ENHANCING INTERACTION IN MIXED REALITY

The Impact of Modalities and Interaction Techniques on the User Experience in Augmented and Virtual Reality

DISSERTATION

an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

vorgelegt von

PASCAL KNIERIM

München, den 20. Januar 2020

Dissertation

Enhancing Interaction in Mixed Reality

The Impact of Modalities and Interaction Techniques on the User Experience in Augmented and Virtual Reality

Pascal Knierim
Human-Centered Ubiquitous Media Group
Institut für Informatik
Ludwig-Maxmilians-Universität München
pascal.knierim@ifi.lmu.de

2012 ACM Computing Classification
• Human-centered computing ~Human computer interaction (HCI) ~Interaction paradigms ~Mixed / augmented reality

ISBN-13: 979-8594425187

Publication date
14. December 2020

Erstgutachter: Prof. Dr. Albrecht Schmidt

Zweitgutachter: Prof. Dr. Morten Fjeld

Drittgutachter: Prof. Dr. Kai Kunze

Tag der Disputation: 13. Oktober 2020

ABSTRACT

With continuous technological innovation, we observe mixed reality emerging from research labs into the mainstream. The arrival of capable mixed reality devices transforms how we are entertained, consume information, and interact with computing systems, with the most recent being able to present synthesized stimuli to any of the human senses and substantially blur the boundaries between the real and virtual worlds. In order to build expressive and practical mixed reality experiences, designers, developers, and stakeholders need to understand and meet its upcoming challenges. This research contributes a novel taxonomy for categorizing mixed reality experiences and guidelines for designing mixed reality experiences.

We present the results of seven studies examining the challenges and opportunities of mixed reality experiences, the impact of modalities and interaction techniques on the user experience, and how to enhance the experiences. We begin with a study determining user attitudes towards mixed reality in domestic and educational environments, followed by six research probes that each investigate an aspect of reality or virtuality. In the first, a levitating steerable projector enables us to investigate how the real world can be enhanced without instrumenting the user. We show that the presentation of in-situ instructions for navigational tasks leads to a significantly higher ability to observe and recall real-world landmarks. With the second probe, we enhance the perception of reality by superimposing information usually not visible to the human eye. In amplifying the human vision, we enable users to perceive thermal radiation visually. Further, we examine the effect of substituting physical components with non-functional tangible proxies or entirely virtual representations. With the third research probe, we explore how to enhance virtuality to enable a user to input text on a physical keyboard while being immersed in the virtual world. Our prototype tracked the user's hands and keyboard to enable generic text input. Our analysis of text entry performance showed the importance and effect of different hand representations. We then investigate how to touch virtuality by simulating generic haptic feedback for virtual reality and show how tactile feedback through quadcopters can significantly increase the sense of presence. Our final research probe investigates the usability and input space of smartphones within mixed reality environments, pairing the user's smartphone as an input device with a secondary physical screen.

Based on our learnings from these individual research probes, we developed a novel taxonomy for categorizing mixed reality experiences and guidelines for designing mixed reality experiences. The taxonomy is based on the human sensory system and human capabilities of articulation. We showcased its versatility and set our research probes into perspective by organizing them inside the taxonomic space. The design guidelines are divided into user-centered and technology-centered. It is our hope that these will contribute to the bright future of mixed reality systems while emphasizing the new underlining interaction paradigm.

ZUSAMMENFASSUNG

Mixed Reality (vermischte Realitäten) gehen aufgrund kontinuierlicher technologischer Innovationen von reinen Forschungsarbeiten langsam in den Massenmarkt über. Mit der Einführung von leistungsfähigen Mixed-Reality-Geräten verändert sich die Art und Weise, wie wir Unterhaltungsmedien und Informationen konsumieren und wie wir mit Computersystemen interagieren. Verschiedene existierende Geräte sind in der Lage, jeden der menschlichen Sinne mit synthetischen Reizen zu stimulieren. Hierdurch verschwimmt zunehmend die Grenze zwischen der realen und der virtuellen Welt. Um eindrucksstarke und praktische Mixed-Reality-Erfahrungen zu kreieren, müssen Designer und Entwicklerinnen die künftigen Herausforderungen und neuen Möglichkeiten verstehen. In dieser Dissertation präsentieren wir eine neue Taxonomie zur Kategorisierung von Mixed-Reality-Erfahrungen sowie Richtlinien für die Gestaltung von solchen.

Wir stellen die Ergebnisse von sieben Studien vor, in denen die Herausforderungen und Chancen von Mixed-Reality-Erfahrungen, die Auswirkungen von Modalitäten und Interaktionstechniken auf die Benutzererfahrung und die Möglichkeiten zur Verbesserung dieser Erfahrungen untersucht werden. Wir beginnen mit einer Studie, in der die Haltung der nutzenden Person gegenüber Mixed Reality in häuslichen und Bildungsumgebungen analysiert wird. In sechs weiteren Fallstudien wird jeweils ein Aspekt der Realität oder Virtualität untersucht. In der ersten Fallstudie wird mithilfe eines schwebenden und steuerbaren Projektors untersucht, wie die Wahrnehmung der realen Welt erweitert werden kann, ohne dabei die Person mit Technologie auszustatten. Wir zeigen, dass die Darstellung von in-situ-Anweisungen für Navigationsaufgaben zu einer deutlich höheren Fähigkeit führt, Sehenswürdigkeiten der realen Welt zu beobachten und wiederzufinden. In der zweiten Fallstudie erweitern wir die Wahrnehmung der Realität durch Überlagerung von Echtzeitinformationen, die für das menschliche Auge normalerweise unsichtbar sind. Durch die Erweiterung des menschlichen Sehvermögens ermöglichen wir den Anwender:innen, Wärmestrahlung visuell wahrzunehmen. Darüber hinaus untersuchen wir, wie sich das Ersetzen von physischen Komponenten durch nicht funktionale, aber greifbare Replikate oder durch die vollständig virtuelle Darstellung auswirkt. In der dritten Fallstudie untersuchen wir, wie virtuelle Realitäten verbessert werden können, damit eine Person, die in der virtuellen Welt verweilt, Text auf einer physischen Tastatur eingeben kann. Unser Versuchsdemonstrator detektiert die Hände und die Tastatur, zeigt diese in der vermischen Realität an und ermöglicht somit die verbesserte Texteingaben. Unsere Analyse der Texteingabequalität zeigte die Wichtigkeit und Wirkung verschiedener Handdarstellungen. Anschließend untersuchen wir, wie

