60% and +60% of the baseline condition length. Interactions with larger virtual objects would not have been possible in the rather small FOV of the HoloLens 2. Due to the size constraint, not all upper or lower limits of size variations may have been detected. In addition, no precise limits have been established; to do so would require specific methods.

Conclusion

This study investigated the extent to which lighting conditions affect how size differences are perceived between a virtual object and its physical representation used for interaction. For this purpose, it has been examined, under three different indoor lighting conditions, to what extent the physical object can deviate in size from its virtual representation in the AR view without significantly degrading presence, usability, and performance.

The results show that the environmental illuminance influences the visual perception of the virtual objects. The virtual objects appear more transparent under brighter lighting conditions and thus make the physical object behind them appear clearer. This different visual perception also has an influence on the size range in which a physical object can differ from its virtual counterpart. In the dark condition, size variations are already possible within a certain range. With increasing brightness these ranges become larger, so that it is possible to work with even larger or smaller objects compared to the virtual object. However, the results also show that presence and usability decrease with increasing luminance, but this must be accepted for applications in realistic indoor lighting conditions.

4.2.4 Investigation of Shape Variations

Since a physical object used to control the virtual object may differ from it in more than just size, other factors must also be considered to determine the extent to which the objects may differ from each other. One crucial factor here is the shape. It must be determined whether the physical and virtual objects must be identical in shape or whether an abstraction of the virtual object's shape can also be used as a tangible prop. This would allow one physical prop to be used for a variety of virtual objects in different use cases.

We therefore investigated in a further study whether it is possible to use a physical object for interaction that differs in shape from its virtual representation without having a strong negative impact on presence, usability and performance.

This study and its results have already been summarized in a paper and published as arXiv preprint [65]. The aim is to submit them also to one of the next HCI conferences. The co-authors helped to implement and to conduct the study and assisted in writing the paper.

Hypotheses

Since the studies on size differences have shown that small size differences are hardly perceived and therefore have little impact on usability (disturbance ratings), presence (AR/TAR presence ratings), and performance (task completion time), we assumed that this would also be the case for very small shape variations, but that there would be strong negative effects above a certain degree of difference. Similarly, we assumed that there could be a strong deterioration in performance if there were very large differences in shape between the virtual and the physical object. We therefore hypothesized the following:

- H1: Shape differences between virtual and physical objects cannot be properly detected for very similar shapes.
- H2: The shape of the virtual and physical object can differ within a certain range without significantly degrading “AR presence” and “TAR presence”.
- H3: The shape of the virtual and physical object can differ within a certain range without significantly worsening usability.
- H4: The level of difference between the shape of the virtual and physical object has an impact on performance.

Study Tasks

To test the above hypotheses, we defined two study parts in which a virtual object was represented by various physical props that served to interact with it. In each part of the study, the tasks had to be solved as quickly and as precisely as possible. One part of the study consisted of a 2D puzzle task. Here, the virtual 3D objects had to be independently adjusted to the respective matching 3D target

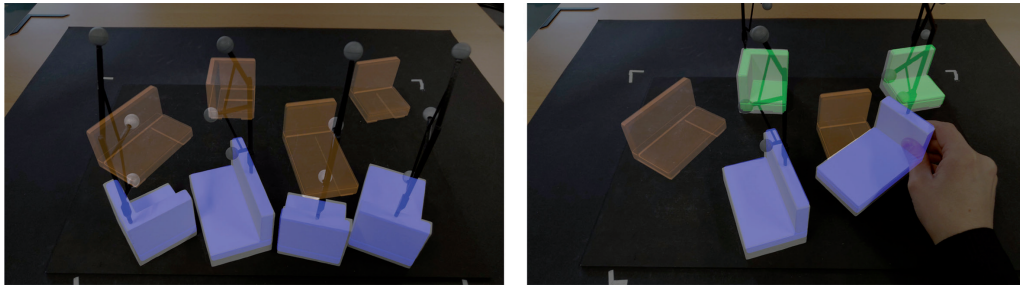


Figure 4.15: HoloLens screenshots of study part 1 in condition *B*: Matching virtual sofa parts (blue) to appropriate 3D targets (orange).

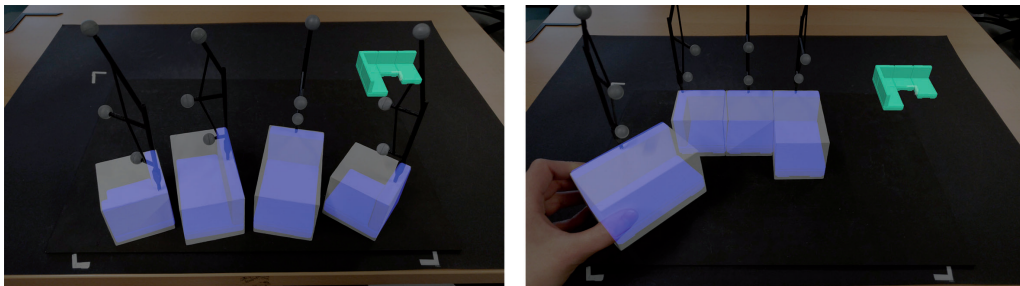


Figure 4.16: HoloLens screenshots of study part 2 in condition *D*: Assembly of the virtual sofa parts (blue) according to a 3D miniature template (turquoise).

objects, which were displayed at an appropriate distance as an AR overlay on the workspace (see Figure 4.15). The other part of the study was to assemble the individual virtual objects into a composite object anywhere on the workspace based on a given miniature 3D model displayed in the upper right corner of the AR view (see Figure 4.16).

In both parts of the study, we made sure to include interactions with multiple objects so that the props would need to be grasped and moved more often, for example, to better measure the influence of the shape of the physical object when grasping [85]. All tasks consisted of simple subtasks, such as grasping, lifting, moving, and placing, all of which are necessary to solve high-level tasks [104].

Participants

20 study participants were recruited (16 male, 4 female) in an age range from 19 to 51 years ($M = 26.95$, $SD = 8.204$). Those outside our institution received 10 Euro for their participation. 19 participants were right-handed, one was left-handed, and all of them had normal or corrected-to-normal vision. On a 7-point Likert scale, participants reported relatively low experience with AR ($M = 2.35$, $SD =$

1.309) and very low experience with AR glasses ($M = 1.25$, $SD = 0.444$) where 1 means they had never used such systems and 7 means they used them regularly.

Study Setup

The study took place in a darkened room to ensure that lighting conditions during the study were not influenced by external factors (such as sunlight) and that the same lighting conditions prevailed for all participants. A medium brightness was chosen, which was perceived as natural by the participants according to the results of the lighting variation study, but which was not too bright, as the overlays were perceived as more comfortable by the participants in darker lighting conditions.

The room was illuminated with the fluorescent tubes of the ceiling lighting. The measured luminance on the tabletop facing upwards was $98.5 lx$ and from the HoloLens camera perspective looking diagonally downwards onto the interaction surface it was $14.8lx$.

10 OptiTrack Prime^X 13 cameras were used to track the physical props and the HoloLens. These were installed on a truss with a dimension of about 4 m x 6 m and directed to the center of the tracking area. The table where the participants sat while working on the tasks was located in the center of this area.

The tasks were performed on a black plate on a black background, which was large enough so that both the physical and the virtual objects were always visible against a black background.

The objects interacted with were white with black marker trees for tracking (see Figure 4.17). The HoloLens 2 used for the study was set to maximum brightness and the opacity of the overlays was set to 100%. For the color of the overlays we chose a medium-dark blue (#001FFF), since the pre-test showed that under the given lighting conditions, the virtual and physical objects are visible equally well with this color choice. Participants were placed at the table so that they all had approximately the same viewing angle (approx. 45 degrees) of the objects they had to interact with to solve the tasks.

Whenever a task was solved, participants had to complete three questionnaires: an AR presence questionnaire, a TAR presence questionnaire, and a shape perception questionnaire (see Appendix B). These questionnaires were answered in AR using a tangible proxy object that served as a pen (see the paragraph *Tangible AR Questionnaire* in Section 5.2.2). The AR and TAR presence questionnaires



Figure 4.17: Two-seater in shape condition *A* with attached marker tree.

were the same as used in the study on lighting variations (see Section 4.2.3). The shape perception questionnaire consisted of three questions. It assessed the perceived shape differences between the virtual and the physical object as well as the disturbance during grasping and interaction (see Appendix B.2.3).

The questionnaires were all rated using 7-point Likert scales. After finishing all tasks, the participants received a final paper questionnaire asking for demographic information and additionally for a ranking of the different physical shapes (see Appendix B.3.3).

Design and Procedure

The study was designed as a within-subjects experiment. In total, five different shape conditions were tested. The order of the shape conditions in both parts of the study was counterbalanced by a Williams design Latin square of size five [161]. Half of the participants started with the first part of the study and the other half with the second part of the study. Figure 4.18 shows the shape variations of the physical proxy object. Condition *A* represents the baseline where the physical object is an exact replica (3D model) of the virtual object. Condition *B* is an abstraction of the 3D model, where seat height and seat depth match those of

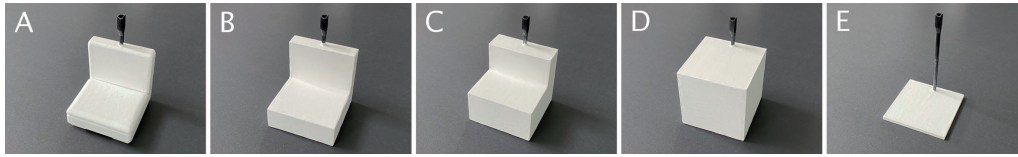


Figure 4.18: Investigated shape variations of the physical proxy object: matching 3D model (A), abstracted 3D model (B), abstraction of a standard sofa (C), cuboid (D), plane (E).

	3D Model <i>A</i>	Abstracted 3D Model <i>B</i>	Abstraction of Standard Sofa <i>C</i>	Cuboid <i>D</i>	Plane <i>E</i>
Seat Height	2.1	2.1	3.2	6.0	0.3
Backrest Depth	1.2	1.2	2.0	6.0	0.0

Table 4.5: Dimensions of seat heights and backrest depths of the different shape variations measured in centimeters.

the virtual object. Condition *C* is an abstraction of a standard sofa. To determine the standard dimensions, the seat heights and seat depths of all available IKEA¹³ sofas were averaged (status as of 2020-01-18). Condition *D* (Cuboid) corresponds to the bounding box of the sofa part and Condition *E* to the base area. Except for Condition *E* – because it is flat – the external dimensions correspond to those of the virtual object, so that only the varied shapes have an influence on the results.

In the study, participants had to interact with four different sofa parts: a one-seater, a two-seater, a corner piece and a chaise longue. We chose a height, width and depth of 6 cm as the size for the one-seater and the corner piece, so that the objects are easy to grasp by hand [36]. The two-seater and the chaise longue differed in that the two-seater was 10 cm wide and the chaise longue was 10 cm deep (see Figure 4.19). The seat heights and backrest depths of the different abstractions can be found in Table 4.5.

The proxy objects for abstractions *B*, *C*, *D* and *E* were laser-cut from MDF boards for environmental reasons. They were assembled into the sofa pieces and then painted several times with white paint. The 3D models of abstraction *A* were 3D printed and afterwards painted several times with the same white paint to guarantee the same feeling when touching the objects. In addition to the surface, care was also taken to ensure that the individual sofa parts had almost the same weight (+/- 1 gram) in all abstraction conditions (except condition *E*), as well as a

¹³<https://www.ikea.com/de/de/> (last accessed: 2023-08-15)

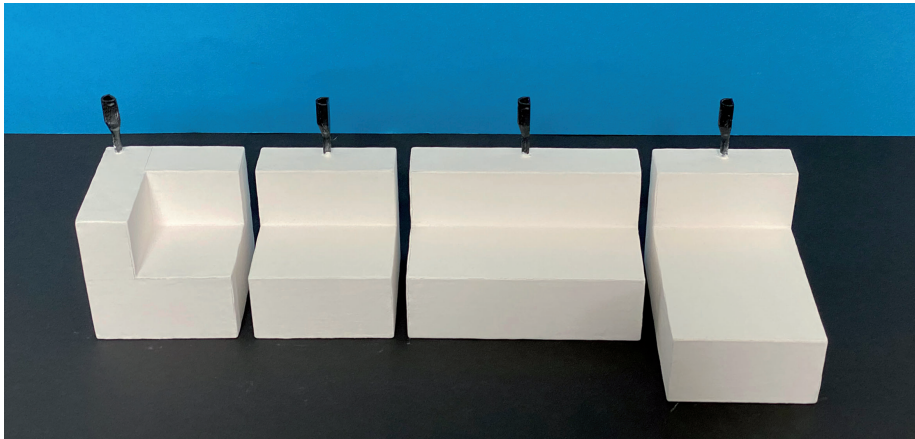


Figure 4.19: Sofa parts used in the study. From left to right: corner piece, one-seater, two-seater and chaise longue in shape condition C.

similar weight distribution (main weight in the lower back area; except conditions *D* and *E*). All sofa parts were equipped with adapters on which the appropriate marker trees could be mounted. In this way, only four different marker trees with three reflective markers each were needed. For each physical sofa part, a fine calibration of the corresponding virtual object was performed in Unity¹⁴ to ensure an accurate overlay of the virtual object. A total of 20 different sofa parts was produced in this way, with which the participants interacted during the study.

In our study we wanted to investigate how much the shape of the physical object used for interaction can differ from its virtual representation without a significant worsening of presence, usability, and performance. We also wanted to find out whether it makes a difference to arrange the sofa parts individually or to combine them into a composite object.

Part 1 In part 1 of the study, the sofa parts had to be arranged individually. Virtual 3D targets to which the virtual overlays of the physical sofa parts had to be matched were displayed at four predefined positions. The physical objects were randomly arranged at the bottom of the plate. Likewise, the mapping of the virtual sofa parts to the four positions as well as the rotation of the virtual target objects was randomized. To ensure that the rotation of the targets was always clearly visible without the need to move one's head, care was taken to ensure that the sofa parts were never shown predominantly from behind. Therefore, we

¹⁴<https://unity.com> (last accessed: 2023-08-15)

excluded a direct view of the back as well as rotations up to 60 degrees in any direction from it.

In all five shape conditions, the four different objects had to be interacted with. The task was to align all four virtual overlays with the displayed 3D targets (see Figure 4.15). The order in which the objects could be assigned was flexible. As soon as an object was aligned precisely enough, its overlay turned green. The task was completed as soon as all overlays were green. An object was considered correctly aligned if errors below a threshold of 4 mm in horizontal distance, 5 mm in vertical distance, and 3° in the angle between virtual overlay and virtual target were detected constantly for 0.5 seconds. Pre-tests showed that the choice of these values made the task sufficiently complex, but still feasible without too much effort.

Part 2 In the second part of the study, the four sofa pieces had to be put together to form a specific sofa combination. Also in this task, the physical objects were randomly arranged at the bottom of the plate at the beginning of each trial and randomly rotated. The sofa combination that had to be recreated was displayed as a virtual 3D object in small format in the upper right area on the plate (see Figure 4.16). It could be built anywhere on the plate and rotated as desired. Once the sofa combination was correctly assembled, all overlays turned green at the same time and the task was completed. Two individual elements were considered to be correctly assembled if they were connected with the correct sides and these had a maximum distance of 2 mm from each other and were at an angle of less than 7.5° to each other. In addition, the offset of the contact edges had to be less than 6 mm when aligned parallel to each other. As soon as these conditions were no longer met for 0.1 seconds, the connection between the parts was released. We also determined these values in pre-tests to ensure that the task would not be too easy to solve, but also would not cause frustration among the participants.

Results

We investigated the effects of shape variations between the tangible and the corresponding virtual object on each of the measured dependent variables reported below with a uniform procedure. First, we applied a Friedman test with four degrees of freedom and a significance level of $\alpha = 0.05$ to compare the samples collected in the five shape conditions with each other, separately for part 1 and part 2 of the study. We report the p -value along with the test statistic

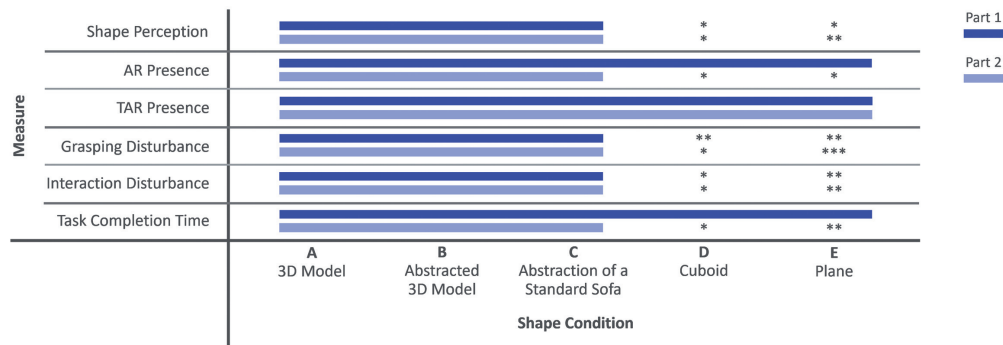


Figure 4.20: Summary of significant differences of the shape conditions compared to the shape-matching condition A as a baseline marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$) for the two study parts. The blue bars indicate the ranges without significant difference.

χ^2 . If a significant influence of the shape condition on the dependent variable was detected, we used Wilcoxon’s signed-rank tests ($dof = 19, \alpha = 0.05$) to compare each shape abstraction to condition A (3D model) as a baseline. We report Bonferroni-corrected p -values, the test statistic W and the matched pairs rank-biserial correlation r as an effect size. Figure 4.20 shows an overview of our results obtained with the post-hoc tests. Except for task completion time, no significant difference was found between part 1 and part 2 of the study when comparing the measures.

Shape Perception The rating of to what extent the shapes of the virtual and physical objects matched was influenced significantly by which shape abstraction was used for both parts of the study. The Friedman test confirmed this for part 1 ($\chi^2 = 18.006, p = 0.001$) and 2 ($\chi^2 = 34.183, p < 0.001$). Wilcoxon’s signed-rank test revealed that in part 1 of the study, conditions C (abstraction of a standard sofa) ($W = 48, p = 0.399, r = -0.439$) and B (abstracted 3D model) ($W = 46.5, p = 1.0, r = -0.114$) did not differ significantly from the shape-matching condition A, while abstractions E (plane) ($W = 20, p = 0.018, r = -0.766$) and D (cuboid) ($W = 24, p = 0.029, r = -0.719$) were found to match significantly less well to the virtual shape. Equally, in part 2, conditions C ($W = 22, p = 0.664, r = -0.436$) and B ($W = 21.5, p = 1.0, r = -0.218$) did not show a significant difference to A, while conditions E ($W = 6, p = 0.002, r = -0.93$) and D ($W = 24, p = 0.03, r = -0.719$) did. The results show that in our study setup slight differences of the proxy object compared to the virtual object were often not noticed.