man Virtualität berühren kann, indem wir generisches haptisches Feedback für virtuelle Realitäten simulieren. Wir zeigen, wie Quadrokopter taktiles Feedback ermöglichen und dadurch das Präsenzgefühl deutlich steigern können. Unsere letzte Fallstudie untersucht die Benutzerfreundlichkeit und den Eingaberaum von Smartphones in Mixed-Reality-Umgebungen. Hierbei wird das Smartphone der Person als Eingabegerät mit einem sekundären physischen Bildschirm verbunden, um die Ein- und Ausgabemodalitäten zu erweitern.

Basierend auf unseren Erkenntnissen aus den einzelnen Fallstudien haben wir eine neuartige Taxonomie zur Kategorisierung von Mixed-Reality-Erfahrungen sowie Richtlinien für die Gestaltung von solchen entwickelt. Die Taxonomie basiert auf dem menschlichen Sinnessystem und den Artikulationsfähigkeiten. Wir stellen die vielseitige Verwendbarkeit vor und setzen unsere Fallstudien in Kontext, indem wir sie innerhalb des taxonomischen Raums einordnen. Die Gestaltungsrichtlinien sind in nutzerzentrierte und technologiezentrierte Richtlinien unterteilt. Es ist unser Anliegen, dass diese Gestaltungsrichtlinien zu einer erfolgreichen Zukunft von Mixed-Reality-Systemen beitragen und gleichzeitig die neuen Interaktionsparadigmen hervorheben.

PREFACE

This thesis is the result of the research I carried out at the University of Stuttgart and the LMU Munich. Numerous discussions, exchange with researchers and practitioners at conferences, workshops, and lab visits inspired and shaped the decisions presented in this work. Since some of the research required varying expertise, this thesis has been done in close collaboration with partners from the LMU Munich, University of Stuttgart, and partners within the *FeuerWeRR* and *Be-Greifen* projects delivering expert knowledge from their respective fields. Further, I supervised undergraduate student projects, Bachelor and Master theses that help to realize my ideas, prototypes, and evaluations. Many of these collaborations and theses resulted in publications that are a core part of this work. To emphasize these collaborations, I chose to write this thesis using the scientific plural ("we"). When applicable, all contributing authors of the resulting publication are cited at the beginning of each chapter, including the publication's reference.

ACKNOWLEDGMENTS

Working on this thesis became one of the most challenging but also exciting journeys. In the past years, I had the privilege of meeting and working with several outstanding people that influenced my research and shaped me as an individual. Their professional input and even more importantly, their personal support made this time so unique and valuable. With these acknowledgments, I want to point out all those beautiful people, and I sincerely apologize to everybody I missed to mention.

First and foremost, I want to thank my supervisor **Albrecht Schmidt** for infinite patience, constant support, and useful advice when needed. His passion for research and unbroken flow of ideas build the foundation of this thesis. I am grateful for continually challenging me, not only professionally but also on ski slopes and beach running tracks. I owe gratitude to my two external reviewers **Morten Fjeld** and **Kai Kunze**, for fruitful discussions and their valuable time.

I want to thank **Enrico Rukzio** and **Florian Schaub** for introducing me to the field of human-computer interaction and promoting me towards this fantastic journey in the first place. Many thanks go to **Tilman Dinger** for explaining academic paper writing to me, **Stefan Schneegaß**, for teaching me how to claim office space, numerous BBQ sessions, and countless discussions. **Markus Funk**, the most passionate drone pilot for his forward leading AR projects. Further thanks go to **Paweł Woźniak** for his motivation in dark times, perspectives, and most valuable feedback when needed. **Yomna Abdelrahman**, for the delicious food, being a fantastic project partner and most important discussions about *rampen*. **Lars Lischke**, for his cafe support and being a great office mate. **Sven Mayer**, for infinite memorable experiences at conferences. **Francisco Kiss**, as a valued project partner and electrical engineer. **Niels Henze** and **Valentin Schwind**, for the countless projects and Christmas cards we did together.

I want to thank the entire hciLab at the University of Stuttgart for making the time there unforgettable: **Dominik Weber**, **Huy Viet Le**, **Alexandra Voit**, **Rufat Rzayev**, **Patrick Bader**, **Céline Coutrix**, **Passant El.Agroudy**, **Miriam Greis**, **Bastian Pfleging**, **Romina Poguntke**, **Mauro Avila**, and **Katrin Wolf**.

During my Ph.D. I frequently traveled, which resulted in meeting extraordinary people at conferences or my internship at Microsoft Research. I want to thank **Steve Hodges**, **Nicholas Chen**, and **James Scott**, for showcasing how to combine research and industry fantastically. Further, I thank **Gavin Wood** for countless discussions at Salisbury Arms.

I want to thank **Martin Strzys**, **Sebastian Knapp**, **Katja Hintz**, and all external collaborators from the project *Be-Greifen*.

At the LMU Munich, I got the chance to meet and work with many talented old and new faces. I want to thank **Thomas Kosch**. A fearless runner, drone, and nerf gun operator. An extraordinary researcher and fantastic van driver, I spent hours and hours on german highways always freezing but never bored due to in-depth discussions and great music. I thank **Matthias Hoppe** and **Jakob Karolus**, for all the good (AR) arguments during coffee breaks, and **Tobias Benz** for relativizing a skewed work-life balance, his optimism, and *gin*spiration.

Extended thanks to the ever-growing group of motivated HCUM people that positively influenced my time at the LMU: **Fiona Draxler**, **Sebastian Feger**, **Jesse Grootjen**, **Tonja Machulla**, **Lewis Chuang**, **Florian Lang**, **Ville Mäkelä**, **Luke Haliburton**, **Francesco Chiossi**, **Lauren Thevin**, and **Steeven Villa**. Besides, I thank **Florian Bemmann**, **Malin Eiband**, **Ceenu George**, **Christina Schneegaß** and **Pino**, **Nađa Terzimehić**, **Sarah Völkel**, **Andreas Butz**, **Linda Hirsch**, **Kai Holländer**, **Carl Oechsner**, **Beat Rossmy**, **Christian Mai**, and the entire *MIMUC* troop for making the office that bright.

Special thanks go to **Anja Mebus** for keeping us all connected to reality and protecting us from outdated inventory lists. Further, I thank **Murielle Naud-Barthelmeß** for keeping track of all the messed up timesheets, and **Christa Feulner**, **Franziska Schwamb**, and **Rainer Fink** for their administrative work.

I was lucky to meet and work with countless talented students: notably, the works of **Gabrielle LaBorwit**, **Steffen Maurer**, **Maximilian Rauh**, **Matthias Hoppe**, and **Dimiti Hein**, whose output also made their way into this thesis.

Lastly, I want to thank **Jens Nickels** and **Thomas Pfeiffer** for their encouragement and regularly breaking my scientific filter bubble.

To finally conclude this list of remarkable people, I want to thank and honor my parents **Christiane** and **Ralf Knierim** for their limitless support, patience, and love. **Chantal Knierim** for being always there for me.

I thank **Mechthild** and **Volker Röhlig** for their trust in my work. I thank **Amrei Knierim** for her unconditional love, her selflessness that enables me to travel and work during unconventional times, and for sharing ideas and dreams. Without your support, this work would not have been possible. I thank **Jonna Knierim** for showing me that the world is way more than research. I thank my whole family and friends for their prolonged encouragement and kindness when I was not there.

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

List of Acronyms

ANOVA	analysis of variance
API	Application Programming Interface
AR	Augmented Reality
BMBF	Federal Ministry of Education and Research (German: Bundesministerium für Bildung und Forschung)
CAVE	cave automatic virtual environment
CV 1	Consumer Version 1
DFKI	German Research Center for Artificial Intelligence
DK	Developer Kit
DOF	Degrees of Freedom
FOV	Field of View
FPS	Frames per Second
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GUI	Graphical User-interface
HCI	Human-Computer Interaction
HMD	Head-Mounted Display
HTCE	Heat and Temperature Conceptual Evaluation
IMU	Inertial mMeasurement Unit
IR	Infrared Radiation
LCD	Liquid Crystal Display
MR	Mixed Reality
MRE	Mixed Reality Experience
POI	Point of Interest
PQ	Presence Questionnaire
RTLX	Raw NASA Taskload Index
RV	Reality-Virtuality
SDK	Software Developer Kit
STEM	Science, Technology, Engineering and Mathematics

SUS	System Usability Scale
TCT	Task Completion Time
TLX	Task Load Index
TUK	Technical University of Kaiserslautern
UI	User Interface
VR	Virtual Reality
WPM	Words per Minute

Chapter 1

Introduction

An early appearance of what we now see in virtual and augmented reality came in the science fiction literature of the early nineteenth century.. Novelists imagined people connected with machines that overlaid all human senses and immersed them in virtual worlds [165], or wore devices such as electronic glasses [254] that mapped data onto the real world. The concept of Augmented Reality (AR) and Virtual Reality (VR) has stayed alive and gradually turned science-fiction into reality aided by increasingly sophisticated technology: Early science fiction used painted scenery and written tales to convey ideas of imagined worlds. Later, photography and film helped creators to share their visions, ideas, and creativity. AR and VR have proliferated over the last years, and today's interactive and powerful three-dimensional simulations allow us to immerse in virtual worlds and experience environments limited only by the imagination of the creator.

The advances in technology enable a paradigm shift in how we consume, interact with, and explore information. Technology fosters highly personalized virtual realities through windows, films, or canvases. Artists, developers, and designers can manipulate any aspect of the environment in real-time and dynamically present digital stimuli. AR and VR technology have become a medium that presents information and immerses users like never before, so that today, users can go beyond reading and watching to interactively explore the information. This shift fundamentally changes digital entertainment [201] and how we work and learn with digital media. Simultaneously, new challenges regarding these interactive systems arise. The considerable extension of input and output modalities substantially increases the potential of these new devices. Information presentation is no longer limited to visual and auditory impulses; existing devices can display synthesized information to any of the human senses. Likewise, new sensors allow for manifold user input beyond explicit device input. In synergy, highly integrated demonstration and interaction with digital information will increasingly enable the technology to blend unobtrusively into the background while providing great utility to the users.

This thesis systematically investigates how to enhance interaction and perception of reality in mixed reality environments. We demonstrate how to integrate novel input and output concepts into enriched experiences. We approach the challenge of blending realities from both ends of the spectrum, reality, and virtuality. We proclaim the potential of extended realities through several research probes and potential application scenarios and build a better understanding of how to enhance Mixed Reality (MR) experiences. We propose best practices on the integration of input and output modalities for compelling experiences based on the developed probes.

1.1 Vision

In this section, we describe our vision of how VR and AR experiences will influence and shape our future in the professional and private domain. Innovative technology always captivates and attracts audiences. If novel technology gains acceptance, it can fundamentally change the behavior of a society. For instance, the invention of the automobile completely changed the paradigm of mobility. Along with enabling people to travel, it drastically changed how business operates and influenced the economy.