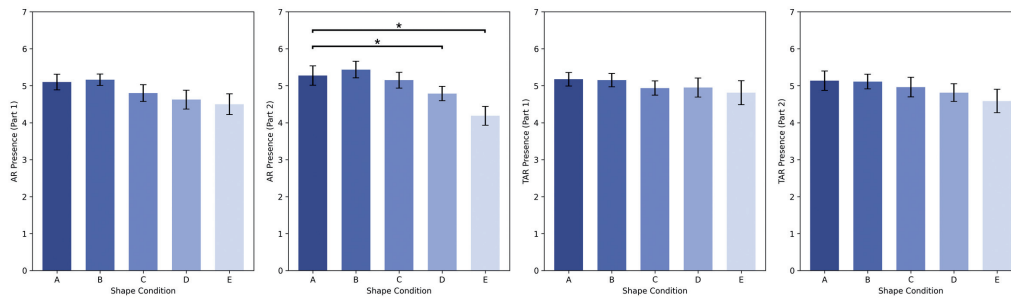


Figure 4.21: Mean AR presence and TAR presence ratings from 1 to 7 with marked standard errors for each shape condition. Significant differences from the baseline condition *A* are marked with * ($p < 0.05$).

Presence According to the Friedman tests, AR presence scores are significantly influenced by the shape condition only in part 2 ($\chi^2 = 27.04, p < 0.001$), not in part 1 ($\chi^2 = 3.024, p = 0.554$). Wilcoxon's signed-rank test revealed for part 2 that both conditions *E* ($W = 19, p = 0.028, r = -0.752$) and *D* ($W = 31, p = 0.041, r = -0.674$) led to significantly lower ratings than the baseline *A* in AR presence (see Figure 4.21). Regarding TAR presence, a significant influence of the shape condition could be detected neither for part 1 nor for part 2 of the study.

Disturbances Regarding disturbance during grasping the objects, the Friedman test revealed significant effects both for study part 1 ($\chi^2 = 36.105, p < 0.001$) and 2 ($\chi^2 = 50.119, p < 0.001$). The post-hoc test then showed that conditions *E* ($W = 0, p = 0.003, r = 1.0$) and *D* ($W = 5, p = 0.009, r = 0.905$) had significantly higher values of disturbance during grasping than the baseline *A* in part 1. For part 2, the results are similar with only conditions *E* ($W = 0, p < 0.001, r = 1.0$) and *D* ($W = 8, p = 0.035, r = 0.824$) standing out with higher ratings. Disturbance during interaction with the objects behaves similarly. In part 1 ($\chi^2 = 27.594, p < 0.001$) and part 2 ($\chi^2 = 37.407, p < 0.001$), a significant influence of shape condition on the ratings was found. Again for part 1, only conditions *E* ($W = 8, p = 0.003, r = 0.906$) and *D* ($W = 18, p = 0.031, r = 0.735$) and for part 2, likewise conditions *E* ($W = 0, p = 0.001, r = 1.0$) and *D* ($W = 8.5, p = 0.022, r = 0.838$) were rated significantly higher in disturbance during interaction than condition *A* (see Figure 4.22). Overall, the results show a tendency for the disturbance ratings to increase with increasing level of abstraction and that small differences between the virtual and physical object are tolerated.

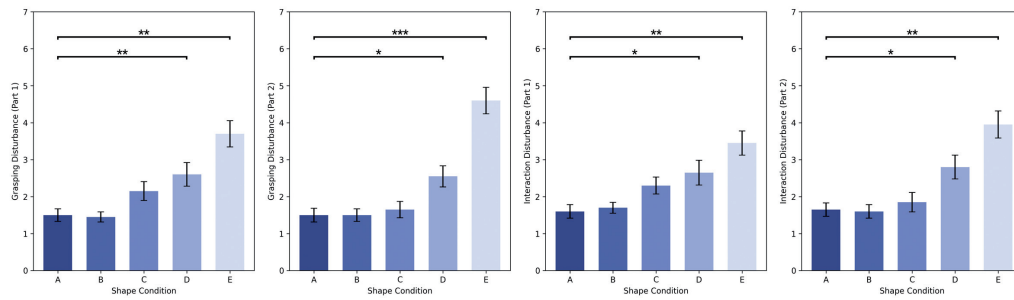


Figure 4.22: Mean disturbance ratings from 1 to 7 with marked standard errors for each shape condition. Significant differences from the baseline condition *A* are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

Task Completion Time The Friedman test indicates a significant influence of the shape condition on the measured task completion times only in study part 2 ($\chi^2 = 18.08, p = 0.001$), not in part 1 ($\chi^2 = 6.8, p = 0.147$). Investigating part 2 in more detail, both conditions *E* ($W = 21, p = 0.003, r = 0.8$) and *D* ($W = 28, p = 0.011, r = 0.733$) showed a significantly increased task completion time compared to baseline condition *A* in the post-hoc tests. These results imply that at least slightly abstracted proxy objects can be used without significantly degrading performance.

Final Questionnaire The study participation was finished by completing a final questionnaire where participants ranked the five shape abstractions based on three criteria: feeling of realism, easiness, and overall preference. For each ranking, the condition with the highest score was given 5 points, the second ranked condition was given 4 points, and so on until the last condition received 1 point. With this approach, a sum of scores was computed for each shape condition which is reported in Table 4.6. In all three categories, there is a clear trend that the position in the ranking worsens as the level of abstraction increases.

Discussion

We hypothesized that shape differences between the virtual and physical object cannot be properly detected for very similar shapes (H1). Figure 4.20 shows that for both parts of the study conditions *B* and *C* were not rated significantly differently compared to the baseline condition *A* in our study setup, while conditions *D* and *E* were. This shows that many participants did not perceive small differences between the physical and the virtual objects, supporting our hypothesis H1.

	3D Model	Abstracted 3D Model	Abstraction of Standard Sofa	Cuboid	Plane
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
Realism	91	80	61	39	29
Easiness	83	73	61	53	30
Preference	82	71	59	54	34

Table 4.6: Summed scores for each shape condition regarding realism, easiness and preference obtained by the rankings of all participants.

We expected that small shape differences between the physical and virtual object would not be perceived. Therefore, we also assumed that the shape of the physical object could be varied to a certain degree without significantly worsening the AR presence and the TAR presence (H2). AR presence worsened significantly only in part 2 for the levels of abstraction *D* and *E*. In part 1, however, no significant difference in AR presence was found for abstractions *B* to *E* compared to the baseline condition *A*. In Figure 4.21, a slight trend can be seen in part 1 that the presence decreases with increasing degree of abstraction. In our chosen study setup, however, no significant deterioration was detectable. Likewise, the TAR presence scores only slightly decrease with increasing abstraction level (see Figure 4.21). For none of the shape conditions could a significantly lower TAR presence, compared to the baseline condition *A*, be determined. The results suggest that it is possible to use an object simplified in shape as a tangible proxy to manipulate a virtual object without significant degradation in presence, thus supporting our hypothesis H2.

We also hypothesized that the shape of the virtual and physical object can differ within a certain range without significantly worsening usability (H3). Figure 4.20 shows that in our study setup the shape could be abstracted up to the condition *C* without a significant worsening of the grasp and interaction disturbance ratings. The results show a clear trend that disturbance increases with increasing abstraction of the proxy object and that small abstractions are possible without significantly degrading usability, which strengthens our hypothesis H3.

In addition to shape perception, presence and disturbance, we also evaluated the influence of the shape condition on the task completion time for both parts of the study separately. Again, we assumed that shape differences between the physical and virtual object are possible to a certain degree without a significant deterioration of the task completion time, but that the choice of the degree

of abstraction has an influence on the performance (H4). Hypothesis H4 was partially supported by our results. While for study part 1 no significant influence of the shape condition on the task completion time could be found, an influence was found for part 2. Here, the completion time worsened significantly starting from the abstraction level *D* compared to the baseline condition *A*. This implies that small differences in the shape compared to the virtual overlay are possible without significantly worsening the task completion time. Study part 2, where the assembled sofa combination had to be built, was more difficult for the participants when using the highly abstracted shapes *D* and *E*, which is reflected in the time taken to solve the task. One assumption for this is that with the physical props whose shapes still made it possible to recognize and feel their orientation, it was possible to assemble the sofa combination without having to concentrate intensively on the virtual overlays. This is particularly the case because a sofa is an everyday object, where it is clear in which orientation two parts are logically connected to each other.

Overall, the results suggest that it is possible to use objects with abstracted shapes as physical representations of different virtual objects. However, a too-strong abstraction, which does not allow users to recognize the original shape anymore, leads to significant deteriorations regarding the perceived presence, usability and – depending on the task – even performance.

Limitations

We investigated the extent to which it is possible to use an object with a different shape as a tangible for the manipulation of a virtual object. The results indicate that at least small abstractions of the shape are possible without significantly degrading performance, presence, and usability. Currently, the results are only based on investigations with sofa elements, which were abstracted to a cuboid and further to a plate. These results have to be consolidated by further investigations with other different objects, which are abstracted to other basic shapes, e.g. a sphere or a cylinder.

In addition, it is not possible to determine from the results exactly to what degree an abstraction is possible. This can only be determined with the help of special procedures, such as the up/down staircase procedure [35]. However, as in this study, the results would depend on the self-defined levels of abstraction.

Conclusion

In this study, we investigated the extent to which a physical object used as a tangible for interaction with a virtual object can deviate in shape from its virtual counterpart. For this purpose we determined the influence of five different shape variations on presence, usability and performance. The baseline condition represented a detailed replica in the form of a 3D print of the virtual object, which was abstracted in defined stages up to the base area of the virtual object.

The results imply that it is possible to use a simplified proxy object for interaction without significantly degrading performance, presence, and usability. In our study setup, this was the case up to the condition “abstraction of a standard sofa”, where the shape was still at least roughly similar to that of the virtual object. Especially for the assembly task, the shape conditions “cuboid” and “plane” performed significantly worse than the baseline condition.

In our study, we solely examined the influence of the shape variation by keeping the size factor the same for all conditions. An exception is the height in the condition “plane”, since this is a flat object. If a set of acceptable physical objects is to be chosen to represent a multitude of virtual objects, then also the interplay between size and shape of the proxy object will play a crucial role, which needs to be examined in detail.

4.3 Summary

In this chapter, we first explained how we determined presence in AR using self-designed questionnaires adapted for this purpose. We then presented the study design and the general procedure applied during our studies. In different studies, we investigated how much a physical proxy object used for interaction can differ from its virtual counterpart in shape and size, and how the environmental lighting influences possible size deviations.

Our results indicate that for interaction with virtual objects, physical objects can be used that vary in size to some degree. This variation range is wider for virtual objects larger than the physical object. Similarly, we found that environmental illuminance affects the perception of objects. In brighter lighting conditions, virtual objects are perceived more transparently in OST AR. In addition, brighter lighting increased the range in which the objects could deviate in size without significantly degrading the presence and usability ratings. The usability and

presence ratings are overall worse in brighter lighting conditions. However, this cannot be avoided when working under natural lighting conditions. The results also suggest that an abstracted form of the virtual object is suitable as a proxy object. However, this deviation must not be too large; the shape should at least roughly correspond to that of the virtual object.

Chapter 5

Frameworks for Controlling Visual and Haptic Perception

In the context of this dissertation, two frameworks were developed that make it possible to influence visual and haptic perception, respectively. By controlling the perception in a targeted way, it is possible to measure the corresponding influence on defined metrics, such as usability or performance. This is necessary to perform the studies described in Chapter 3 and Chapter 4.

The first framework developed in this context enables the control and inspection of a user's gaze direction through visual cue stimuli in an instrumented environment (see Section 5.1). The second framework enables the visualization of virtual overlays in AR glasses as overlays on physical proxy objects used for interaction, as well as their evaluation (see Section 5.2).

5.1 Framework for Adaptive Gaze Guidance

The framework for adaptive gaze guidance enables the personalized guidance of the gaze direction of people in instrumented environments. By selectively triggering actuators, visual cues can be displayed at predefined locations. By means of eye tracking, the current direction of the person's gaze can be determined. In this way, this framework makes it possible to measure the influence of different cues in a real environment on people's gaze direction.

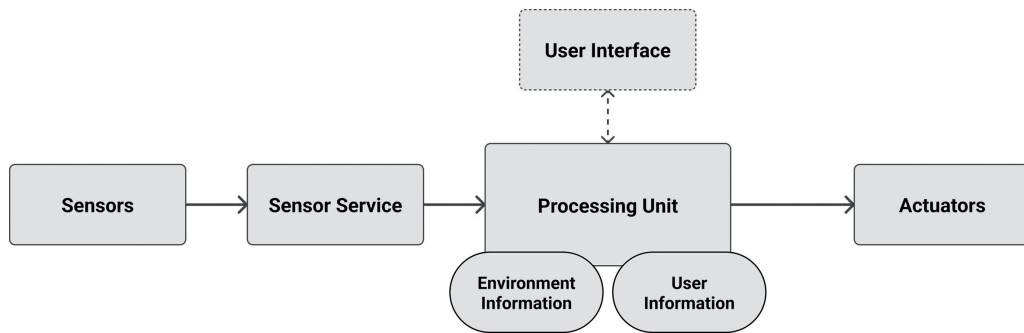


Figure 5.1: Main components of the framework for adaptive gaze guidance.

In the following, we first introduce the concept of the framework and then demonstrate the framework using the prototypical implementation we used for the studies in Section 3.2.2.

5.1.1 Concept

The gaze direction framework consists of 4 main components that can be extended with an optional user interface (see Figure 5.1). The core component of the framework is the processing unit, which triggers appropriate cue stimuli on suitable actuators based on environmental and user information and taking into account current sensor information. In the following, the individual components of the framework are presented.

Sensors

Sensors are used to determine the direction of the user's gaze. A very precise statement about where someone is looking can be made using eye trackers. These can be stationary in the instrumented environment or mobile in the form of eye tracking glasses. In addition to eye tracking, it is also possible to determine where someone is likely to be looking based on the orientation of their head. The approximate head pose can be determined inexpensively, e.g. with the help of depth cameras, or an exact orientation can be achieved with expensive motion capturing systems.

Sensor Service

The sensors are connected to the Sensor Service, which receives the raw data from the sensors. This data is preprocessed by the Sensor Service into a self-defined but consistent format and then forwarded to the Processing Unit.

Actuators

Actuators installed in the environment are used to guide the direction of gaze. To direct visual attention, auditory and visual cues are most suitable; these can be triggered by a wide variety of actuators. For example, a flashing LED on an object, or an object illuminated by a lamp, can attract attention. Likewise, playing sounds through loudspeakers at appropriate positions in the environment can target visual attention to objects. In the studies we conducted (see Section 3.2.2), we only investigated the influence of visual cue stimuli on gaze direction.

Processing Unit

The Processing Unit is the core of the framework and decides when, where and which information stimulus is displayed. Among other things, it has access to the environment information. This means that it knows exactly the locations and types of actuators in the environment and which different cues they can send out. In addition to the environment information, the Processing Unit also has knowledge of the user information. On the one hand, it knows which objects in the room are of interest to the respective user. On the other hand, it knows the user's preferred cues, or even which cues work particularly well for the user. This information enables the system to adapt optimally to the respective user. The Processing Unit represents the control system that determines the appropriate stimulus based on the gaze direction data it receives from the Sensor Service, as well as the user data and the environmental data, and instructs the appropriate actuator to display the respective stimulus. Additionally, the Processing Unit is able to evaluate and learn the reactions to sent-out cues, so that the system can improve over time.

User Interface

The User Interface is an optional element of the gaze direction framework. It can be used to control the actuators in the environment. For example, the environ-

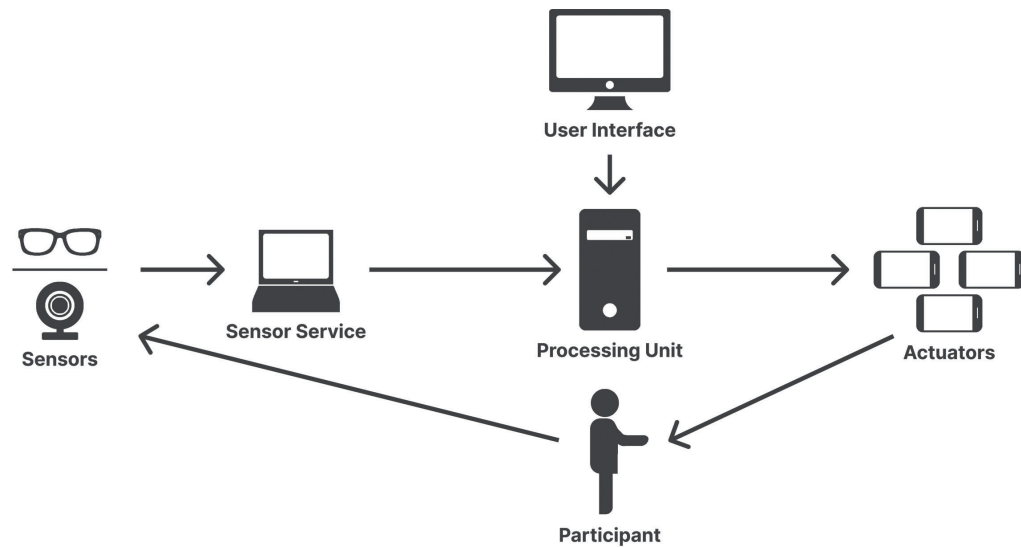


Figure 5.2: Detailed architecture of the framework for adaptive gaze guidance.

ment can be displayed together with its actuators and specific commands can be sent to individual or multiple actuators.

5.1.2 Prototypical Implementation

To perform the studies described in Section 3.2.2, the concept of the adaptive gaze direction guidance framework described in Section 5.1.1 was implemented prototypically. Figure 5.2 visualizes the relationships between the different implemented components. The sensors we used in our studies to determine gaze direction were the mobile eye tracker Pupil Pro [72] to measure exactly where people were looking (see Figure 5.3, left), and the OptiTrack motion capturing system to determine head orientation for approximate gaze direction (see Figure 5.3, right). For the studies, a shopping shelf with 16 product compartments was built and instrumented accordingly (see Figure 5.4). We used Android smartphones as actuators, which also functioned as digital price tags. The Processing Unit represented the server in the framework and was implemented in Java. The communication between the sensors and the Processing Unit as well as between the actuators and the Processing Unit was done with the Event Broadcasting Service (EBS) of Kahl et al. [70] via UDP. The environment information required for the studies as well as participant information was provided as xml files. A front end was implemented, which made it possible to send single or connected visual cue stimuli to specified actuators and to start study runs.

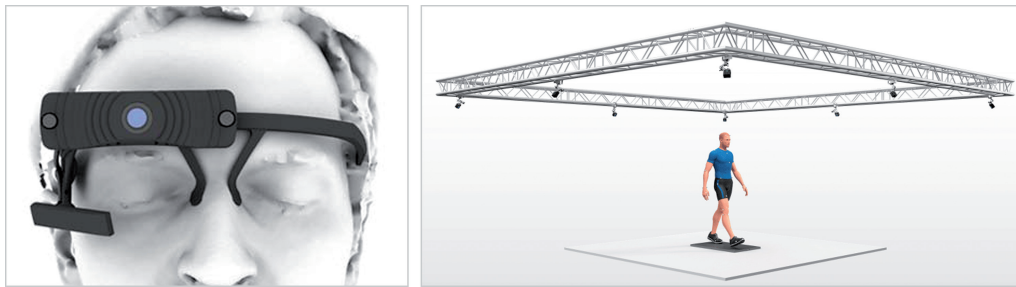


Figure 5.3: Pupil Pro eye tracking headset (left). Adapted from [72]. Example OptiTrack motion capturing setup (right).¹⁵

In the following, the prototypical implementation of the components and the setup for the studies is explained in detail.

Sensors and Sensor Service

We implemented the framework for adaptive gaze guidance with both eye tracking and motion capturing to compare the two types of sensors in terms of their suitability for gaze guidance.

Eye Tracking The eye tracking system Pupil Pro uses eye tracking glasses to record both the positions of the pupils and the environment (see Section 2.1.4) and can use this information to determine at which position in a predefined area the participant is looking. In total, the shelf was divided into 8 different predefined areas, each containing two shelf compartments located one above the other. The recognition of the currently viewed area was done with the help of 48 2D markers, which were attached to the shelf (see Figure 5.4). If a person's gaze falls within one of these areas, the eye tracker returns the predefined name of this area (e.g. shelf area 1) and the normalized coordinates (i.e. (x,y) values between 0 and 1) in this area. For all other areas, values outside the normalized range are provided to the connected sensor service. The sensor service translates the sensor data received into the corresponding shelf compartment. A y -value of 0.5 represents the separation value between the two compartments of a shelf area. It then sends the determined information to the processing unit as an EBS event.

Motion Capturing The OptiTrack motion capturing system is used to determine the orientation of the participant's head in order to determine the partici-

¹⁵Source: <https://optitrack.com> (last accessed: 2023-08-15)



Figure 5.4: Representation of the shelf, which was created to guide the gaze. Smartphones, of which only the displays are visible, serve as price tags/actuators. 2D barcodes are used by the eye tracker to localize the environment.

participant's approximate line of sight. The motion capturing system determines the position and rotation of objects using reflective markers. To determine head orientation, we installed four of these markers on a cap that we placed on participants' heads during the study. Based on the position and rotation data of the head as well as the exact position and orientation of the shelf, the Sensor Service determines to which area of the shelf the participant's head is aligned. The determined shelf compartment is then transmitted to the Processing Unit as an EBS event.

Actuators

The framework was implemented using smartphones as actuators. These were attached to each of the shelf compartments of the instrumented shopping shelf and, in addition to displaying the visual cues, also served to display the product

information, including the price. A total of 16 Samsung Galaxy S3 minis were attached to the 16 compartments, as well as four additional smartphones located in the top bar of the shelf, whose flashing lights could be used to illuminate the product compartments below. Decorative strips were attached to the shelf so that only the displays of the smartphones were visible (see Figure 5.4). The smartphones were coupled together using USB hubs, so that only one USB cable was needed to connect all the phones to the PC and to be able to load the Android application implemented specifically for this purpose. The developed application is able to receive all events sent by the processing unit to display visual cues and execute them accordingly. A large number of cues were implemented, which could be transmitted to the smartphones via EBS. Depending on the stimulus, the duration, intensity and frequency could also be set. The following list shows an overview of the most important implemented cues:

- light-dark flashing price tags
- color flashing price tags
- blurred price tags
- illuminated shelf compartments
- displayed website
- played sound
- text-to-speech output

In addition to the cue stimuli, other events could be transmitted to the smartphones, e.g., to stop the cue stimuli or to turn off the smartphones.

Processing Unit

The prototypical development of the processing unit was done in Unity. The initial user information and the environment information required for the studies were stored in xml files that were read in at the beginning of each study run. Since our studies were to start in a non-trained baseline condition, only the ID was read in as user information. However, more detailed information about the individual, such as the preferred cue stimuli, can be stored in the aforementioned xml document. In the environment information file, it is noted in detail which product

is located at which shelf position, including additional product information, such as the price. The Processing Unit knows the structure of the shelf and which actuator (i.e. which smartphone) is located at which shelf compartment. This information is needed to be able to send out targeted information stimuli.

The Processing Unit is connected to the Sensor Service and receives information from it about the shelf compartment the user is currently looking at. Depending on which shelf compartment the user's attention should be directed to and how far away it is from the current point of view, the Processing Unit determines the intensity with which the stimulus should be displayed. User preferences can also be taken into account when selecting the stimulus and the intensity. We implemented the framework in such a way that the stimulus can be switched if directing attention with the previously used stimulus did not work. In this way, the system can adaptively adjust to the particular user. By knowing which stimuli work well or do not work for a person, the gaze guidance can be optimized by directly sending a suitable stimulus with appropriate intensity to the actuators the next time. As described in the previous section, the cues were sent via events that were transmitted to the smartphones by the EBS. A separate EBS event was implemented for each cue. The Processing Unit provided the EBS server to which the smartphones and the Sensor Service were connected as clients.

User Interface

In the process of the prototypical implementation, a user interface was also developed, which was implemented in the form of a web interface (see Figure 5.5). With the web interface, it is possible to trigger specific stimuli on individual or several product compartments. To do this, the product compartments for which the stimulus is to be triggered are first selected in the left-hand area of the interface. This is done by switching off or on individual switches or even entire rows with the help of the check mark shown. In addition, it is also possible to switch all actuators off or on in the "Actuator Status" area at the top right. Once the corresponding actuators have been selected, the event to be sent can be selected in the right-hand area of the user interface. In this case, it is possible to make the price tags flash light-dark or in a color at a specified frequency, to play a sound, to illuminate the shelf compartments with the help of the smartphones' flashlights, or to transmit a text that is read out by the actuators. The "Send Event" button then sends the specified stimulus to all selected shelf compartments. All active stimuli can be stopped using the button located in the upper right-hand

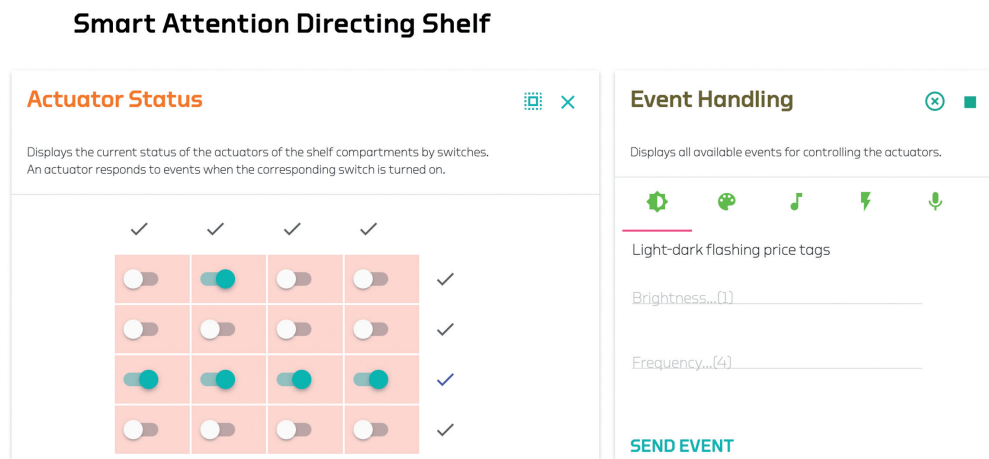


Figure 5.5: Web interface for controlling the prototypically implemented framework realized as a smart attention directing shelf.

corner of the “Event Handling” area. The button to the left of this allows all smartphones to be switched off so that this does not have to be done on each smartphone individually.

5.1.3 Conclusion

The developed framework is capable of displaying cueing stimuli to guide gaze direction in the environment. It can also detect users’ reactions to the emitted stimuli and respond to their gaze direction changes in a targeted manner. In Section 3.2.2, we presented two studies using the framework. By displaying various forms of cues on smartphones and detecting gaze direction using eye tracking and motion capturing, we determined the suitability of these sensors and actuators for guiding gaze direction. However, the presented framework allows the use of arbitrary actuators, which enable the output of visual or auditory signals. Likewise, any sensors that can detect the direction of the user’s gaze can be used to detect the direction of gaze. With the help of the developed framework, it is thus possible to conduct a variety of studies to specifically investigate the effect of different cue stimuli in an instrumented environment. The User Interface enables simple targeted control of the connected actuators. In addition to sending out specific events to display cues, it could also be used to start individual predefined study runs, for example.

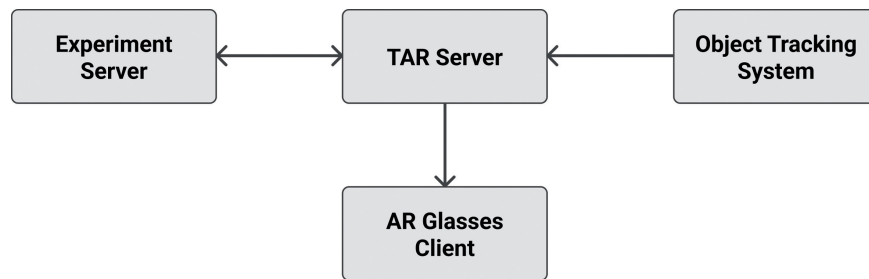


Figure 5.6: Main components of the framework for AR proxy interaction.

5.2 Framework for AR Proxy Interaction

The framework for AR proxy interaction is used to display virtual objects as overlays on physical objects in AR glasses. By selecting different overlays that cover the physical object, it is possible to measure the degree of difference between the physical object and the virtual overlay that results in performance losses or a massive deterioration in usability.

In the following, the concept of the framework is presented first. Subsequently, the example implementation of this framework for the execution of our studies from Chapter 4 will be explained in detail.

5.2.1 Concept

The framework for AR proxy interaction consists of 4 components, the Experiment Server, the TAR Server, the Object Tracking System and the AR Glasses Client (see Figure 5.6).

The central component of the framework is the TAR Server, which receives data from both the Object Tracking System and the Experiment Server, processes it, and then makes it available to the AR Glasses Client. The individual components are briefly introduced below.

Object Tracking System

The Object Tracking System provides information about the positions and rotations of the physical objects with which the user interacts, as well as information about the position and rotation of the AR glasses used. Position information must be available continuously so that the overlays can later be displayed at the correct position in the AR view of the glasses at any time.

Experiment Server

The Experiment Server is used to simplify the execution of studies. It contains the information about the individual subtasks that must be performed by each individual participant and the order in which they must be completed in each case. This is especially helpful if each subject has to complete the tasks in a different order, e.g. to ensure a balanced study design. The Experiment Server makes it possible to inform the TAR Server one by one about individual tasks that are to be performed. Once a task is completed, the Experiment Server receives the measurement results back from the TAR Server and stores them in files for later evaluation. In addition to the tasks, information on questionnaires can also be transmitted. Their results are also forwarded from the TAR Server to the Experiment Server for storage.

AR Glasses Client

The AR Glasses Client receives information from the TAR Server about which overlay it should display at which position, with which rotation and in which color. The transmitted information is processed and the corresponding objects are displayed in the AR view. The AR Glasses Client is only used to visualize objects in the AR view. No data processing takes place in order to minimize the load on the glasses.

Since depth perception is different for each person, e.g. due to the distance between the eyes, an appropriate calibration to the eyes must be made. Many of the AR glasses on the market already offer such functionality. The procedure for calibrating to the eyes has to be done separately for each user. The mentioned procedures work quite well, but they may somewhat exceed their limits for people with widely spaced eyes or very thick lenses, as well as contact lens wearers, so that a manual correction in the TAR Server is still necessary for these specific cases.

TAR Server

The TAR Server represents the core of the framework and performs many different process steps. Among other things, it allows us to perform a fine calibration to the eyes of the respective user, if the automatic calibration of the AR glasses did not succeed. It also provides a way to check the head-desk distance and to

automatically start and stop a task, which enables accurate timekeeping. In addition, it receives position and rotation information of the physical proxy objects and the AR glasses. Since this position information is available in the respective coordinate system of the Object Tracking System, a coordinate transformation first takes place in the TAR Server so that the data is compatible with the position information of the AR glasses. This ensures that the overlays can be displayed at the correct positions in the AR view of the glasses.

The Experiment Server provides the TAR Server with information about the current task to be performed. The TAR Server knows where the physical objects are located in space and thus informs the AR Glasses Client where to display which overlay.

During the execution of the task, the TAR Server continuously receives the updated position data of the proxy objects from the Object Tracking System and checks whether the given task (e.g. placing the objects at a certain position) has been solved. If this is the case, it can, for example, send different color information for a certain object to the AR Glasses Client so that it is also visually visible that the task has been solved. Once the task has been solved, the Experiment Server is also informed accordingly. In addition, it receives measurement results from the TAR Server, e.g. how long it took to solve the respective task.

In addition to processing AR proxy interaction tasks, questionnaires can also be completed in tangible AR using the framework. The Experiment Server informs the TAR Server about the corresponding questionnaire to be completed by the user. The AR Glasses Client receives – similar to the processing of tasks – the information as to which overlay/question it should display at which position. With the help of a physical object, which serves as a kind of pen, the questionnaires can be filled out. After completion, the results are then transmitted from the TAR Server to the Experiment Server.

5.2.2 Prototypical Implementation

The concept described in Section 5.2.1 was implemented to enable conducting the studies described in Chapter 4.

The Object Tracking System chosen was a motion capturing system from the company OptiTrack, including the associated application called “Motive” for processing and streaming the data. The AR Glasses Client was implemented using a HoloLens 2 application and the TAR Server was implemented in Unity.

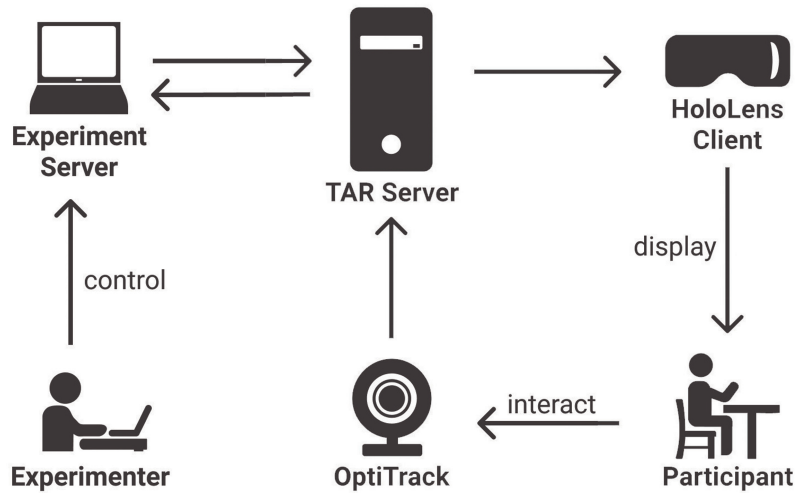


Figure 5.7: Detailed architecture of the framework for AR proxy interaction.

Communication between the OptiTrack software Motive and the TAR Server was done using the NatNet¹⁶ networking protocol, which uses UDP and allows sending and receiving data in real time. Information from the TAR Server to the HoloLens client is sent via MQTT. The data exchange between Experiment Server and TAR Server takes place via TCP. Figure 5.7 visualizes the interaction of the individual components of the framework.

In the following, the implementation of the components as well as the communication between them is explained in detail.

OptiTrack – Object Tracking System

For our studies, we chose OptiTrack’s stationary motion capturing system to track the positions. This is very accurate and allows for a fast detection of the objects. Compared to tracking based on image targets, which we implemented for testing purposes using Vuforia, it is significantly more robust to dropouts that occur, for example, when markers are occluded during interaction with objects. Especially when it came to fast interaction with objects, as was necessary in our studies, the implementation with OptiTrack proved to be much more robust.

Object Tracking The OptiTrack system uses cameras to detect reflective elements within the tracked area. The camera information is received and processed

¹⁶<https://optitrack.com/software/natnet-sdk> (last accessed: 2023-08-15)

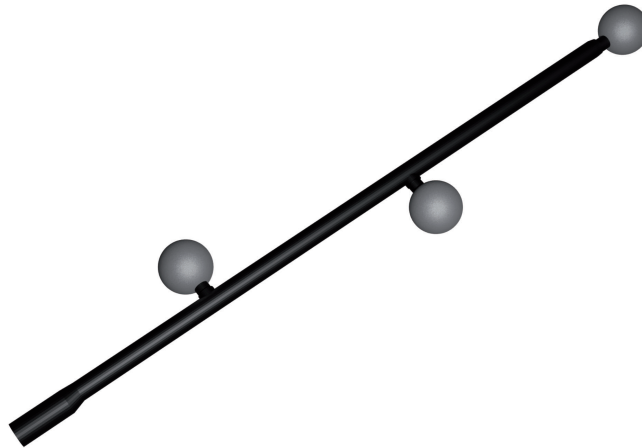


Figure 5.8: Example of a marker tree.

by the associated application Motive. Motive allows several of these reflective elements to be combined into a so-called rigid body with an associated pivot point, whose position and rotation data can then be broadcast via NatNet.

In order for the markers to be tracked in space, we first had to carefully calibrate the system and align the OptiTrack coordinate system. In addition, the objects to be interacted with had to be provided with reflective markers in order to be able to track them.

Marker Arrangement The recognition of objects in OptiTrack is done with the help of reflective markers, as described above. A rigid body must consist of at least three of these markers. For our implementation we used reflective spheres with a diameter of 11 mm. In order for an object/rigid body to be properly detected, several aspects must be considered when placing these reflective markers. For example, they must be far enough away from each other to be recognized as individual markers by the OptiTrack system. We therefore had to keep a distance of more than 4.5 cm between the spheres when placing them. To ensure that the markers interfere as little as possible when interacting with the objects, we attached them to marker trees that we placed on the objects. To ensure that the objects do not interfere with each other, we had to make sure that the distance between the marker spheres of one object and the spheres of the other objects was greater than 4.5 cm at all times. The trees were therefore placed at the center of the objects and mainly built up in height. However, this also made the trees a little more unstable, which leads to stronger wobbling/shaking, especially when

placing the objects on the table. Figure 5.8 shows an example of a model of a marker tree in use. In addition, it had to be verified that the orientation of the markers of a rigid body is unambiguously defined from each side. Furthermore, it had to be ensured that the marker trees of the objects used differ sufficiently from each other so that at no time is a different rigid body falsely recognized.