Due to the rapid development in science and technology, similar periods of change can be witnessed in the domain of connected mobile devices. With the presentation of the first iPhone back in 2007 and its constant development since, many paradigms changed. Before the smartphone era, we used physical maps, dedicated devices to listen to music, and stationary computers to browse the internet. Today, smartphones are used everyday and on-the-go for texting, navigation, social media, calling, browsing the internet, and taking pictures. Customized apps can significantly extend the functionally of these devices as more and more advanced sensors are integrated.

Peak smartphone use has not been reached yet [189]; however, the next generation of mobile devices that will radically change how people carry out daily activities, learn, and interact with computing systems is approaching. In the near future, VR and AR systems will grow more similar to one another, consequently becoming known as MR systems. We expect that wearable MR systems will gradually substitute today's smartphones. As the next step beyond Weiser's vision [255], the technology can be woven into our clothes and accessories and will replace the functionalities of our smartphones.

In the professional domain, MR systems can empower employees to work together even from ad-hoc or distant locations [91], enhance learning with step-by-step instructions to help employees to acquire new skills, improve efficiency and reduce errors [71]. For creative tasks, immersive visualizations can foster discussions and simplify the exploration of new layouts and products [195] before they are built as a physical model.

Also, in our daily lives, MR systems will map an analysis of the physical world around us to present useful services. Information will be available everywhere and blend indistinguishably into the real world. This includes information according to the location and context, as well as very personal information. As a chaperone, MR experiences will help us to navigate the environment, support decision making, and interact with the environment. We also expect a substantial change in entertainment and education. MR is the first technology that can displace people into the experience of others, creating new means to communicate, learn, and explore [117].

We assume that future MR systems are going to amplify all of the human senses and not be limited to visual overlays. Although many technological advances still need to be set in place before this vision can turn into reality, including hardware, software, and social aspects, and potential concerns as privacy, information overload, and immaturity. In consequence, MR technology is very likely to be strongly interwoven with people and become an extension of themselves.

When we allow for intuitive interaction and convincing presentation, the future for the advent of MR is bright. We believe the presented research in this thesis and future research will shape the technology towards an appealing and useful component of our everyday life.

1.2 Research Questions

We expect that AR and VR have the potential to become ubiquitously available. Novel devices will progressively penetrate the technology markets and potentially

RESEARCH QUESTIONS	
	USER-CENTERED
RQ1	What are the users' attitudes towards the introduction of AR at home?
RQ2	What are the scenarios that are most promising for domestic AR applications?
RQ3	What are the constraints and opportunities for future systems?
	MIXED REALITY EXPERIENCE-CENTERED
RQ4	How to enhance reality by blending information in the real world?
RQ5	How to enhance virtual reality by adding tangibility?
RQ6	How to provide flexible haptic input and output in mixed reality?
	INTERACTION-CENTERED
RQ7	How to interact with digital artifacts in mobile mixed reality?
RQ8	How to design a taxonomy for mixed reality experiences?

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

replace smartphones due to their enhanced usability and versatility. To fully understand the potential and identify the limitations of this emerging technology, research is required, and the arising fundamental challenges have to be addressed. Summarizing, our research approach is based on three major research topics investigating the essential challenges from a *user-centered* and a *interaction-centered* perspective. In between resides the investigation of MR systems organized as an *experience-centered* extension. An overview of the three themes and the specific research questions addressed in this thesis is presented in Table 1.1.

Upcoming MR experiences can be used in a wide variety of different locations and situations. Understanding the individual challenges and requirements helps to design and develop fascinating and meaningful experiences for the user. The first group of user-centered research questions focuses on the potential of MR in domestic environments **(RQ1)**. Further, we research potential constraints and opportunities of future MR systems **(RQ2, RQ3)**.

The second group of research questions focuses on the unique challenges and opportunities of MR systems and how these experiences can be extended. Therefore, we approach MR experiences from the real and virtual extrema of the reality-virtuality continuum [175]. We investigate the advantages and limitations of the facilitated technology on both ends of the spectrum. We developed research probes that blend and substitute real and virtual artifacts and introduce new interaction modalities. For a precise investigation of the effects, we developed research probes extending reality **(RQ4)**, and virtuality **(RQ5)**, as well as blending experiences **(RQ6)**. We investigate the performance and effects of the particular extension in a delimited application scenario with each probe.

After understanding the different aspects of dynamically blending and extending realities, we address *interaction-centered* challenges. There is considerable modality support for interaction in MR, but technological and environmental factors influence the suitability of input modalities. Thus we explore how to interact with these digital artifacts in mobile MR environments **(RQ7)**. In a final step, we explore how to design a novel taxonomy for mixed reality experiences **(RQ8)** based on the human abilities.

The scope of the thesis embraces the exploration of the challenges and opportunities of MR applications. The main focus lies in overcoming the limitations and complexity of MR systems and applications and contributing to the advancement of input and output modalities. The overarching theme is assessing the impact of the superimposed virtual realm and the perceived reality on the user experience to guide the ideation, development, and evaluation of MR applications.

1.3 Methodology

The field of MR is not only driven by constant research, but also by advances in technology. Virtual reality experiences [99] and Head-Mounted Displays (HMDs) [241] were already suggested in the early 1960s, and since then, devices have become more reliable, lighter, and more affordable, allowing for more extensive studies, sometimes even outside of labs. However, there is still no agreed understanding of how MR experiences can affect our lives positively. To develop meaningful MR applications we followed a user-centered design approach [4] as one of our principal methodologies. User-centered design is an iterative design process in which developers start with a general approach and focus on users' feedback and needs in each iteration of the continuous design process. We further apply design thinking methods, including paper prototypes, wizard-of-oz prototypes, and digital mock-ups, to simulate future systems and ideas. During the development process of our research probes, we evaluated these in formative user studies. With our empirical research practices, we foster a better understanding of how humans interact in MR. Consequently, our findings provide knowledge that enables us to create improved input and output modalities for future systems.