Broadcasting The broadcasting feature is used to make the position data of the rigid bodies available outside Motive. Here one can specify to which address the data should be streamed. If – as in our case – the data is to be processed by the same computer, the LocalLoop functionality must be used so that the data is available via localhost. The rigid bodies must be provided for this in advance with IDs, in order to be able to assign them later in the TAR Server again.

Experiment Server

The Experiment Server is used to manage studies regarding AR proxy interactions with AR glasses. In our studies, it was used to sequentially transmit the current task via TCP to the TAR Server. We read in the respective task sequences from text files in which we had previously stored them to guarantee the balancing of the tasks. To ensure easy interaction with the Experiment Server, we created a graphical user interface that allows us to send out individual tasks and tells us which task is being worked on.

Figure 5.9 shows a screenshot of the Experiment Server interface used in the study on lighting variations, which was created using Java Swing¹⁷. In the upper area it is possible to set which participant is concerned, which study part he or she is currently working on and which specific task he or she is performing. The option to select the current task number makes it possible to continue from a specific task even in the case of unforeseen program errors or disconnections. It is possible to see which task is currently in progress and whether it is still being processed or has already been completed. As long as the subject is still working on the task, the box on the right appears in green and as soon as the TAR Server has returned results and the experimenter has to take action to start the next task, the color changes to red. The checkbox next to “Calibrate” can be used to select whether or not to display the cuboid object for manual eye calibration (see the paragraph *Manual Eye Calibration* in Section 5.2.2), and this information can be

¹⁷<https://docs.oracle.com/javase/8/docs/technotes/guides/swing/> (last accessed: 2023-08-15)

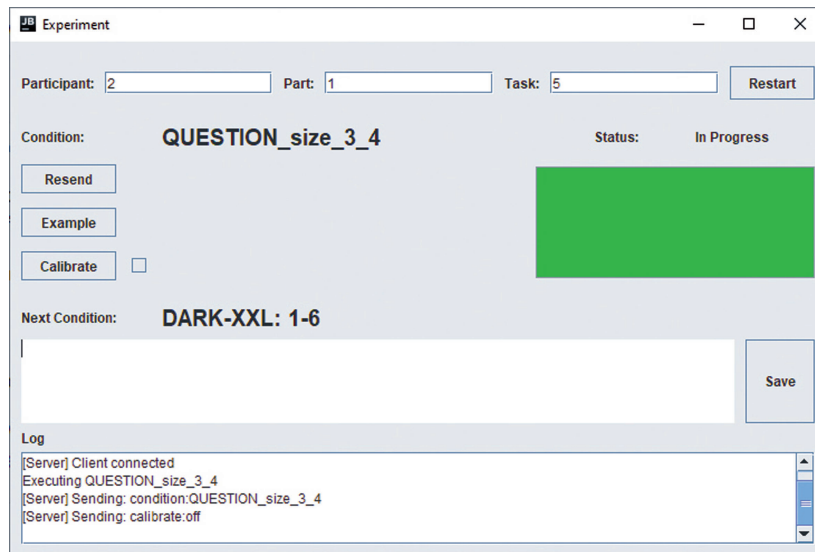


Figure 5.9: Screenshot of the graphical user interface of the Experiment Server for the lighting variations study.

sent to the AR Glasses Client via “Resend”. The “Example” button is used to display the example task that all participants must perform at the beginning of each study part. In the lower part of the user interface it is visible which task has to be performed next. Furthermore, there is a comment field where notes on verbal comments of the study participants on an individual task can be entered and saved. Finally, the current connection status of the server or the information sent out can be seen.

The results returned by the TAR Server, such as the measured time in performing the last task, were stored in log files. These were structured in such a way that a separate file was created for each subject for each measured variable. For example, each participant had one file with time measurements, one with the answers to the AR presence questionnaire and one with the answers to the TAR presence questionnaire, etc., in order to enable an easy evaluation afterwards.

HoloLens Client

For our studies, we chose the HoloLens 2 as the AR Glasses Client. We have also done test implementations for the HoloLens 1¹⁸ and the Magic Leap 1. However, the Magic Leap 1 was not usable for our studies in which the participants had to

¹⁸<https://docs.microsoft.com/de-de/hololens/hololens1-hardware> (last accessed: 2023-08-15)

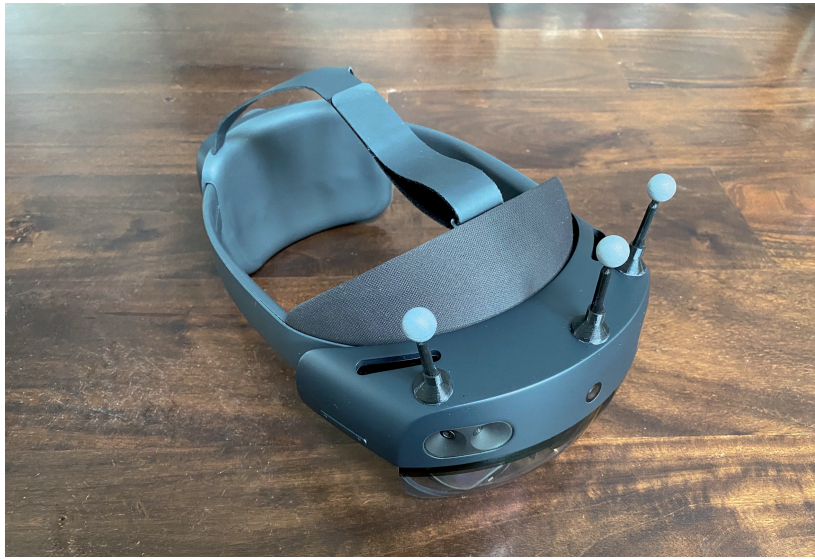


Figure 5.10: HoloLens 2 equipped with reflective markers.

arrange objects on the table in front of them, because it has a fixed near-clipping plane of 37 cm. Therefore, objects that were closer were not displayed, or cut off. The HoloLens 1 was not an option for the studies either, since with $30^\circ \times 17.5^\circ$ it has a significantly smaller FOV than the HoloLens 2¹⁹ with $43^\circ \times 29^\circ$ has. This would not have allowed, for example, studies on size variations with three different interaction objects simultaneously.

The HoloLens 2 is used to display the virtual objects in the AR view. The position of the glasses in the room is set using the position of the markers on the glasses (see Figure 5.10) determined by OptiTrack and the positional tracking of the HoloLens is deactivated accordingly. Since it is not possible to deactivate the HoloLens' rotational tracking, the rotation data from the OptiTrack system cannot be included at this point. Instead, each time the application is launched, the HoloLens is aligned in space so that its rotation matches the camera in the Unity scene. The exact alignment of the glasses in the room is achieved by a construction on a table that is precisely aligned through the use of markers.

At the beginning of each study, an automatic eye calibration was performed. With the HoloLens 2 application available for this purpose, the user must track gems displayed in the AR view with his or her eyes without moving his or her head. The HoloLens 2 then informs the user whether the eye calibration was successful. However, our manual eye calibration in the TAR Server (see the para-

¹⁹<https://www.microsoft.com/de-DE/hololens/hardware> (last accessed: 2023-08-15)

graph *Manual Eye Calibration* in Section 5.2.2) was always performed regardless of this result, as it turned out that the calibration does not always work well, e.g. for people wearing glasses, and we had to make sure that a very high precision was given. Apart from that, HoloLens 2 is only used to display predefined objects. It receives information from the TAR Server regarding what has to be displayed. In total, the HoloLens client can handle three types of input: commands, rigid bodies and questionnaires. Commands include activating/deactivating of the calibration object, activating/deactivating of targets and virtual props or coloring them green when a task is completed, setting the size of props and targets, and other commands for debugging.

The information about the rigid bodies includes the ID, the position and the rotation of the objects. In addition, it is transmitted whether the objects should be visible or not and whether they should be displayed in green, which represents that the mapping of virtual object and target was successful.

The information that is transmitted to display the questionnaires is quite diverse, e.g. to be able to jump back and forth between individual questions.

TAR Server

As mentioned in Section 5.2.1, many different processes are performed in the central component, the TAR Server. It performs the coordinate system transformation, sends data to the HoloLens, and checks whether the respective task has been completed. In addition, the TAR Server enables manual fine calibration of the system to the eyes of the respective user. In the following, these processes are described in detail on the basis of the prototypical implementation for the studies from Chapter 4.

Manual Eye Calibration To ensure that the virtual overlays are in the same position as their physical counterparts, the first step is to check whether the automatic eye calibration of the HoloLens2 has worked. For this purpose, users receive an example object on which a suitable AR overlay is displayed. In the case of our studies, the calibration was performed by means of a calibration triangle with a cuboid in the center, which is overlaid by a corresponding AR overlay (see Figure 5.11). If the virtual object is not in exactly the right place, it can be moved according to the user's specifications using the developed Unity application until an optimal overlay is ensured. Once the application has been accurately adjusted

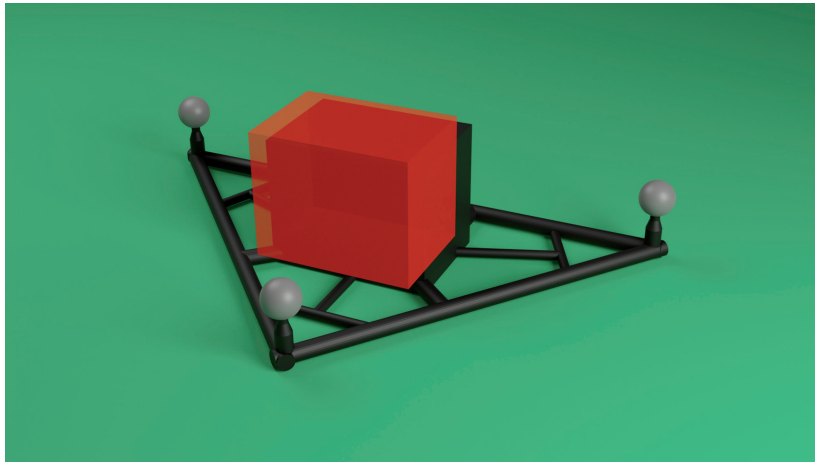


Figure 5.11: Physical example object for participant's eye calibration. The shifted overlay (red box) is adjusted until it fits the black box.

to the participant's eyes, all other AR overlays will later also be displayed in the correct position.

Head-Desk Distance Check To ensure that all subjects have approximately the same viewing angle of the interaction surface, the chair and the table are positioned in the same way for all participants. Additionally the distance between the head and the tabletop has to be adjusted to guarantee a similar view. For this purpose, the framework allows the calculation of the distance between a marker on the tabletop and a marker on the HoloLens. This distance is checked at the beginning of each study session and is displayed in the Unity application of the TAR Server. The test person has to adjust his or her chair height until the desired value is displayed here.

Field-of-View Check An important requirement for conducting AR proxy interaction studies is that the overlays on the proxy objects are visible at all times. To ensure this, the FOV of the test person must be aligned properly. With the help of the framework, it is possible to display the outer edges of the FOV of the HoloLens 2 so that the participant can adjust his or her FOV to a predefined region in the real world. Figure 5.12 illustrates the adjustment of the FOV (white corners in the AR overlay) to the interaction surface (white corners on the table).

Automatic Task Start and Timing In order to enable meaningful time measurements during the execution of tasks, it should be ensured that the participants



Figure 5.12: Adjustment of the FOV to the interaction area.

can only see the objects as well as their associated overlays and possible targets as soon as the time measurement has been started. Therefore, the props were placed covered with a box in front of the subjects (see Figure 5.13). The framework enables the automatic display of the AR overlays as soon as the box is lifted and the objects are detected by the OptiTrack system. At the same moment, the time measurement also starts; this ends automatically as soon as the task is completed. In this way, exact time measurement is performed for each task.

Coordinate System Transformation Motive and Unity use different coordinate systems, so a transformation from Motive’s right-handed Y-up coordinate system to Unity’s left-handed Y-up coordinate system must first be performed. This functionality is already provided by Motive’s streaming client for Unity, which we use.

HoloLens Communication The communication between the TAR Server and the HoloLens was done via TCP for our studies. For the lighting variations study and the shape variations study, the data was sent to the HoloLens application via MQTT broadcast to make the connection even more stable. Mosquitto²⁰ was used as the MQTT broker. This decoupling between the TAR Server and the HoloLens client makes it possible to restart only one component in case of problems and to split the information into individual channels for clarity.

²⁰<https://mosquitto.org> (last accessed: 2023-08-15)



Figure 5.13: Physical props covered with a box in front of the participant.

To keep the load on the HoloLens as low as possible, all communication is one-way. The TAR Server publishes its information to the MQTT broker, which sends the messages to the HoloLens application. The messages in this case are grouped according to priority. Regarding the rigid bodies, the tangible props have a high priority, the calibration triangle a default priority and the targets a priority of the type “trigger”. Messages with priority “trigger” are only sent if a corresponding trigger has been fired. Position information of the tangible props and the calibration triangle are sent 60 times per second.

For each rigid body, its unique ID, the position divided into the three coordinate axes x , y , and z , the rotation as quaternion defined by four values (analogous to Unity) and the visibility information (visible/invisible/green) of the rigid body are sent. In addition, commands are sent that contain information about the size of the virtual objects or whether the calibration cuboid should be displayed. The information for displaying the questionnaires is transferred via JSON string.

The HoloLens client interprets the corresponding information and visualizes it. Feedback is not sent back to the TAR Server. Since the TAR Server takes care of the verification of the executed tasks (see the paragraph *Task Verification* in Section 5.2.2), a communication back to the TAR Server is not necessary, which saves valuable resources of the HoloLens.

Experiment Server Communication With the help of the Experiment Server, the TAR Server can be informed about which task is to be processed next or which question of a questionnaire is to be answered next. In addition, the information as to whether the calibration should be active or not can also be sent.

A possible message to send a task is, e.g., `condition:1_2_3-4`, which means: lighting condition 1, size condition 2, target arrangement 3 and target arrangement 4. For each combination of size and lighting condition, participants must solve two tasks, so two target arrangements are sent.

To ask a specific question from a questionnaire, e.g., `QUESTION.size.1.2` is sent for the question on lighting condition 1 and size condition 2 in the size perception questionnaire, or `QUESTION.lighting.1.5` for the question about lighting condition 1 and size condition 5 in the lighting perception questionnaire. The TAR server now knows the current task and thus knows at which position which target has to be displayed. It forwards this data, along with the current object's position information from OptiTrack, to the HoloLens.

Task Verification By knowing at which positions the virtual targets are located, the TAR Server knows at which positions the physical objects must be located for the task to be considered complete. Therefore, in each frame the object positions and rotations provided by OptiTrack are compared with those of the corresponding target positions. Since exact matching of target position and rotation is nearly impossible, appropriate thresholds were set for each task by which a physical object can deviate from the target position in order to solve the task. The thresholds for possible differences between position or rotation of the target and the corresponding physical object were chosen in such a way that there is a good balance between solvability and complexity. On the one hand, this should prevent users from becoming frustrated quickly because it takes too long to solve the task; on the other hand, the task should also be difficult enough, i.e. involve fine-tuning of the object placements, so that differences between the individual conditions can be determined. We considered differences in rotation and position separately. We used different thresholds for each study since we also changed the OptiTrack cameras and the software version of Motive after our study on size variations. The change from OptiTrack Flex 3 to OptiTrack Prime^X 13 cameras and Motive version 1.15 to Motive version 2.3 allowed for a more precise and more stable positioning. The thresholds used for the respective studies are given in the corresponding sections in Chapter 4.



Figure 5.14: Excerpt from the in-AR questionnaire: Selection with the help of the interactive pen (left) and selected option (right).

Tangible AR Questionnaire In addition to handling AR proxy interaction tasks, the framework additionally enables questionnaire completion in AR using a physical proxy object. In this case, the communication between Experiment Server, TAR Server, and HoloLens Client is the same as when processing tasks. Instead of individual targets, however, the user is successively shown questions from a questionnaire that are to be answered by means of a 7-point Likert scale. A pen-like physical object overlaid with a suitable AR overlay serves as the input medium for this purpose. For displaying the questions in the AR view we adapted the framework of Feick et al. [29] for questionnaires in VR to our needs. Figure 5.14 shows two screenshots taken with the HoloLens while answering a question. The question is displayed at the top and the corresponding answer options below it. In the bottom area there are buttons to jump back to the last question or to display the next question. If the “pen” is placed on an answer option, this selection is highlighted in blue accordingly. Changing the selection is possible at any time. If all questions of a questionnaire have been completed, the TAR Server sends the results to the Experiment Server, which stores them for later evaluation.

5.2.3 Conclusion

The implemented framework enables the visualization of virtual overlays on physical objects used to interact with the virtual content. A precise object tracking system in combination with the manual fine calibration to the eyes enables a very exact placement of the overlays in the AR view of the glasses. The ability to use arbitrary virtual objects as overlays on the physical props allows for a variety of studies to investigate differences between the virtual object and its physical counterpart. In Chapter 4, we presented studies that examine differences with

respect to a single feature, such as size or shape. However, the framework can be used to evaluate any kind of differences between objects in user studies. The integrated possibility to answer questions in AR with the help of a tangible pen represents a more pleasant way of filling out questionnaires without leaving the AR experience. Likewise, the graphical user interface of the Experiment Server additionally enables helpful support for study execution. This could be extended with a variety of additional features that would further simplify study design, execution, and analysis, such as an integrated tool to assist in creating a balanced study design, which would eliminate the need to read in text files created manually in advance.

5.3 Summary

In this chapter, two different frameworks were introduced to investigate the perception of visual augmentations of reality. The first framework makes it possible to emit visual stimuli in instrumented environments and to determine a person's reaction to a stimulus by evaluating the person's gaze direction at runtime. Through this, the framework also enables adaptation, e.g. of the stimuli, to the respective user. The second framework provides a way to evaluate the perception of virtual objects overlying physical proxy objects used for interaction. Using the framework, it is possible to accurately place virtual overlays on their physical counterparts and evaluate the perception of the objects during interaction using Tangible AR questionnaires. Both frameworks thus provide a basis for measuring the impact of visual augmentations on the perception of the environment.

Chapter 6

Conclusion

In this concluding chapter, we first provide an overview of the research conducted in this dissertation and highlight its major contributions. We then discuss possible directions for further work beyond the scope of this thesis. Finally, we close the thesis with some concluding remarks.

6.1 Summary

The goal of this thesis was to investigate how our visual-haptic perception of reality can be influenced by digital augmentations. We wanted to find out how a person's gaze direction can be influenced by a digitally augmented reality (**RQ1**) and how digital augmentations on real 3D objects influence the perception of these objects (**RQ3**). Furthermore, the aim was to show how suitable visual cues and visualization types for gaze guidance can be determined (**RQ2**) and how it is possible to test the influence of visual augmentations of reality on our visual-haptic perception of objects (**RQ4**).

To answer the research questions, we designed and implemented frameworks and conducted a series of studies. In the following, a summary of the dissertation work is given.

1. **Subtle Gaze Guidance in Real and Real-size Scaled Environments (RQ1, Chapter 3):** In Chapter 3, we investigated whether and with which methods it is possible to direct visual attention using subtle visual cue stimuli. We

determined how well different visual stimuli are suited as well as investigated different methods for determining gaze direction for their suitability regarding gaze direction guidance. For the first time, we also investigated subtle gaze direction guidance in OST AR and determined which visual stimuli are suitable for this purpose.

- 2. Framework for Evaluating Gaze Guidance Methods in Instrumented Environments (RQ2, Chapter 5):** In Section 5.1 we presented a framework that allows us to test which visual and auditory stimuli in the environment are suitable for directing visual attention. The framework enables the control of actuators located in the environment as well as the detection of the user's gaze direction. This allows the framework to be used to evaluate the effectiveness of different stimuli for directing gaze.
- 3. Influence of the Appearance of Visual Overlays on the Visual and Haptic Perception of Proxy Objects. (RQ3, Chapter 4):** In Chapter 4, we investigated the impact of visual overlays on the visual and haptic perception when interacting with proxy objects. We determined how much the visual overlying object can deviate from the underlying physical object without serious degradation of the feeling of presence, the usability, and the performance. In studies we identified possible deviations with respect to size and with respect to shape of the objects, as well as investigated the influence of environmental illumination on possible size deviations.
- 4. Framework for Evaluating Visual-haptic Perception of Visually Overlaid Proxy Objects in OST AR. (RQ4, Chapter 5):** In Section 5.2 we presented a framework that enables the evaluation of visual-haptic perception of visually overlaid proxy objects. The framework enables the determination of the position of the proxy object in space and the exact placement of the virtual object at the position of the physical object in the 3D AR view. This allows for investigations related to possible discrepancies between the virtual object and the physical object in OST AR. The framework also enables questionnaires to be answered in OST AR using tangibles, thereby ensuring that the AR experience is not disrupted when answering questionnaires.

6.2 Major Contributions

In Section 1.4 we have already given an overview of the contributions to the research fields. In the following, we summarize the major contributions of this thesis by addressing the research questions.

RQ1: How can a person's gaze direction be influenced by a digitally augmented reality?

Our studies reveal that people's gaze direction can be guided by visual stimuli that are displayed in the environment. Not only are obvious cues suitable for guiding gaze in the real world, but also subtle cues that are hardly noticed or not noticed at all. We found that less obvious cues are perceived as less disturbing, but have a poorer success rate in successfully directing the gaze to the target region. Subtlety is achieved, for example, by adaptively adjusting the intensity of the stimulus depending on the distance of the current gaze direction to the target region. For successful gaze guidance with adaptive visual stimuli, a coarse gaze direction, e.g., by determining the head orientation, is sufficient.

The outcomes show that subtle stimuli can be used to direct people's attention to target areas. By using OST AR, people can be individually pointed to information that is relevant to them. When OST AR has become commonplace, it will be possible to reduce obvious visual cue stimuli in the environment and thus decrease information overload.

RQ2: How can we test which actuators and sensors are suitable for directing visual attention?

An important contribution of this work is the development of the adaptive gaze guidance framework designed during the dissertation period. With the help of the newly developed framework, it is possible to investigate the suitability of surrounding cue stimuli for guiding gaze direction. This is done by determining and evaluating the reactions that occur in the form of changes in gaze direction in response to the emitted cue stimuli. In addition to the integration of different actuators that emit the cue stimuli, various sensors for gaze direction determination can also be incorporated in order to assess their impact on the adaptive gaze guidance. The information about the gaze direction, together with the information about the user and the environment, allows the control system to

determine an appropriate visual stimulus and transmit it to a specific actuator. The reactions to the emitted stimuli can be learned and thus taken into account by the control system in the future.

By providing descriptions of the framework, which explain the individual components and the interconnection between them, we aid other researchers in quickly building their own implementation of the framework. Since the framework allows the connection of different hardware, researchers can adapt the framework to their own hardware and then use it to determine suitable visual cues for gaze guidance and suitable sensors for gaze direction detection for their own use cases.

RQ3: How do digital extensions of reality in the form of overlays on real 3D objects influence the visual-haptic perception of these objects?

The results of our research show that visual augmentations superimposed on physical proxy objects influence their visual-haptic perception. Large discrepancies between virtual and physical objects in terms of size and shape significantly worsen usability, performance and the feeling of presence. However, small differences in size and shape should be possible without major losses, especially since these are often perceived neither visually nor haptically, as our studies have shown. We found that environmental lighting affects the visual-haptic perception of digitally overlaid objects. In the investigations under low environmental lighting, the virtual overlays were visible almost exclusively, which makes this setting comparable to VR and VST AR. With increasing room lighting, the virtual overlays were perceived more transparently by our study participants and the underlying physical objects became more clearly visible. Our results revealed that with increasing room illumination, larger size differences between virtual and physical object are feasible without significantly worsening usability, presence, and performance ratings.

The findings demonstrate that digital augmentations superimposed on objects change the visual and haptic perception of these objects, which must be taken into account when implementing OST AR applications. Proxy objects used to interact with virtual content do not necessarily have to be an exact replica of the virtual object. They can differ at least slightly in size and shape, as these differences are usually not perceived visually or haptically. However, their size and shape should not differ significantly from the virtual object, as this would have a negative impact on the user experience. These findings are of great value for further investigations in this field, since one can already narrow down

possible differences to a rough range. In this way, more targeted investigations of potential differences can be conducted, such as determining precise limits or exploring whether a physical object can deviate in both size and shape.

RQ4: How can we evaluate whether virtual overlays can differ from their physical 3D counterparts used for interaction?

Our contribution associated with this research question is the development of the framework for evaluating the visual-haptic perception of visually overlaid proxy objects. The framework makes it possible to track physical objects in space and to display visual augmentations in OST HMDs on top of these physical objects. Additionally, the framework can be used to answer questions about the experienced visualizations directly in OST AR with the help of a proxy object without the need to leave the OST AR environment. The framework thus makes it possible to visually and haptically experience and evaluate the differences between virtual overlays and their corresponding physical counterparts.

The framework presents the hardware and software components required to conduct studies on the perception of differences between virtual and physical objects in OST AR, and shows how they need to be linked together. The descriptions of the framework assist other researchers in implementing this framework with their own hardware and thus enable them to conduct further investigations on possible differences between virtual and physical objects, for example in terms of differences in material or texture.

6.3 Future Work

The findings we have presented in this thesis represent a starting point for future research. They provide new opportunities, but also leave open challenges. In the following, we summarize the opportunities for future work that we already discussed in the individual chapters.

Visual Stimuli for Gaze Guidance

We have conducted a number of studies to examine which stimuli are suitable for subtly guiding gaze in real-world environments. In doing so, we have only been able to investigate a relatively small set of possible visual stimuli. Our focus here

was on already known methods, but many other stimuli for subtle guidance of visual attention are imaginable and should be investigated in subsequent studies.

We used two different sensors to determine the gaze direction. It would make sense to investigate other sensors, e.g., to determine whether an even coarser gaze direction is sufficient for gaze guidance. In our investigations, we only looked at how well the gaze guidance worked, but we did not investigate whether, for example, the use of a less precise sensor leads to the stimuli being perceived more clearly and thus being less subtle for the subjects. This needs to be investigated in further studies.

Another point to be investigated in future work is the question to what extent stimulus intensities that are adaptively adjusted to the user lead to better results and also ensure that the stimuli are perceived more subtly. In our study on the projected shopping shelf, we found that different individuals perceived the stimuli at very different intensities. Thus, some participants probably did not perceive some stimuli at all, while others could see them quite well. If the stimulus intensity would be adjusted to the particular participant so that he or she just perceives it, the stimulus would be as subtle as possible for him or her and it would be more likely that he or she would perceive the stimuli during the study. In our studies, it was important that the participants did not know that stimuli would be displayed, meaning that no adaptation to the user could take place here in advance. However, a general study of the extent to which adaptive stimulus intensities add value would be very interesting for the future.

In order to be able to make clear statements about which stimuli are suitable in which situations, and how well they are suited, larger-scale studies are necessary in which a large number of stimuli are compared. Here it is important to assess how subtle a stimulus was perceived or whether it was perceived at all.

Proxies for Interaction in Optical See-through AR

In this work, we investigated whether a proxy object used to interact with virtual content can deviate to some degree from the virtual representation displayed on it in OST AR. Exact bounds are to be determined in future work using, for example, the up/down staircase procedure.

When investigating shape differences between the virtual object and the physical object, we have so far only considered one well known object, which has been abstracted via a cuboid to a plane. In order to be able to make a more general

statement about the extent to which abstractions are possible, more objects must be examined in follow-up studies.

So far, we have only considered size and shape differences between the virtual object and the physical object. Further studies should investigate the influence of other features, such as texture or material. Furthermore, it is important to investigate the effect of feature combinations, e.g. when the proxy object differs from its virtual counterpart in both size and shape.

Our investigations on the influence of environmental lighting on possible size differences between the virtual object and the physical object were carried out under strict laboratory conditions so that all participants had the same study conditions and the results were not influenced by the weather situation. Nevertheless, studies in natural environments will be necessary in the future, as significantly brighter lighting conditions are to be expected and need to be investigated.

In our studies in OST AR, we used Microsoft's HoloLens 2 because at the time of the studies, these were the only glasses with a relatively large FOV that allowed us to display overlays at short distances. In the future, it will be important to investigate the impact of environmental illumination on the perception of objects while using different, more advanced devices. For example, if the FOV of OST AR HMDs becomes significantly larger, it could be investigated how perception differs when interacting with real objects in a room, instead of just interacting at a table.

It would also be necessary to investigate how the perception of objects changes when the OST AR HMD adapts to the environmental lighting. For example, the Magic Leap 2²¹, which is relatively new to the market, supports dynamic dimming. Thanks to dynamic dimming, it is possible to automatically dim the environment and thus ensure clear visibility of the virtual content even in brighter environments. In addition to dim the whole environment, it is also possible to dim only the parts of the display where virtual objects are located in order to increase their visibility. The dimming also enables the rendering of black, a capability that could not be achieved with previous OST AR HMDs. In studies, for example, this could be used to make the reflective markers less noticeable against the black background.

²¹<https://www.magicleap.com/news/magic-leap-2-optics-breakthroughs-empower-enterprise-ar> (last accessed: 2023-08-15)

Another HMD already announced is the Vision Pro mixed reality headset from Apple²², which uses VST AR to bring virtual content into the environment. According to early hands-on reports²³, the video see-through functionality is said to be significantly better than that of similar devices, but still not perfect enough to provide a true sense of reality. Although mixed reality headsets will continue to evolve in this regard, it is likely that they will remain limited to specific indoor use cases. When it comes to getting assistance at any time in normal everyday life, OST AR glasses are indispensable, as only they provide a view of actual reality, which is especially crucial for interactions in road traffic. However, to make AR glasses a reality for everyday use, some technical advances are still needed to increase the FOV and make the glasses smaller and lighter.

6.4 Closing Remarks

In this work, we have explored how we can use subtle stimuli to effectively draw people's attention to personally relevant content instead of obvious visual stimuli, which increasingly overwhelm us in everyday life. Additionally, we set a starting point for considering the impact of environmental visual stimuli on the perception of reality.

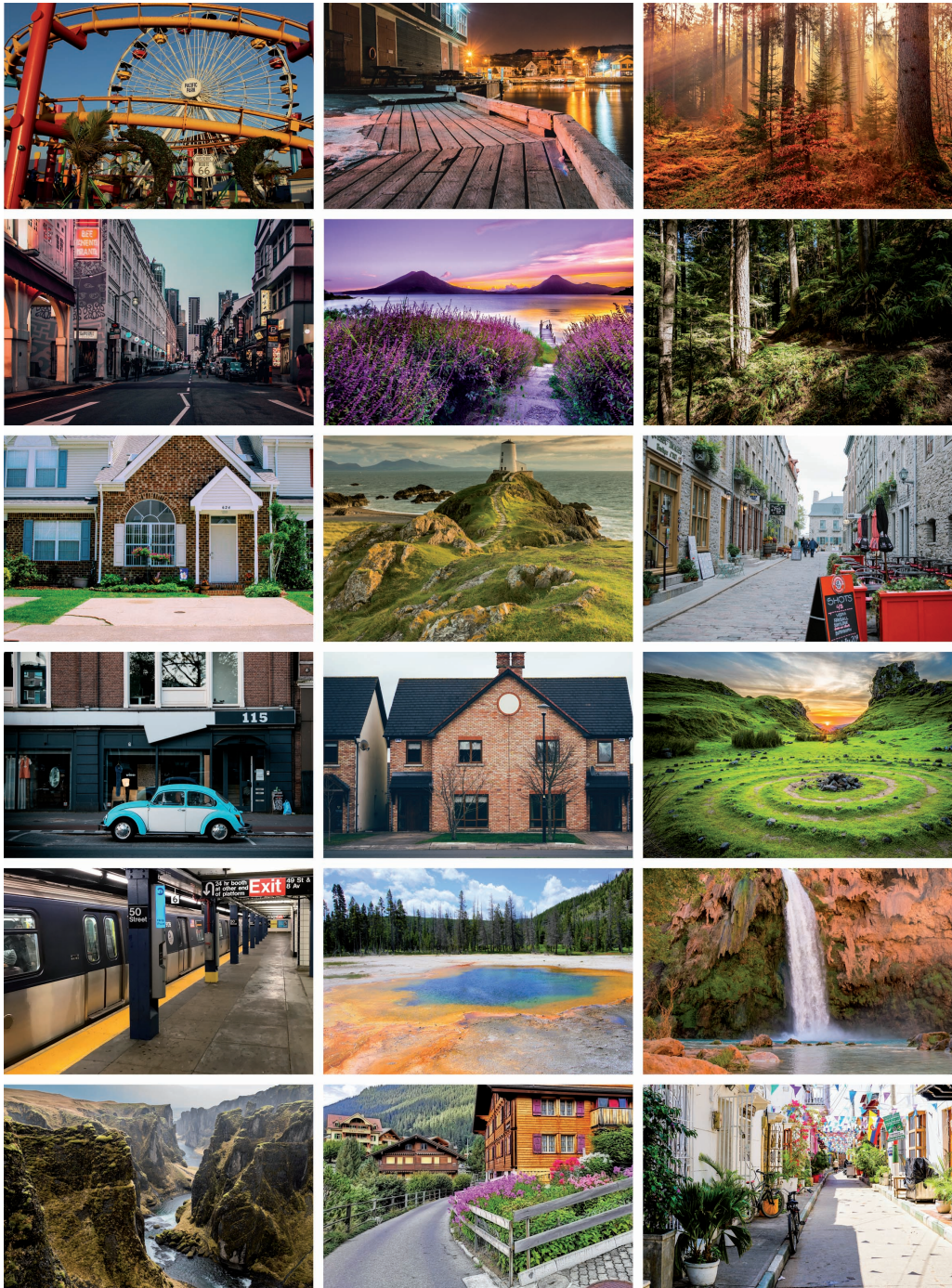
We would like to inspire others, for example when developing AR applications or creating digital content, to consider the impact of these visual representations and how they influence perception. We would also like to encourage others to take up our research and develop it even further. The use of AR glasses will probably increase strongly in the next few years. It would be beneficial if by then it is researched what influence visualizations in AR glasses have on the perception of the environment, how visualizations should be presented, and what effect external influences have on the perception of the environment.

²²<https://www.apple.com/apple-vision-pro/> (last accessed: 2023-08-15)

²³<https://mixed.de/apple-vision-pro-erste-hands-ons/> (last accessed: 2023-08-15)

Appendix A

Landscape Images



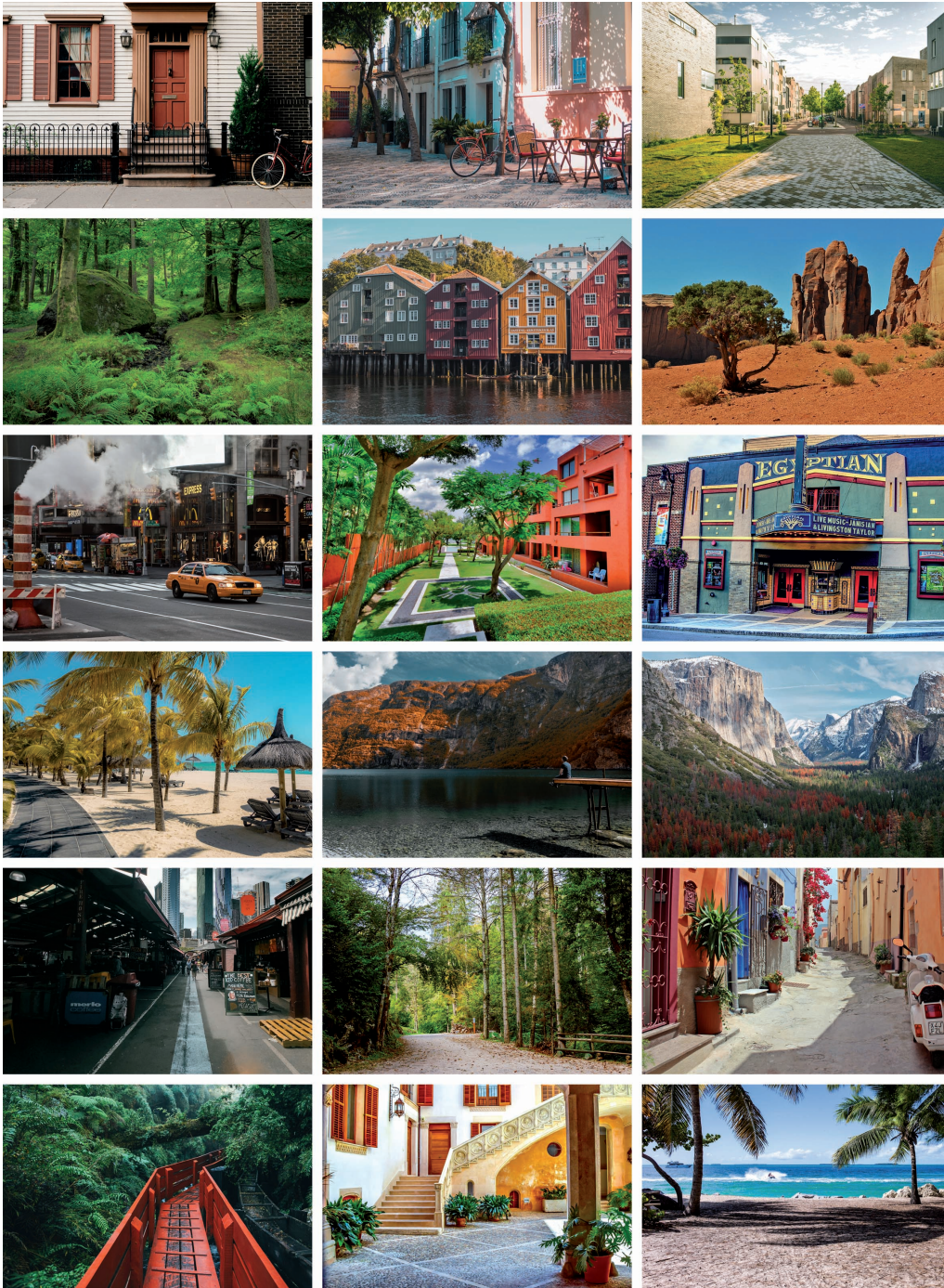


Figure A.1: Landscape images used in the study on investigations of subtle gaze guidance in AR. Source: <https://unsplash.com/de> (last accessed: 2023-08-15).