1.3.1 Research Probes

Throughout the thesis, we build and evaluate several prototypes to understand the opportunities and challenges of MR experiences development and to what extent these experiences can be enhanced by blending with digital or physical artifacts. Each of the research probes was developed following an iterative user-centered design process. For the design of novel concepts of input and output for mixed reality systems, we also applied participatory design methods and focus groups. An overview of the built research probes is presented in Table 1.2.

1.3.2 Evaluation

With the development of the presented research probes, we pursue one overarching objective: building a better understanding of how to enhance MR experiences. To that extent, the novel user interfaces and MR experiences are evaluated in controlled lab experiments and field studies.

In each study, we use a variety of qualitative and quantitative measures to evaluate the effects of the research probe. Quantitative measures include but are not limited to performance measures, e.g., task complexation time, error rate, memorability, or perceived workload. Further, we collect more specific measures like presence in VR experiences. All qualitative data were analyzed using conventional statistical methods and tools [63].

We gathered qualitative data through semi-structured interviews, focus groups, or online questionnaires. For qualitative data analysis, we used grounded theory in the course of obtaining common themes. More specifically, we utilized an open coding approach.

1.3.3 Ethics

The research and user studies presented in this thesis were conducted within the Federal Ministry of Education and Research (German: Bundesministerium für Bildung und Forschung) (BMBF) funded projects Be-Greifen and FeuerWeRR. Within the Be-Greifen project, ethical, legal, and social implications were addressed in active collaboration with Wulf Loh and Tobias Stoerzinger from the Institute for Philosophy at the University of Stuttgart. Every development and conducted study was intended to make the resulting technological solutions valuable for everyone.

Each of the studies was conducted in line with the declaration of Helsinki. Starting in May 2019, we further followed the General Data Protection Regulation (GDPR) 2016/679 that regulates data protection and privacy.

1.4 Research Context

The research leading to this thesis and beyond was carried out between autumn 2014 and summer 2019 at the University of Stuttgart in the Human-Computer Interaction group and the LMU Munich in the Human-Centered Ubiquitous Media group. Many of the projects were done in collaboration with project partners that influenced the work presented in this thesis.

1.4.1 Be-Greifen

Most of the research presented in this thesis was conducted as part of the BMBF-funded the project Be-Greifen[1]. The project started in July 2016 with a term of three years. The project aims to investigate how to support the fusion of the real and digital world to create a strong connection between lab experiments and theory to support learners in the Science, Technology, Engineering and Mathematics (STEM) subjects. Together with Martin Strzys, Sebastian Kapp from the Technical University of Kaiserslautern (TUK) and the German Research Center for Artificial Intelligence (DFKI) we explored and evaluated how to plan, design, and implement intuitive user interfaces to interact with scientific experiments through a combination of tangible user interfaces and augmented reality [135, 237, 238]. Part of this research also included the dynamic adaption of the visualization and interaction to the level of knowledge and learning requirements of the student [138].

[1] http://begreifen.dfki.de/

Within the same BMBF call, which focuses on assistive systems for learning, we collaborated with partners from the projects KoBeLU[2] and MAL[3]. Together with Thomas Kosch (KoBeLU) we examined how assistive systems and mixed realities can support knowledge transfer in education and which challenges need to be tackled. In cooperation with Tanja Döring (MAL) and others, we organized a workshop [52] on tangible interaction and brought together experts in this domain to discuss how tangible interaction and interactive systems can support learning [147].

1.4.2 FeuerWeRR

Further research presented in this thesis was also conducted as part of the BMBF-funded project FeuerWeRR. The research was conducted between March 2015 and February 2018. The main objective of the research consortium was to develop and evaluate a new mobile thermal imaging camera that links depth information of the environment with the thermal data. This real-time fusion of data would increase firefighters' safety and allow citizens at risk of fire to be rescued more quickly.

The central research goal was to develop a radar sensor and the integration into a hand-held camera that combines these two data sources. We focused in particular on different visualizations and interaction concepts [6, 7, 8] and the fusion of temperature data with the acquired depth information to enable firefighters effortless process all data while keeping the overall cognitive load on a minimum.

1.4.3 Selected Collaboration

Alongside the defined projects Be-Greifen and FeuerWeRR, we investigated in additional collaborations the potential of Mixed Realities beyond the scope of this thesis.

---

[2] http://www.kobelu.de/

[3] http://mal-projekt.de/

Collaboration with University of Copenhagen

Together with Sebastian Boring and Markus Löchtefeld (Aalborg university), we organized a workshop on interactive displays through mobile projection [263] at the University of Copenhagen. We discussed with experts, to what extent mobile projections would change future mobile interactions and information visualization [262]. Furthermore, the first discussions inspired us to develop and study the capabilities of drone-carried projectors [141].

Collaboration with Microsoft Research Cambridge

In close cooperation with Steve Hodges and the Sensors and Devices Group at Microsoft Research, we shaped an understanding of how users can improve productivity and convenience with a secondary screen integrated into a smartphone display cover [109]. During the development, several limitations and challenges were tackled to allow users to work with a smartphone more productively. These results were filed in various patents stating concepts for data transfer, power harvesting in a mobile context, and attention-based interaction [83, 84, 101, 102, 243]

University of Stuttgart

Alongside the research presented in this thesis, we realized various projects at the University of Stuttgart. We developed several virtual environments to investigate the effects of avatar representation on presence [222, 220, 221]. We carried out this work mainly with Valentin Schwind and Niels Henze.

Other research was conducted in collaboration with Markus Funk, Thomas Kosch, Lars Lischke, Paweł W. Woźniak, and Katrin Wolf. Some of the work presented here is further based on collaborations with undergraduate students whom we supervised during the time at the University of Stuttgart in their seminars, Bachelor and Master theses [11, 75, 93, 113, 122, 131, 168, 190, 256].