Appendix B

Questionnaires

B.1 Presence Questionnaires

B.1.1 AR Presence Questionnaires

Participant: Condition: Task:

AR Presence Questionnaire:

1) Please rate on a scale of 1 to 7 how much you felt you were in a place without visual overlays, where 7 is your normal feeling of being in a place.
I had the feeling of being in the unaltered reality:

1 = not at all	2	3	4	5	6	7 = very much

2) To what extent were there times during the experience when the visual overlays were reality for you?
There were times during the experience where the visual overlays were the reality for me...

1 = at no time	2	3	4	5	6	7 = almost all the time

3) During the time of the experience, which was the strongest on the whole, your sense of being in an unchanged reality or the feeling being in a changed reality?
I had a stronger sense of...

1 = being in a changed reality	2	3	4	5	6	7 = being in an unchanged reality

4) Consider your memory of being in the augmented environment. How similar are your visual memories of the displayed objects compared to memories of other objects you have seen today?
My memories of the displayed objects are similar to the memories of real objects that I've seen today...

1 = not at all	2	3	4	5	6	7 = very much

Score (mean):

Figure B.1: Paper-based AR presence questionnaire used in study on size variations.

AR presence questionnaire:

Note: This questionnaire was displayed in AR and filled out using a tangible pen-like object.

1) Please rate on a scale of 1 to 7 how much you felt you were in a place without visual overlays, where 7 is your normal feeling of being in a place.

1 = not at all 2 3 4 5 6 7 = very much

2) To what extent were there times during the experience when the visual overlays were reality for you?

1 = at no time 2 3 4 5 6 7 = almost all the time

3) During the time of the experience, which was the strongest on the whole, your sense of being in an unchanged reality or the feeling being in a changed reality?

1 = being in a changed reality 2 3 4 5 6 7 = being in an unchanged reality

4) Consider your memory of being in the augmented environment. How similar are your visual memories of the displayed objects compared to memories of other objects you have seen today?

1 = not at all similar 2 3 4 5 6 7 = very similar

Figure B.2: AR presence questionnaire used in studies on lighting variations and shape variations.

B.1.2 TAR Presence Questionnaires

Participant: Condition: Task:

TAR Presence Questionnaire:

1) Please rate on a scale of 1 to 7 how much you felt you were interacting with the visual overlays, where 7 is the feeling of actually interacting with the digital visualizations.
I felt I was interacting with the visual overlays in reality:

1 = not at all	2	3	4	5	6	7 = very much

2) To what extent were there times during the experience when the visual overlays felt real during the interaction?
There were times during the study when the visual overlays felt real:

1 = at no time	2	3	4	5	6	7 = almost all the time

3) During the time of the experience, which was the strongest on the whole, the feeling of interacting with the visual overlays or the feeling of interacting with the models below?
I had a stronger feeling that I was interacting with the...

1 = models below	2	3	4	5	6	7 = displayed overlays

4) Consider your time in the augmented environment. How similar are your memories of the interactions with the objects in the augmented environment compared to memories of interactions with other objects that you performed today?
My memories of interactions with the objects in the augmented environment are similar to the memories of interactions with real objects that I have performed today.

1 = not at all	2	3	4	5	6	7 = very much

Score (mean):

Figure B.3: Paper-based TAR questionnaire used in study on size variations.

TAR presence questionnaire:

Note: This questionnaire was displayed in AR and filled out using a tangible pen-like object.

1) Please rate on a scale of 1 to 7 how much you felt you were interacting with the visual overlays, where 7 is the feeling of actually interacting with the digital visualizations.

1 = not at all 2 3 4 5 6 7 = very much

2) To what extent were there times during the experience when the visual overlays felt real during the interaction?

1 = at no time 2 3 4 5 6 7 = almost all the time

3) During the time of the experience, which was the strongest on the whole, the feeling of interacting with the visual overlays or the feeling of interacting with the physical models below?

1 = interacting with physical models 2 3 4 5 6 7 = interacting with visual overlays

4) Consider your time in the augmented environment. How similar are your memories of the interactions with the objects in the augmented environment compared to memories of interactions with other objects that you performed today?

1 = not at all similar 2 3 4 5 6 7 = very similar

Figure B.4: TAR presence questionnaire used in studies on lighting variations and shape variations.

B.2 Usability Questionnaires

B.2.1 Size Perception Questionnaires

Participant: Condition: Task:

Size Perception Questionnaire

1) Please rate the size of the virtual objects in comparison to the corresponding physical models.
The virtual objects were:

1 much smaller	2	3	4 equal-sized	5	6	7 much larger

2) If you observed any difference in size, please rate how disturbing it was for grasping the objects. If you did not observe any difference, please choose 1.
The size difference was ... for grasping:

1 not disturbing at all	2	3	4	5	6	7 extremely disturbing

3) If you observed any difference in size, please rate how disturbing it was during the interaction with the objects. If you did not observe any difference, please choose 1.
The size difference was ... during interaction:

1 not disturbing at all	2	3	4	5	6	7 extremely disturbing

4) If you made any further observations, please describe them below.

Figure B.5: Paper-based size perception questionnaire.

Size perception questionnaire:

Note: This questionnaire was displayed in AR and filled out using a tangible pen-like object.

1) Please rate the size of the virtual overlays compared to the corresponding physical objects.

1 = much smaller virtual overlays	2	3	4 = equal-sized virtual overlays	5	6	7 = much larger virtual overlays
--	----------	----------	---	----------	----------	---

2) Please rate how disturbing it was to grasp the objects under the just experienced size condition.

1 = not disturbing at all	2	3	4	5	6	7 = extremely disturbing
----------------------------------	----------	----------	----------	----------	----------	---------------------------------

3) Please rate how disturbing it was to interact with the objects under the just experienced size condition.

1 = not disturbing at all	2	3	4	5	6	7 = extremely disturbing
----------------------------------	----------	----------	----------	----------	----------	---------------------------------

Figure B.6: Size perception questionnaire used in study on lighting variations.

B.2.2 Lighting Perception Questionnaire

Lighting questionnaire:

Note: This questionnaire was displayed in AR and filled out using a tangible pen-like object.

1) Please rate how natural you found the lighting.

1 = not natural at all	2	3	4	5	6	7 = extremely natural
-----------------------------------	----------	----------	----------	----------	----------	--------------------------------------

2) Please rate how pleasant you found the lighting.

1 = not pleasant at all	2	3	4	5	6	7 = extremely pleasant
--	----------	----------	----------	----------	----------	---------------------------------------

3) Please rate the transparency of the virtual overlays.

1 = not transparent at all	2	3	4	5	6	7 = extremely transparent
---	----------	----------	----------	----------	----------	--

4) Please rate how disturbing it was to grasp an object under the lighting condition you just experienced.

1 = not disturbing at all	2	3	4	5	6	7 = extremely disturbing
--	----------	----------	----------	----------	----------	---

5) Please rate how disturbing it was to interact with the objects under the lighting condition you just experienced.

1 = not disturbing at all	2	3	4	5	6	7 = extremely disturbing
--	----------	----------	----------	----------	----------	---

Figure B.7: Lighting perception questionnaire.

B.2.3 Shape Perception Questionnaire

Shape Perception Questionnaire

Note: This questionnaire was displayed in AR and filled out using a tangible pen-like object.

1) Please rate how good the shapes of the physical props fitted to the shapes of the virtual overlays.

1	2	3	4	5	6	7
not fitting at all						completely fitting

2) Please rate how disturbing it was to grasp the objects under the just experienced shape condition.

1	2	3	4	5	6	7
not disturbing at all						extremely disturbing

3) Please rate how disturbing it was to interact with the objects under the just experienced shape condition.

1	2	3	4	5	6	7
not disturbing at all						extremely disturbing

Figure B.8: Shape perception questionnaire.

B.3 Concluding Questionnaires

B.3.1 Size Variations Study

Participant: _____ Date: _____

Concluding Questionnaire

1. Gender

- m
 f
 d

2. Age: ____

3. Do you wear glasses or contact lenses?

- No, I don't need vision aids and I have normal vision
 Yes, I wear vision aids and I have corrected-to-normal vision
 I have an uncorrected vision defect: _____

4. Dominant hand?

- right
 left

5. I don't have a haptic perception disorder.

- Yes, this statement is correct
 No, I suffer from such a disorder

6. How often do you use Augmented Reality Technologies? (e.g. Smartphone Apps, AR-Glasses, etc.)

1 never	2 infrequently	3	4	5	6	7 regularly

7. How often do you use AR-Glasses?

1 never	2 infrequently	3	4	5	6	7 regularly

8. If you already have experience with Augmented Reality, please list the devices you used (e.g. Smartphone, Hololens, Google Glass, Magic Leap, etc.):

Participant: _____ Date: _____

9. How often do you work by hand? (e.g. with tools)

1 never	2 infrequently	3	4	5	6	7 regularly

10. Did you feel sick after your time in the augmented reality?

1 not at all	2	3	4	5	6	7 very sick

11. During the experiment you encountered seven levels of size difference between the real and corresponding virtual objects. These conditions are listed below on the left. On the right, please rank these conditions according to how real the interaction felt with this size difference. Start with highest realism as 1. and end with lowest realism as 7.

very much smaller considerably smaller slightly smaller matching slightly larger considerably larger very much larger	} } } } } } }	virtual size	_____ 1. _____ 2. _____ 3. _____ 4. _____ 5. _____ 6. _____ 7. _____
---	---------------------------------	--------------	--

12. Please indicate in your ranking in 11. up to which condition you would describe the interaction as pleasant. For example, mark the border between the two conditions where you would draw the line.

Participant: _____ Date: _____

13. Now think about how easy it was to solve the tasks with each of the size differences and rank them accordingly below. Start with the easiest one as 1. and end with the most challenging as 7.

very much smaller	}	virtual size	_____
considerably smaller			1.
slightly smaller			2.
matching			3.
slightly larger			4.
considerably larger			5.
very much larger			6.
			7.

14. Please indicate in your ranking in 13. up to which condition you would describe the interaction as efficient. For example, mark the border between the two conditions where you would draw the line.

15. Further comments:

Figure B.9: Concluding questionnaire of the size variations study.

B.3.2 Lighting Variations Study

Participant: _____ Date: _____

Concluding Questionnaire

1. Gender

- m
 f
 d

2. Age: ____

3. Do you wear glasses or contact lenses?

- No, I don't need vision aids and I have normal vision
 Yes, I wear vision aids and I have corrected-to-normal vision
 I have an uncorrected vision defect: _____

4. Dominant hand?

- right
 left

5. I don't have a haptic perception disorder.

- Yes, this statement is correct
 No, I suffer from such a disorder

6. How often do you use Augmented Reality Technologies? (e.g. Smartphone Apps, AR-Glasses, etc.)

1 never	2 infrequently	3	4	5	6	7 regularly

7. How often do you use AR-Glasses?

1 never	2 infrequently	3	4	5	6	7 regularly

8. If you already have experience with Augmented Reality, please list the devices you used (e.g. Smartphone, Hololens, Google Glass, Magic Leap, etc.):

Participant: Date:

9. How often do you work by hand? (e.g. with tools)

1 never	2 infrequently	3	4	5	6	7 regularly

10. Did you feel sick after your time in the augmented reality?

1 not at all	2	3	4	5	6	7 very sick

During the experiment you experienced three different lighting conditions, which led to three different degrees of transparency of the virtual overlays listed below:

- A: very transparent / bright lighting
- B: medium transparent / medium lighting
- C: (nearly) not transparent / dark lighting

11. Please rank the condition where the interaction felt most real with 1 and the condition where the interaction felt least real with 3.

1. _____ 2. _____ 3. _____

12. Please rate the condition where the interaction was the easiest with 1 and the condition where the interaction was the least easy with 3.

1. _____ 2. _____ 3. _____

13. Please rate the condition you liked most with 1 and the condition you liked least with 3.

1. _____ 2. _____ 3. _____

Participant: Date:

14. Further observations regarding lighting/transparency:

15. Further comments:

Figure B.10: Concluding questionnaire of the lighting variations study.

B.3.3 Shape Variations Study

Participant: _____ Date: _____

Concluding Questionnaire

1. Gender

- m
 f
 d

2. Age: ____

3. Do you wear glasses or contact lenses?

- No, I don't need vision aids and I have normal vision
 Yes, I wear vision aids and I have corrected-to-normal vision
 I have an uncorrected vision defect: _____

4. Dominant hand?

- right
 left

5. I don't have a haptic perception disorder.

- Yes, this statement is correct
 No, I suffer from such a disorder

6. How often do you use Augmented Reality Technologies? (e.g. Smartphone Apps, AR-Glasses, etc.)

1 never	2 infrequently	3	4	5	6	7 regularly

7. How often do you use AR-Glasses?

1 never	2 infrequently	3	4	5	6	7 regularly

8. If you already have experience with Augmented Reality, please list the devices you used (e.g. Smartphone, Hololens, Google Glass, Magic Leap, etc.):

Participant: _____ Date: _____


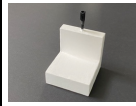
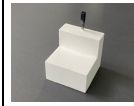


9. How often do you work by hand? (e.g. with tools)

1 never	2 infrequently	3	4	5	6	7 regularly

10. Did you feel sick after your time in the augmented reality?

1 not at all	2	3	4	5	6	7 very sick

During the experiment you experienced five different shape conditions:

A 3D Model	B Abstracted 3D Model	C Abstraction of a Standard Sofa	D Cuboid	E Plane
				

11. Please rank the condition where the interaction felt most real with 1 and the condition where the interaction felt least real with 5.

1. _____ 2. _____ 3. _____ 4. _____ 5. _____

12. Please rate the condition where the interaction was the easiest with 1 and the condition where the interaction was the least easy with 5.

1. _____ 2. _____ 3. _____ 4. _____ 5. _____

Participant: _____ Date: _____

13. Please rate the condition you liked most with 1 and the condition you liked least with 5.

1. _____ 2. _____ 3. _____ 4. _____ 5. _____

14. Further observations regarding shape variations:

15. Further comments:

Figure B.11: Concluding questionnaire of the shape variations study.

List of Figures

1.1	Digital advertising in Tokyo causing visual information overload.	2
1.2	Presentation of the structure of the thesis.	8
2.1	The human eye with the cornea, lens, optic nerve, retina, and fovea. The central human retina (macula) with an idealized view of its blood vessels and the optic disc.	12
2.2	Distribution of rods and cones on the retina.	13
2.3	Illustration of the human visual field.	15
2.4	Illustration of fixations and saccades.	19
2.5	Experimental conditions investigated by Seeliger et al.	21
2.6	Attention funnels directing the gaze of the user to the red target objects.	22
2.7	Visualization of several modulation thresholds used by Veas et al.	23
2.8	Sample image of the experiment by Hata et al.	24
2.9	Example of the luminance modulation.	25
2.10	Gaze distribution for an unmodulated and modulated image. . .	26
2.11	Study design of Grogorick et al.	29
2.12	Different contrast levels investigated by Lu et al.	30
2.13	The Reality-Virtuality Continuum by Milgram and Colquhoun. .	31
2.14	The TAR controller applications <i>Magic Paddle</i> and <i>MagicCup</i>	32
2.15	Example TAR applications: Cubical User Interface, TAR for prod- uct usability assessment, and Augmented Foam.	33
2.16	Visuo-haptic Reality-Virtuality Continuum developed by Jeon and Choi.	34
2.17	Layer of the skin with four different mechanoreceptors.	37
2.18	Virtual substitutions of the mug and physical props used in the study of Simeone et al.	40
2.19	Study setup of the experiment by de Tinguy et al.	41
2.20	Physical props used in the study of Kobeisse and Holmquist. . . .	42
2.21	Virtually overlaid everyday proxy objects determined by the algo- rithm of Hettiarchichi and Wigdor.	43
3.1	Illustration of the study task on the TV screen.	49
3.2	Illustration of the peripheral and foveal tasks.	51

3.3	Visualization of the distances from which the objects were seen (blue), the color was correctly recognized (red; only colored primitives) or the details were correctly recognized (green).	52
3.4	Presentation of the results of the method of limits for determining the cue intensity.	55
3.5	Example of one rendered shelf used in the study.	57
3.6	Task completion times for all records and only for records with correct answers in study part 1 and study part 2.	61
3.7	Correctly solved tasks in study part 1 and study part 2.	63
3.8	Task completion times and correctly solved tasks grouped by visual cue stimuli and by sensor mode.	69
3.9	Ratings regarding helpful assistance, disturbance, favor and everyday usability of the visual cues.	70
3.10	NASA-TLX ratings of the tasks without visual cues, with static visual cues and with adaptive visual cues for assistance.	77
3.11	Illustration of the study setup.	80
3.12	Visualization of the red modulation spots and the grey areas where no modulations should be displayed and an example image with the modulation spot placed at the left door.	83
3.13	Number of fixations of the target, total fixation time and time to first fixation grouped by visual cue stimuli.	85
3.14	Heatmap visualization for the different modulation variants.	87
4.1	Participant's perspective for part 1 of the study with size condition M. HoloLens view of study part 2 with size condition XS.	98
4.2	Study setup: The participant's interaction area positioned at the center of the tracking zone.	99
4.3	Physical props with attached marker trees on top.	100
4.4	Size variations of the virtual overlays (white) compared to the physical proxy objects (black).	101
4.5	Shapes of the tangible objects.	102
4.6	Ranges of size conditions without significant difference from baseline condition M for each measure, divided into the two study parts.	105
4.7	Task completion times in seconds for each size condition in part 2 of the study.	107
4.8	Screenshots of the HoloLens 2 during the execution of the study in size conditions XL and XXS.	114