LMU Munich

From October 2017, we continued our research from the Human-Centered Ubiquitous Media group, which is part of the institute for informatics at the Ludwig-Maximilians-Universität München. In various projects, we studied together with Matthias Hoppe and Thomas Kosch the effects of interactive tangible components in MR environments [137, 142, 192], how to amplify our vision by altering time perception [129, 139] and the integration of MR systems into sport [264]. Additional projects and works that are shown here are based on seminars, Bachelor, and Master thesis we supervised at the LMU Munich [5, 67, 82, 100, 149, 205].

RESEARCH PROBES

Prototype	Description	Chapter
	**Photo Elicitation App** To understand how people imagine AR in their daily lives, we developed an Android application to support photo-elicitation interviews. The application allows users to annotate photographs taken with the smartphone or tablet to envision AR use-cases. Through photo supported interviews, we obtained a better understanding of the potential, challenges, and opportunities AR in everyday life has to offer. [145]	Chap. 3
	**Quadcopter Mounted Projector** To enhance reality in a mobile outdoor scenario, we envisioned a projector drone, and we combined a remote-controlled quadcopter with a wireless programmable projector. Using this research probe, we investigated the effects of projected in-situ navigation cues compared to traditional cues presented on a smartphone. [141]	Chap. 4
	**Augmented Thermal Perception** Through a handheld or wearable AR display, we fused real-time sensor data with the real world to amplify the human senses and enhance the perception of reality. We combined AR capable smartphones, AR glasses with thermal imaginer to provide the wearer with in-situ thermal vision. We also investigated the effects of substituting the physical components of lab experiments with tangible replicas and virtual representations within this research probe. [135].	Chap. 5
	**Typing in Virtual Reality** To overcome the challenging task of text input while being immersed in virtual environments, we developed a prototype that tracks the user's hands and a physical keyboard and visualizes them in VR. The prototype comprises a consumer VR HMD and a precise low latency motion capturing system. Our application simulates a virtual working environment while haptic. In a study, we investigated the achievable text entry speed and the effect of hand representations and transparency on overall typing performance, workload, and presence. [144]	Chap. 6
	**Typing in Mobile Virtual Reality** Based on the experience gained from the development of the previous research probe, we developed a mobile prototype by replacing all specialized or stationary parts. The mobile research probe comprises only off-the-shelf hardware, including an AR capable smartphone, a VR viewer, and a physical keyboard. The setup is enabling text input in the mobile virtual reality context and allows widespread use beyond the lab. [137]	Chap. 6.7
	**Haptic Feedback in Virtual Reality** With *TactileDrone*, we envisioned and developed one of the first tactile feedback providing quadcopters for VRenvironments. While the user is visually and acoustically immersed in VR, small quadcopters simulate objects providing haptic feedback to the user. The mini-quadcopters are tracked by an optical marker tracking system and controlled wirelessly through our application. [140]	Chap. 7
	**Interaction in MR** With the*SmARtphone Controller* probe, we envision a ubiquitous and highly flexible input controller for MR environments. Repurposing a smartphone allows for multimodal input via touch, gestures, or movement of the device, while the display can act as a secondary high-resolution secondary screen. With this probe, we investigated the feasibility and usability of this in- and output modality. [133]	Chap. 8

Table 1.2: Research probes developed within the scope of this thesis. Each research prototype is described in detail in a dedicated chapter or sections within this thesis.

1.5 Summary of Research Contributions

This thesis's main contributions are to the field of input and output modalities for Mixed Reality Experiences (MREs). First, we identify challenges and opportunities coupled with potential use-cases for novel MR systems. Employing several research probes, we then explore the effects of different interaction and

Publication	Personal Contribution
Exploring the Potential of Augmented Reality in Domestic Environments [145]	I came up with the original research idea and led the study design. Further, I developed the research probe and conducted the study. I led an ample amount of the analyses of the interviews and the resulting publication.
Challenges and Opportunities of Mixed Reality Systems in Education [138]	I came up with the original research idea and was the leading author of the resulting publication.
Quadcopter-Projected In-Situ Navigation Cues for Improved Location Awareness [141]	I was involved in the primary research ideation process, supervised the research probe's development, and planned the user-study. I further made contributions to all sections of the final publication.
Tangibility is overrated: Comparing learning experiences of physical setups and their virtual equivalent in augmented reality [134]	The original idea for this research was deduced from the funding project proposal. I developed the first prototype and supervised the refinement. Further, I planned and analyzed the user study. I am the lead author of the resulting publication.
Look Inside: UnderstandingThermal Flux Through Augmented Reality[135]	I adapted the idea based on the project's proposal, created the research probe's architecture, and developed the first prototype, including visualization and interaction concept.
Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands [144]	I developed the original research idea and study design together with Dr. Valentin Schwind; I supervised the implementation of the research probe and the planning and conductance of the study; I analyzed the collected data and was the leading author of the publication.
Opportunities and Challenges of Text Input in Portable Virtual Reality [137]	Based on previous research [144], I initiated this research idea. I supervised and actively participated in the development and created the study design. I led the data analysis and was the lead author of the resulting publication.
Tactile Drones: Providing Immersive Tactile Feedback in Virtual Reality through Quadcopters [140]	I was involved in the ideation process of this research stream. I was the lead developer and author of this publication.
VRHapticDrones: Providing Haptics in Virtual Reality through Quadcopters [115]	Based on the previous research idea [140], I developed with Matthias Hoppe this research proposal. I supervised and participated in the development and study design process. I was actively co-authoring this submission.
The SmARtphone Controller: Leveraging Smartphones as Input and Output Modality for Improved Interaction within Mobile Augmented Reality Environments [133]	I came up with the original research idea. I planned the study together with Dr. Thomas Kosch. Further, I analyzed and interpreted the study data and took the lead during the publication's preparation.

Table 1.3: Personal contributions to the core publications of this thesis.

presentation techniques and contribute design recommendations for building more advantageous MR experiences. Lastly, we contribute a new taxonomy for classifying MR experiences based on the input and output modalities.

Research Prototypes

This thesis contributes to a diverse set of research probes. We envisioned, developed, and evaluated the prototypes to answer the research questions presented in Table 1.1. All prototypes were built using the latest off-the-shelf hardware and are described in detail to enable reproducibility. An overview of the prime research probes is presented in Table 1.2.

Contributing Publications

This thesis's core contributions are based on research published at international conferences with a competitive peer-reviewing process. All publications benefited from close collaboration, ideation, and discussions with the stated co-authors. However, for the core publications that lead to this thesis, I took the lead during the development, study design, analysis, and writing as clarified in Table 1.3. In each chapter or a specific section, the corresponding publication is stated at the very beginning.

1.6 Thesis Outline

This thesis contains eleven chapters. An overview of all chapters and how each chapter relates to the former is depicted in Figure 1.1. The first two chapters introduce the topic and provide an extensive description of virtual, augmented, and mixed realities. The third chapter covers the requirements from an end-user perspective and introduces mixed reality systems in domestic environments. The following five chapters (Chapters 4 - 8) are the central part of the thesis presenting the different research probes. The research probes cover the aspects of the extending reality (Chapter 4 and 5) and virtuality (Chapter 6 and 7). In Chapter 9 and 10, we present the overarching results of the thesis. First, we describe a new taxonomy for mixed reality experiences based on all the previously presented components and sets our research probes into perspective. Later, we illustrate the design recommendations we derived through all development conducted throughout this thesis. With the last chapter, we conclude the thesis and present the summary of research contributions as well as remaining future work.

Chapter 1 - Introduction. The first chapter motivates the topic of this thesis and unfolds our vision for mixed reality experiences. Further, the chapter contains the research context, research questions and used methodology followed by a summary of research contributions. The chapter closes with this brief outline.

Chapter 2 - Virtual, Augmented, and Mixed Realities In this chapter, we define the different realities and explain the different characteristics. Further we elaborate the concepts of presence and immersion. In the following we give a brief introduction to the history of mixed realities. Last, different input and output modalities and tools supporting the development of MR are discussed. The chapter closes with a discussion of current challenges faced by industry and research.

Chapter 3 - Understanding the Users In this chapter, we explore how MR could enhance everyday interaction from a user's perspective. We conducted an online survey and investigated the attitudes towards domestic AR. We further explored the opportunities for AR at home in a technology probe. We show that users are eager to benefit from on-demand information, assistance, enhanced sensory perception, and play offered by AR across many locations at home. Based on the results we present the derived functional and non-functional requirements for MR systems.

Chapter 4 - Enhancing Reality With this probe we explore how we can enhance reality through in-situ projection. As an application scenario, we evaluate visual navigation systems that usually constrain the users to shift their attention to the navigation system and then map the instructions to the real world. We suggest using in-situ navigation instructions that are presented directly by augmenting reality using a projector-quadcopter. In a user study we evaluate the concept of enhancing reality. We show that using in-situ instructions for navigation leads to a significantly higher ability to observe and recall real-world points of interest.

Chapter 5 - Perceiving Reality After focusing on enhancements of the virtual and the real world, we investigate in this chapter how to effectively blend both worlds and create functional and non-functional representations. We examine the effect of substituting the physical components of a lab experiment with tangible replicas and virtual representations. The results of the user study indicate that the substitution reduces the experiment setup duration without affecting knowledge transfer. We conclude with further insights for creating complex mixed reality learning environments.

Chapter 6 - Enhancing Virtuality Virtual Reality presents a challenge when entering text to computing systems. Since neither the user's hands nor the physical input devices are directly visible, conventional desktop peripherals are very slow, imprecise, and cumbersome. In this chapter we present an apparatus that enhances virtual reality environments. Our apparatus tracks the user's hands and a physical keyboard, and visualizes them in VR, t Thus, enabling users to interact comfortably with the computing system. Results of our text input study indicate that users almost reach outside-VR typing performance. We conclude that to enable high typing performance, the optimization of the visualization of hands in VR is important, especially for inexperienced typists.

Chapter 7 - Touching Virtuality In this chapter, we propose drones to provide tactile stimulation in VR, hence enhancing the haptic sensation. While the user is visually and acoustically immersed in VR, small drones simulate objects that provide active or passive tactile feedback to the user. In a user study, we demonstrate that haptic feedback provided by drones significantly increases users' sense of presence compared to vibrotactile controllers and interactions without additional haptic feedback. The chapter closes with a discussion of unobtrusive and flexible feedback provided by drones as well as insights for future VR systems enhanced with haptic feedback.

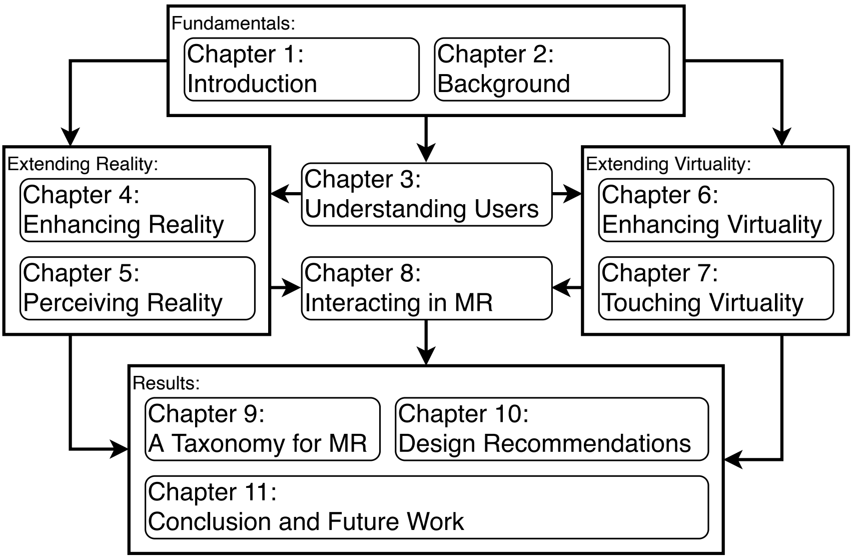


Figure 1.1: Outline and grouping of this thesis' chapters including the relation between the different chapters.