4.9	Illustrations of the laboratory setup in our three lighting conditions Dark, Medium and Bright.	115
4.10	Physical props with attached marker trees on top.	116
4.11	Size variations of the virtual overlays (blue) compared to the physical proxy objects (white).	118
4.12	Summary of significant differences of the size conditions compared to the size-matching condition M as a baseline.	120
4.13	Mean transparency ratings regarding the overlays, mean task completion times in seconds and naturalness and pleasantness of the environmental lighting with marked standard errors for each lighting condition.	121
4.14	Mean presence and disturbance ratings from 1 to 7 with marked standard errors for each lighting condition.	122
4.15	HoloLens screenshots of study part 1 in condition B : Matching virtual sofa parts (blue) to appropriate 3D targets (orange).	130
4.16	HoloLens screenshots of study part 2 in condition D : Assembly of the virtual sofa parts (blue) according to a 3D miniature template (turquoise).	130
4.17	Two-seater in shape condition A with attached marker tree.	132
4.18	Investigated shape variations of the physical proxy object: matching 3D model (A), abstracted 3D model (B), abstraction of a standard sofa (C), cuboid (D), plane (E).	133
4.19	Sofa parts used in the study.	134
4.20	Summary of significant differences of the shape conditions compared to the shape-matching condition A as a baseline.	136
4.21	Mean AR presence and TAR presence ratings from 1 to 7 with marked standard errors for each shape condition.	137
4.22	Mean disturbance ratings from 1 to 7 with marked standard errors for each shape condition.	138
5.1	Main components of the framework for adaptive gaze guidance.	144
5.2	Detailed architecture of the gaze guidance framework.	146
5.3	Pupil Pro eye tracking headset. Example OptiTrack motion capturing setup.	147
5.4	Representation of the shelf, which was created to guide the gaze.	148
5.5	Web interface for controlling the prototypically implemented framework realized as a smart attention directing shelf.	151
5.6	Main components of the framework for AR proxy interaction.	152

5.7	Detailed architecture of the framework for AR proxy interaction.	155
5.8	Example of a marker tree.	156
5.9	Screenshot of the graphical user interface of the Experiment Server for the lighting variations study.	158
5.10	HoloLens 2 equipped with reflective markers.	159
5.11	Physical example object for participant's eye calibration.	161
5.12	Adjustment of the FOV to the interaction area.	162
5.13	Physical props covered with a box in front of the participant.	163
5.14	Excerpt from the in-AR questionnaire.	165
A.1	Landscape images used in the study on investigations of subtle gaze guidance in AR.	177
B.1	Paper-based AR presence questionnaire used in study on size variations.	180
B.2	AR presence questionnaire used in studies on lighting variations and shape variations.	181
B.3	Paper-based TAR questionnaire used in study on size variations.	182
B.4	TAR presence questionnaire used in studies on lighting variations and shape variations.	183
B.5	Paper-based size perception questionnaire.	184
B.6	Size perception questionnaire used in study on lighting variations.	185
B.7	Lighting perception questionnaire.	186
B.8	Shape perception questionnaire.	187
B.9	Concluding questionnaire of the size variations study.	190
B.10	Concluding questionnaire of the lighting variations study.	193
B.11	Concluding questionnaire of the shape variations study.	196

List of Tables

3.1	Measured task completion times in part 1 and part 2 of the study. Presentation of sample sizes, means, and standard deviations in the three test conditions for all records and only for the records with correct answers.	60
3.2	Correctly solved tasks in part 1 and part 2 of the study. Presentation of sample sizes, means, and standard deviations in the three test conditions.	62
3.3	Measured task processing times overall and separately for the sensor types <i>Eye Tracking</i> and <i>Motion Capturing</i> . Sample sizes, means and standard deviations are shown for the three visual cue conditions.	68
3.4	Measured task completion times. Sample sizes, means, and standard deviations for the two sensor types <i>Eye Tracking</i> and <i>Motion Capturing</i>	69
3.5	Means and standard deviations of the task completion times in the three test conditions.	76
3.6	Percentages of successful target viewing, unrecognized cue stimuli in the successful trials and overall undetected stimuli.	86
4.1	Ranking scores of the size conditions for realism and easiness. . .	108
4.2	Number of classifications for each condition as pleasant and efficient (out of 13).	108
4.3	Lighting intensity in lx , measured on the tabletop pointing upwards and at the HoloLens camera pointing towards the interaction area in the different lighting conditions.	115
4.4	Scores for each size condition in realism, easiness and preference according to participants' rankings.	125
4.5	Dimensions of seat heights and backrest depths of the different shape variations measured in centimeters.	133
4.6	Summed scores for each shape condition regarding realism, easiness and preference obtained by the rankings of all participants. .	139

List of Abbreviations

- AR** Augmented Reality
- BLE** Bluetooth Low Energy
- EBS** Event Broadcasting Service
- FOV** Field of View
- GNSS** Global Navigation Satellite System
- HCI** Human-Computer Interaction
- IMU** Inertial Measurement Unit
- LIDAR** Light Detection and Ranging
- LS** Latin square
- MQTT** Message Queuing Telemetry Transport
- OST AR** Optical See-through Augmented Reality
- RFID** Radio Frequency Identification
- SGD** Subtle Gaze Direction
- SLAM** Simultaneous Localization and Mapping
- TAR** Tangible Augmented Reality
- VR** Virtual Reality
- VST AR** Video See-through Augmented Reality

Bibliography

- [1] F. N. Afif, A. H. Basori, and N. Saari. Vision-based Tracking Technology for Augmented Reality: A Survey. *International Journal of Interactive Digital Media*, 1(1):46–49, 2013.
- [2] D. Alexandrovsky, S. Putze, V. Schwind, E. D. Mekler, J. D. Smeddinck, D. Kahl, A. Krüger, and R. Malaka. Evaluating User Experiences in Mixed Reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–5, 2021.
- [3] U. Ansorge and H. Leder. Wahrnehmung und Aufmerksamkeit. In *Wahrnehmung und Aufmerksamkeit*, pp. 9–25. Springer, 2011.
- [4] J. Aulinas, Y. Petillot, J. Salvi, and X. Lladó. The SLAM Problem: A Survey. *Artificial Intelligence Research and Development*, pp. 363–371, 2008.
- [5] R. T. Azuma. A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments*, 6(4):355–385, 1997.
- [6] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle Gaze Direction. *ACM Transactions on Graphics (TOG)*, 28(4):1–14, 2009.
- [7] G. Ballestin, M. Chessa, and F. Solari. A Registration Framework for the Comparison of Video and Optical See-through Devices in Interactive Augmented Reality. *IEEE Access*, 9:64828–64843, 2021.
- [8] J. Bauer, R. Wichert, C. Konrad, M. Hechtel, S. Dengler, S. Uhrmann, M. Ge, P. Poller, D. Kahl, B. Ristok, et al. ForeSight – User-centered and Personalized Privacy and Security Approach for Smart Living. In *International Conference on Human-Computer Interaction*, pp. 18–36. Springer, 2022.
- [9] J. Bergström, A. Mottelson, and J. Knibbe. Resized Grasping in VR: Estimating Thresholds for Object Discrimination. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 1175–1183, 2019.

- [10] L. Besançon, P. Issartel, M. Ammi, and T. Isenberg. Mouse, Tactile, and Tangible Input for 3D Manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 4727–4740, 2017.
- [11] M. Billinghurst, A. Clark, G. Lee, et al. A Survey of Augmented Reality. *Foundations and Trends in Human–Computer Interaction*, 8(2-3):73–272, 2015.
- [12] M. Billinghurst, H. Kato, and I. Poupyrev. Tangible Augmented Reality. *ACM SIGGRAPH ASIA*, 7(2):1–10, 2008.
- [13] F. Biocca, A. Tang, C. Owen, and F. Xiao. Attention Funnel: Omnidirectional 3D Cursor for Mobile Augmented Reality Platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1115–1122, 2006.
- [14] T. Booth, S. Sridharan, A. McNamara, C. Grimm, and R. Bailey. Guiding Attention in Controlled Real-world Environments. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 75–82, 2013.
- [15] E. Bozgeyikli and L. L. Bozgeyikli. Evaluating Object Manipulation Interaction Techniques in Mixed Reality: Tangible User Interfaces and Gesture. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 778–787. IEEE, 2021.
- [16] M. Burke, A. Hornof, E. Nilsen, and N. Gorman. High-cost Banner Blindness: Ads Increase Perceived Workload, Hinder Visual Search, and Are Forgotten. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(4):423–445, 2005.
- [17] B. T. Carter and S. G. Luke. Best Practices in Eye Tracking Research. *International Journal of Psychophysiology*, 155:49–62, 2020.
- [18] Y. S. Chang, B. Nuernberger, B. Luan, and T. Höllerer. Evaluating Gesture-based Augmented Reality Annotation. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 182–185. IEEE, 2017.
- [19] P. Chen, X. Liu, W. Cheng, and R. Huang. A Review of Using Augmented Reality in Education from 2011 to 2016. In *Innovations in Smart Learning*, pp. 13–18. Springer Singapore, 2017.
- [20] Y. M. Choi. Applying Tangible Augmented Reality for Product Usability Assessment. *Journal of Usability Studies*, 14(4):187–200, 2019.

- [21] X. de Tinguy, C. Pacchierotti, M. Emily, M. Chevalier, A. Guignardat, M. Guillaudeux, C. Six, A. Lécuyer, and M. Marchal. How Different Tangible and Virtual Objects Can Be While Still Feeling the Same? In *2019 IEEE World Haptics Conference (WHC)*, pp. 580–585. IEEE, 2019.
- [22] M. Dorr, T. Martinetz, K. Gegenfurtner, E. Barth, et al. Guidance of Eye Movements on a Gaze-contingent Display. In *Dynamic Perception Workshop of the GI Section "Computer Vision"*, pp. 89–94, 2004.
- [23] T. Düwel, N. Herbig, D. Kahl, and A. Krüger. Combining Embedded Computation and Image Tracking for Composing Tangible Augmented Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2020.
- [24] T. Düwel. TAROC: A Tangible Augmented Reality System for Object Configuration. Bachelor’s thesis, Saarland University, Germany, 2018.
- [25] D. Englmeier, J. Dörner, A. Butz, and T. Höllerer. A Tangible Spherical Proxy for Object Manipulation in Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 221–229. IEEE, 2020.
- [26] M. J. Eppler and J. Mengis. The Concept of Information Overload - A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines (2004). *Kommunikationsmanagement im Wandel*, pp. 271–305, 2008.
- [27] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. Exploring the Limitations of Environment Lighting on Optical See-through Head-mounted Displays. In *Symposium on Spatial User Interaction*, pp. 1–8. ACM, 2020.
- [28] M. W. Eysenck and M. T. Keane. *Cognitive Psychology: A Student’s Handbook*. Psychology Press, 2015.
- [29] M. Feick, N. Kleer, A. Tang, and A. Krüger. The Virtual Reality Questionnaire Toolkit. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 68–69, 2020.
- [30] B. W. Fitzgerald. Using Hawkeye from the Avengers to Communicate on the Eye, 2018.
- [31] O. Fried, E. Shechtman, D. B. Goldman, and A. Finkelstein. Finding Distractors in Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1703–1712, 2015.

- [32] J. L. Gabbard, J. E. Swan, and D. Hix. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence: Virtual and Augmented Reality*, 15(1):16–32, 2006.
- [33] J. L. Gabbard, J. E. Swan, D. Hix, S.-J. Kim, and G. Fitch. Active Text Drawing Styles for Outdoor Augmented Reality: A User-based Study and Design Implications. In *2007 IEEE Virtual Reality Conference*, pp. 35–42. IEEE, 2007.
- [34] J. L. Gabbard, J. E. Swan, D. Hix, R. S. Schulman, J. Lucas, and D. Gupta. An Empirical User-based Study of Text Drawing Styles and Outdoor Background Textures for Augmented Reality. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 11–18. IEEE, 2005.
- [35] M. A. Garcia-Pérez. Forced-choice Staircases with Fixed Step Sizes: Asymptotic and Small-sample Properties. *Vision Research*, 38(12):1861–1881, 1998.
- [36] J. W. Garrett. The Adult Human Hand: Some Anthropometric and Biomechanical Considerations. *Human Factors*, 13(2):117–131, 1971.
- [37] L. A. Gatys, M. Kümmerer, T. S. Wallis, and M. Bethge. Guiding Human Gaze with Convolutional Neural Networks. *arXiv Preprint arXiv:1712.06492*, 2017. Available at: <http://arxiv.org/abs/1712.06492> (last accessed: 2023-08-15).
- [38] Y. Georgiou and E. A. Kyza. The Development and Validation of the ARI Questionnaire: An Instrument for Measuring Immersion in Location-based Augmented Reality Settings. *International Journal of Human-Computer Studies*, 98:24–37, 2017.
- [39] J. J. Gibson. Adaptation, After-effect and Contrast in the Perception of Curved Lines. *Journal of Experimental Psychology*, 16(1):1–31, 1933.
- [40] D. Goldreich and I. M. Kanics. Tactile Acuity is Enhanced in Blindness. *Journal of Neuroscience*, 23(8):3439–3445, 2003.
- [41] E. B. Goldstein. *Blackwell Handbook of Sensation and Perception*. John Wiley & Sons, 2008.
- [42] E. B. Goldstein and L. Cacciamani. *Sensation and Perception*. Cengage Learning, 2021.

- [43] S. Grogorick, G. Albuquerque, and M. A. Magnor. Comparing Unobtrusive Gaze Guiding Stimuli in Head-Mounted Displays. In *ICIP*, pp. 2805–2809, 2018.
- [44] S. Grogorick, G. Albuquerque, J.-P. Tauscher, and M. Magnor. Comparison of Unobtrusive Visual Guidance Methods in an Immersive Dome Environment. *ACM Transactions on Applied Perception (TAP)*, 15(4):1–11, 2018.
- [45] S. Grogorick, M. Stengel, E. Eisemann, and M. Magnor. Subtle Gaze Guidance for Immersive Environments. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 1–7, 2017.
- [46] S. Guest and C. Spence. Tactile Dominance in Speeded Discrimination of Textures. *Experimental Brain Research*, 150:201–207, 2003.
- [47] A. Hagiwara, A. Sugimoto, and K. Kawamoto. Saliency-based Image Editing for Guiding Visual Attention. In *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-based Interaction*, pp. 43–48, 2011.
- [48] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, vol. 52, pp. 139–183. Elsevier, 1988.
- [49] J. Hartcher-O’Brien, C. Levitan, and C. Spence. Extending Visual Dominance over Touch for Input off the Body. *Brain Research*, 1362:48–55, 2010.
- [50] H. Hata, H. Koike, and Y. Sato. Visual Guidance with Unnoticed Blur Effect. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 28–35, 2016.
- [51] A. F. Healy and R. W. Proctor. *Handbook of psychology: Experimental psychology, Vol. 4*. John Wiley & Sons Inc, 2003.
- [52] C. Heeter. Being There: The Subjective Experience of Presence. *Presence: Teleoperators & Virtual Environments*, 1(2):262–271, 1992.
- [53] H. B. Helbig and M. O. Ernst. Optimal Integration of Shape Information from Vision and Touch. *Experimental Brain Research*, 179:595–606, 2007.
- [54] A. Hendrickson. Organization of the Adult Primate Fovea. In *Macular Degeneration*, pp. 1–23. Springer, 2005.

- [55] A. Hettiarachchi and D. Wigdor. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1957–1967, 2016.
- [56] H. G. Hoffman. Physically Touching Virtual Objects Using Tactile Augmentation Enhances the Realism of Virtual Environments. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No. 98CB36180)*, pp. 59–63. IEEE, 1998.
- [57] Y. Huang. Evaluating Mixed Reality Technology for Architectural Design and Construction Layout. *Journal of Civil Engineering and Construction Technology*, 11(1):1–12, 2020.
- [58] iGroup. iGroup Presence Questionnaire. Retrieved June 30, 2023 from <http://www.igroup.org/pq/ipq/index.php>.
- [59] H. Ishii and B. Ullmer. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *CHI '97 Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 234–241. ACM, 1997.
- [60] W. James, F. Burkhardt, F. Bowers, and I. K. Skrupskelis. *The Principles of Psychology*, vol. 1. Macmillan London, 1890.
- [61] S. Jeon and S. Choi. Haptic Augmented Reality: Taxonomy and an Example of Stiffness Modulation. *Presence*, 18(5):387–408, 2009.
- [62] P. Jonczyk. Adaptive Subtle Gaze Direction in Augmented Reality. Master’s thesis, Saarland University, Germany, 2023.
- [63] J. Jonides. Voluntary Versus Automatic Control Over the Mind’s Eye’s Movement. In J. B. Long and A. Baddeley, eds., *Attention and Performance*, vol. IX, pp. 187–204. Erlbaum, 1981.
- [64] M. A. Just and P. A. Carpenter. A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review*, 87(4):329–354, 1980.
- [65] D. Kahl and A. Krüger. Using Abstract Tangible Proxy Objects for Interaction in Optical See-through Augmented Reality. *arXiv Preprint arXiv:2308.05836*, 2023. Available at: <http://arxiv.org/abs/2308.05836> (last accessed: 2023-08-15).

- [66] D. Kahl, M. Ruble, and A. Krüger. Evaluating User Experience in Tangible Augmented Reality. In *Workshop on Evaluating User Experiences in Mixed Reality at CHI '21*. ACM, 2021. Available at: https://www-live.dfki.de/fileadmin/user_upload/import/11565_CHI2021_W13_Kahl.pdf (last accessed: 2023-08-15).
- [67] D. Kahl, M. Ruble, and A. Krüger. Identification of Everyday Proxies for Tangible Augmented Reality. In *Workshop on Everyday Proxy Objects for Virtual Reality at CHI '21*. ACM, 2021. Available at: https://www-live.dfki.de/fileadmin/user_upload/import/11625__CH2021_WS27_Kahl.pdf (last accessed: 2023-08-15).
- [68] D. Kahl, M. Ruble, and A. Krüger. Investigation of Size Variations in Optical See-through Tangible Augmented Reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 147–155. IEEE, 2021.
- [69] D. Kahl, M. Ruble, and A. Krüger. The Influence of Environmental Lighting on Size Variations in Optical See-through Tangible Augmented Reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 121–129. IEEE, 2022.
- [70] G. Kahl, C. Bürckert, L. Spassova, and T. Schwartz. Event Broadcasting Service - An Event-Based Communication Infrastructure. In *Workshop on Location Awareness for Mixed and Dual Reality at IUI '12*. ACM, 2012. Available at: https://www.dfki.de/LAMDa/accepted/Event_Broadcasting_Service.pdf (last accessed: 2023-08-15).
- [71] G. Kahl and D. Paradowski. A Privacy-aware Shopping Scenario. In *Proceedings of the Companion Publication of the 2013 International Conference on Intelligent User Interfaces (IUI)*, pp. 107–108. ACM, 2013.
- [72] M. Kassner, W. Patera, and A. Bulling. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1151–1160, 2014.
- [73] H. Kato, K. Tachibana, M. Tanabe, T. Nakajima, and Y. Fukuda. Magic-Cup: A Tangible Interface for Virtual Objects Manipulation in Table-top Augmented Reality. In *2003 IEEE International Augmented Reality Toolkit Workshop*, pp. 75–76. IEEE, 2003.

- [74] T. Kawashima, K. Imamoto, H. Kato, K. Tachibana, and M. Billingham. Magic Paddle: A Tangible Augmented Reality Interface for Object Manipulation. In *Proc. of ISMR2001*, pp. 194–195, 2001.
- [75] J. L. Kerr. Visual Resolution in the Periphery. *Perception & Psychophysics*, 9(3):375–378, 1971.
- [76] A. Khan, J. Matejka, G. Fitzmaurice, and G. Kurtenbach. Spotlight: Directing Users’ Attention on Large Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 791–798, 2005.
- [77] R. L. Klatzky, S. J. Lederman, and V. A. Metzger. Identifying Objects by Touch: An “Expert System”. *Perception & Psychophysics*, 37:299–302, 1985.
- [78] N. Kleitman and Z. A. Blier. Color and Form Discrimination in the Periphery of the Retina. *American Journal of Physiology – Legacy Content*, 85(2):178–190, 1928.
- [79] S. Kobeisse and L. E. Holmquist. “I Can Feel It in My Hand”: Exploring Design Opportunities for Tangible Interfaces to Manipulate Artefacts in AR. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*, pp. 28–36, 2022.
- [80] C. Koch and S. Ullman. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. In *Matters of Intelligence*, pp. 115–141. Springer, 1987.
- [81] H. Kolb, R. F. Nelson, P. K. Ahnelt, I. Ortuño-Lizarán, and N. Cuenca. The Architecture of the Human Fovea. *Webvision: The Organization of the Retina and Visual System*, 2020. Available at: <https://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/the-architecture-of-the-human-fovea/> (last accessed: 2023-08-15).
- [82] H. Koyuncu and S. H. Yang. A Survey of Indoor Positioning and Object Locating Systems. *IJCSNS International Journal of Computer Science and Network Security*, 10(5):121–128, 2010.
- [83] R. Krueger, S. Koch, and T. Ertl. SaccadeLenses: Interactive Exploratory Filtering of Eye Tracking Trajectories. In *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)*, pp. 31–34. IEEE, 2016.

- [84] A. Krüger, W. Maaß, D. Paradowski, and S. Jansen. Empfehlungssysteme und integrierte Informationsdienste zur Steigerung der Wertschöpfung im stationären Handel. In W. Reinartz and M. Käuferle, eds., *Wertschöpfung im Handel*, pp. 273–292. Kohlhammer Verlag, 2014.
- [85] E. Kwon, G. J. Kim, and S. Lee. Effects of Sizes and Shapes of Props in Tangible Augmented Reality. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pp. 201–202. IEEE, 2009.
- [86] C. Lander, M. Speicher, D. Paradowski, N. Coenen, S. Biewer, and A. Krüger. Collaborative Newspaper: Exploring an Adaptive Scrolling Algorithm in a Multi-user Reading Scenario. In *Proceedings of the 4th International Symposium on Pervasive Displays*, pp. 163–169, 2015.
- [87] N. Lavie. Distracted and Confused?: Selective Attention Under Load. *Trends in Cognitive Sciences*, 9(2):75–82, 2005.
- [88] N. Lavie. Attention, Distraction, and Cognitive Control Under Load. *Current Directions in Psychological Science*, 19(3):143–148, 2010.
- [89] S. J. Lederman. Auditory Texture Perception. *Perception*, 8(1):93–103, 1979.
- [90] S. J. Lederman and R. L. Klatzky. Hand Movements: A Window into Haptic Object Recognition. *Cognitive Psychology*, 19(3):342–368, 1987.
- [91] S. J. Lederman and R. L. Klatzky. Multisensory Texture Perception. 2004.
- [92] S. J. Lederman and R. L. Klatzky. Haptic Perception: A Tutorial. *Attention, Perception, & Psychophysics*, 71(7):1439–1459, 2009.
- [93] S. J. Lederman, G. Thorne, and B. Jones. Perception of Texture by Vision and Touch: Multidimensionality and Intersensory Integration. *Journal of Experimental Psychology: Human Perception and Performance*, 12(2):169, 1986.
- [94] H. Lee, M. Billinghamurst, and W. Woo. Two-handed Tangible Interaction Techniques for Composing Augmented Blocks. *Virtual Reality*, 15(2-3):133–146, 2011.
- [95] W. Lee and J. Park. Augmented Foam: A Tangible Augmented Reality for Product Design. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pp. 106–109. IEEE, 2005.

- [96] P. H. Lindsay and D. A. Norman. *Human Information Processing: An Introduction to Psychology*. Academic Press, 2013.
- [97] J. Liu, G. Huang, J. Hyyppä, J. Li, X. Gong, and X. Jiang. A Survey on Location and Motion Tracking Technologies, Methodologies and Applications in Precision Sports. *Expert Systems with Applications*, 2023. Article 120492.
- [98] Y. Liu, H. Dong, L. Zhang, and A. El Saddik. Technical Evaluation of HoloLens for Multimedia: A First Look. *IEEE MultiMedia*, 25(4):8–18, 2018.
- [99] W. Lu, B.-L. H. Duh, and S. Feiner. Subtle Cueing for Visual Search in Augmented Reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 161–166. IEEE, 2012.
- [100] W. Lu, D. Feng, S. Feiner, Q. Zhao, and H. B.-L. Duh. Evaluating Subtle Cueing in Head-worn Displays. In *Proceedings of the Second International Symposium of Chinese CHI*, pp. 5–10, 2014.
- [101] O. Lukashova-Sanz and S. Wahl. Saliency-Aware Subtle Augmentation Improves Human Visual Search Performance in VR. *Brain Sciences*, 11(3), 2021. Article 283.
- [102] M. Löchtefeld, S. Gehring, D. Paradowski, and A. Krüger. Filtered Reality – Keeping Your Peripheral Vision Clean. In *Workshop on Peripheral Interaction: Shaping the Research and Design Space at CHI 2014*. ACM, 2014. Available at: https://www.dfki.de/fileadmin/user_upload/import/7810_FilteredReality.pdf (last accessed: 2023-08-15).
- [103] A. Mack and I. Rock. Inattentional Blindness: Perception Without Attention. *Visual Attention*, 8:55–76, 1998.
- [104] C. L. MacKenzie. From Manipulation to Goal-directed Human Activities in Virtual and Augmented Environments. In *ICAT '99 Proceedings of the Ninth International Conference of Artificial Reality and Telexistence*, pp. 6–8. The Virtual Reality Society of Japan, Tokyo, 1999.
- [105] R. Malaka, A. Butz, and H. Hußmann. *Medieninformatik: Eine Einführung*. Pearson Deutschland GmbH, 2009.

- [106] C. Malewski, J. Wiesmann, D. Paradowski, and A. Grote. Fächerverbindendes Arbeiten mit Karten-APIs. In T. Bartoschek and J. Schubert, eds., *Geoinformation im Geographieunterricht*, pp. 158–175. Monsenstein und Vannerdat, 2013.
- [107] V. A. Mateescu and I. V. Bajić. Attention Retargeting by Color Manipulation in Images. In *Proceedings of the 1st International Workshop on Perception Inspired Video Processing*, pp. 15–20, 2014.
- [108] E. Matin. Saccadic Suppression: A Review and an Analysis. *Psychological Bulletin*, 81(12):899–917, 1974.
- [109] A. McNamara, R. Bailey, and C. Grimm. Search Task Performance Using Subtle Gaze Direction with the Presence of Distractions. *ACM Transactions on Applied Perception (TAP)*, 6(3):1–19, 2009.
- [110] A. McNamara, T. Booth, S. Sridharan, S. Caffey, C. Grimm, and R. Bailey. Directing Gaze in Narrative Art. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 63–70, 2012.
- [111] R. Mechrez, E. Shechtman, and L. Zelnik-Manor. Saliency Driven Image Manipulation. *Machine Vision and Applications*, 30(2):189–202, 2019.
- [112] E. Mendez, S. Feiner, and D. Schmalstieg. Focus and Context in Mixed Reality by Modulating First Order Salient Features. In *International Symposium on Smart Graphics*, pp. 232–243. Springer, 2010.
- [113] P. Milgram, H. Colquhoun, et al. A Taxonomy of Real and Virtual World Display Integration. *Mixed Reality: Merging Real and Virtual Worlds*, 1(1999):1–26, 1999.
- [114] P. Milgram and F. Kishino. A Taxonomy of Mixed Reality Visual Displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.
- [115] S. Montabone and A. Soto. Human Detection Using a Mobile Platform and Novel Features Derived from a Visual Saliency Mechanism. *Image and Vision Computing*, 28(3):391–402, 2010.
- [116] A. Morris, J. Goodson, and NAVAL AEROSPACE MEDICAL RESEARCH LAB PENSACOLA FL. The Development of a Precision Series of Landolt Ring Acuity Slides. *NAMRL Report*, 1983. Report No. 1303.

- [117] J. Müller, D. Wilmsmann, J. Exeler, M. Buzeck, A. Schmidt, T. Jay, and A. Krüger. Display Blindness: The Effect of Expectations on Attention Towards Digital Signage. In *International Conference on Pervasive Computing*, pp. 1–8. Springer, 2009.
- [118] B. Munsinger, G. White, and J. Quarles. The Usability of the Microsoft HoloLens for an Augmented Reality Game to Teach Elementary School Children. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–4. IEEE, 2019.
- [119] N. C. Nilsson, A. Zenner, A. L. Simeone, D. Degraen, and F. Daiber. Haptic Proxies for Virtual Reality: Success Criteria and Taxonomy. In *Workshop on Everyday Proxy Objects for Virtual Reality at CHI '21*. ACM, 2021.
- [120] C. B. Owen, J. Zhou, A. Tang, and F. Xiao. Display-relative Calibration for Optical See-through Head-mounted Displays. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 70–78. IEEE, 2004.
- [121] D. Paradowski and A. Krüger. Modularization of Mobile Shopping Assistance Systems. In *2013 5th International Workshop on Near Field Communication (NFC)*, pp. 1–6. IEEE, 2013.
- [122] M. I. Posner. Orienting of Attention. *Quarterly Journal of Experimental Psychology*, 32(1):3–25, 1980.
- [123] M. I. Posner, M. J. Nissen, and R. M. Klein. Visual Dominance: An Information-Processing Account of its Origins and Significance. *Psychological Review*, 83(2):157–171, 1976.
- [124] I. Rabbi and S. Ullah. A Survey on Augmented Reality Challenges and Tracking. *Acta graphica: znanstveni časopis za tiskarstvo i grafičke komunikacije*, 24(1-2):29–46, 2013.
- [125] R. A. Rensink. Change Detection. *Annual Review of Psychology*, 53(1), 2002.
- [126] R. A. Rensink, J. K. O’Regan, and J. J. Clark. To See or Not to See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science*, 8(5):368–373, 1997.
- [127] I. Rock and J. Victor. Vision and Touch: An Experimentally Created Conflict Between the Two Senses. *Science*, 143(3606):594–596, 1964.

- [128] M. Ruble. Investigation of Possible Size Differences Between Virtual Overlay and Physical Prop in Optical See-through Tangible Augmented Reality. Bachelor's thesis, Saarland University, Germany, 2020.
- [129] M. Ruble. Tangible Augmented Reality for Virtual Scene Authoring. Master's thesis, Saarland University, Germany, 2022.
- [130] N. Rutsch. Methoden zur Blickrichtungssteuerung am projizierten Einkaufsregal. Bachelor's thesis, Saarland University, Germany, 2015.
- [131] V. Schwind, P. Knierim, N. Haas, and N. Henze. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.
- [132] A. Seeliger, G. Merz, C. Holz, and S. Feuerriegel. Exploring the Effect of Visual Cues on Eye Gaze During AR-Guided Picking and Assembly Tasks. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 159–164. IEEE, 2021.
- [133] A. L. Simeone, E. Velloso, and H. Gellersen. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3307–3316, 2015.
- [134] M. Slater and S. Wilbur. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997.
- [135] M. Smith and A. McNamara. Gaze Direction in a Virtual Environment via a Dynamic Full-image Color Effect. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–2. IEEE, 2018.
- [136] Z. Soleimanitaleb, M. A. Keyvanrad, and A. Jafari. Object Tracking Methods: A Review. In *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 282–288. IEEE, 2019.
- [137] S. S. Sørensen. The Development of Augmented Reality as a Tool in Architectural and Urban Design. *Nordic Journal of Architectural Research, Volume 19, No 4*, 19(4):25–32, 2006.
- [138] S. Sridharan, R. Bailey, A. McNamara, and C. Grimm. Subtle Gaze Manipulation for Improved Mammography Training. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 75–82, 2012.

- [139] X. Sun and A. Varshney. Investigating Perception Time in the Far Peripheral Vision for Virtual and Augmented Reality. In *Proceedings of the 15th ACM Symposium on Applied Perception*, pp. 1–8, 2018.
- [140] I. E. Sutherland et al. The Ultimate Display. In *Proceedings of the IFIP Congress*, vol. 2, pp. 506–508. New York, 1965.
- [141] M. Szemenyei and F. Vajda. Learning 3D Object Recognition Using Graphs Based on Primitive Shapes. In *Workshop on the Advances of Information Technology, Budapest, Hungary*, pp. 187–195, 2015.
- [142] M. Szemenyei and F. Vajda. 3D Object Detection and Scene Optimization for Tangible Augmented Reality. *Periodica Polytechnica Electrical Engineering and Computer Science*, 62(2):25–37, 2018.
- [143] A. M. Treisman and G. Gelade. A Feature-integration Theory of Attention. *Cognitive Psychology*, 12(1):97–136, 1980.
- [144] P. Tuddenham, D. Kirk, and S. Izadi. Graspables Revisited: Multi-touch vs. Tangible Input for Tabletop Displays in Acquisition and Manipulation Tasks. In *CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2223–2232. ACM, 2010.
- [145] H. Uchiyama and E. Marchand. Object Detection and Pose Tracking for Augmented Reality: Recent Approaches. In *18th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, 2012. Available at: <https://inria.hal.science/hal-00751704/document> (last accessed: 2023-08-15).
- [146] C. Ullrich, M. Aust, R. Blach, M. Dietrich, C. Igel, N. Kreggenfeld, D. Kahl, C. Prinz, and S. Schwantzer. Assistenz- und Wissensdienste für den Shopfloor. In *Proceedings of DeLFI Workshops*, vol. 13, 2015. Available at: <https://ceur-ws.org/Vol-1443/paper15.pdf> (last accessed: 2023-08-15).
- [147] C. Ullrich, M. Aust, N. Kreggenfeld, D. Kahl, C. Prinz, and S. Schwantzer. Assistance- and Knowledge-services for Smart Production. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, pp. 1–4, 2015.
- [148] M. Usoh, E. Catena, S. Arman, and M. Slater. Using Presence Questionnaires in Reality. *Presence*, 9(5):497–503, 2000.

- [149] D. W. F. van Krevelen and R. Poelman. A Survey of Augmented Reality Technologies, Applications and Limitations. *International Journal of Virtual Reality*, 9(2):1–20, 2010.
- [150] P. Vávra, J. Roman, P. Zonča, P. Ihnát, M. Němec, J. Kumar, N. Habib, and A. El-Gendi. Recent Development of Augmented Reality in Surgery: A Review. *Journal of Healthcare Engineering*, 2017:1–9, 2017.
- [151] E. E. Veas, E. Mendez, S. K. Feiner, and D. Schmalstieg. Directing Attention and Influencing Memory with Visual Saliency Modulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1471–1480, 2011.
- [152] F. Vega-Bermudez and K. O. Johnson. Fingertip Skin Conformance Accounts, in Part, for Differences in Tactile Spatial Acuity in Young Subjects, but Not for the Decline in Spatial Acuity with Aging. *Perception & Psychophysics*, 66(1):60–67, 2004.
- [153] R. von Wartburg, P. Wurtz, T. Pflugshaupt, T. Nyffeler, M. Lüthi, and R. M. Müri. Size Matters: Saccades During Scene Perception. *Perception*, 36(3):355–365, 2007.
- [154] A. Vovk, F. Wild, W. Guest, and T. Kuula. Simulator Sickness in Augmented Reality Training Using the Microsoft HoloLens. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–9, 2018.
- [155] N. Waldin, M. Waldner, and I. Viola. Flicker Observer Effect: Guiding Attention through High Frequency Flicker in Images. In *Computer Graphics Forum*, vol. 36, pp. 467–476. Wiley Online Library, 2017.
- [156] B. Wang, Y. Han, D. Tian, and T. Guan. Sensor-based Environmental Perception Technology for Intelligent Vehicles. *Journal of Sensors*, 2021:1–14, 2021.
- [157] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2019.
- [158] M. Weiser. The Computer for the 21 st Century. *Scientific American*, 265(3):94–105, 1991.
- [159] T. A. Whitaker, C. Simões-Franklin, and F. N. Newell. Vision and Touch: Independent or Integrated Systems for the Perception of Texture? *Brain Research*, 1242:59–72, 2008.

- [160] C. D. Wickens and J. S. McCarley. *Applied Attention Theory*. CRC Press, 2019.
- [161] E. Williams. Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Australian Journal of Chemistry*, 2(2):149–168, 1949.
- [162] C. Wisultschew, G. Mujica, J. M. Lanza-Gutierrez, and J. Portilla. 3D-LIDAR Based Object Detection and Tracking on the Edge of IoT for Railway Level Crossing. *IEEE Access*, 9:35718–35729, 2021.
- [163] L.-K. Wong and K.-L. Low. Saliency Retargeting: An Approach to Enhance Image Aesthetics. In *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 73–80. IEEE, 2011.
- [164] M. Yokomi, N. Isoyama, N. Sakata, and K. Kiyokawa. Subtle Gaze Guidance for 360° Content by Gradual Brightness Modulation and Termination of Modulation by Gaze Approaching. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 520–521. IEEE, 2021.
- [165] R. A. Young, R. M. Lesperance, and W. W. Meyer. The Gaussian Derivative Model for Spatial-temporal Vision: I. Cortical Model. *Spatial Vision*, 14(3):261–320, 2001.