Chapter 8 - Interacting in Mixed Realities As a last research probe, we developed and evaluated a multimodal interaction concept by pairing a smartphone as an input controller with AR glasses. In a user study, we investigated the effects of the novel input controller on interaction speed, accuracy, and workload in different tasks. We can show that a smartphone-based controller results in significantly faster and more accurate interaction and reduced cognitive workload compared to mid-air gestures. Concluding the chapter, we discuss how future AR systems can benefit from touchscreens as an additional and complementary interaction modality.

Chapter 9 - A Taxonomy for Mixed Reality Experiences To put all research probes into perspective and enabling the classification and comparison of MR experiences, we propose a taxonomy that further helps gaining an overview of the build prototypes. In the taxonomy, the different MR experiences are grouped based on input and output dimensions and the degree of freedom or virtualization respectively.

Chapter 10 - Design Recommendations Through designing and developing mixed reality applications along with conducting numerous user studies we gained considerable experience. In this chapter, we provide design recommendations drawn from our experience for the design and implementation of mixed reality systems.

Chapter 11 - Conclusion and Future Work The last chapter summarizes our research contributions and reflects back on the research questions initially presented in this chapter. Finally, we outline remaining open questions and potential work for future research and development.

Chapter 2

Virtual, Augmented, and Mixed Realities

Augmented and virtual reality have immense potential to shape the way we interact with computing interfaces and perceive our environment. Currently, we consume information through dedicated, sometimes mobile devices that provide content on request, but Mixed Reality (MR) devices are becoming ever more ubiquitous and before long will provide us with information whenever needed. This trend opens a new set of possibilities. In the future, MR devices will extend and amplify our understanding of the environment we live in.

An essential aspect of MR experiences is to fuse digital information with the physical environment. The visualization of media, information, or interfaces is registered in our everyday environment and allows simple consumption. While the current interaction with interactive systems is explicit through dedicated devices such as smartphones, future devices like smart glasses will not be recognized as such and can provide a continuous personalized and private information stream to the user. Contextualized information is embedded in the physical world and will allow for new ways of interaction. Moreover, the next generation of devices will potentially augment additional human senses and allow for substantial changes in the perception of the environment.

In this chapter, we will introduce the most important definitions and terms in the field of augmented, virtual, and mixed realities. We provide a technology-driven and curated history of these simulated interactive experiences and elaborate on the rise and fall of this technology. Additionally, we will describe current developments and highlight selected prototypes and frameworks that support the steady stream of mixed reality experiences and devices becoming gradually mainstream.

2.1 Definitions of Realities

A *Virtual Reality (VR)* is an entirely computer-simulated environment that immerses the user and creates a feeling of *presence*. VR has the potential to lead someone to believe in being part of the simulated environment. Any of the human senses can be stimulated to convey the simulated environment. Today, there are varying definitions of VR. The term can be traced back to several sources in literature [165] and art [99]. Since technology is evolving rapidly, the definition of VR should ideally be detached from any technological advances and emphasis on the fundamental principles and characteristics of VR. Steuer defines VR based on the concept of telepresence that referes to any medium-induced subjective feeling of presence.

> A **Virtual Reality** is defined as a real or simulated environment in which a perceiver experiences telepresence.
>
> Jonathan Steuer [234]

In contrast, the definition by Steven M. LaValle is expansive and driven by human physiology and perception. This definition is neither limited to a specific stimulus nor human species. According to LaValle, it is also about VR when, for example, gerbils are running on a spherical treadmill and control virtual movement through a projected maze [253].

> **Virtual Reality** induc[es] targeted behavior in an organism by using artificial sensory stimulation, while the organism has little or no awareness of the interference.
>
> Steven M. LaValle [235]

A *Augmented Reality (AR)* is a composition of computer-generated digital media, the physical world, and the human user. The media presents stimuli to any part of the human sensory system. Digital media is superimposed on the real world and spatially anchored in the physical environment. Further, the media is interactive and can be controlled and manipulated by the user. AR fosters the inability to distinguish between the real and virtual artifacts and ultimately creates experiences that enhance human perception.

Complementary to AR, Falk et al. defined the term *Amplifying Reality* [57]. In contrast to Augmented Reality, Amplifying Reality targets the publicly visible augmentation of expressions of objects or humans using embedded or wearable computing. Amplifying Reality differs in key aspects to AR so is not included in the scope of this thesis.

> **Augmented Reality** allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. Therefore, AR supplements reality, rather than completely replacing it. Ideally, it would appear to the user that the virtual and real objects coexisted in the same space [...]. AR [are] systems that have the following three characteristics:
>
> 1. Combines real and virtual
> 2. Interactive in real time
> 3. Registered in 3-D
>
> This definition allows other technologies besides HMDs while retaining the essential components of AR.
>
> Ronald Azuma [17]

If the definitions of virtual and augmented reality, or respectively the technologies interweave, a precise classification is not always possible. Milgram defined the AR-VR continuum ranging from the extrema of completely virtual to completely real, namely virtuality and reality [176]. Any computer-simulated experience that enhances or modifies the perception of the environment can be mapped in this space to a certain degree. The described continuum is visualized in Figure 2.1.

Recently [232] investigated the diversity of different MR definitions. Still, there is no established framework or definition. Most research in the field of MR is derived from the definition from Milgram and Kishino work. They define *Mixed Reality* as an environment in which real and virtual objects are presented in

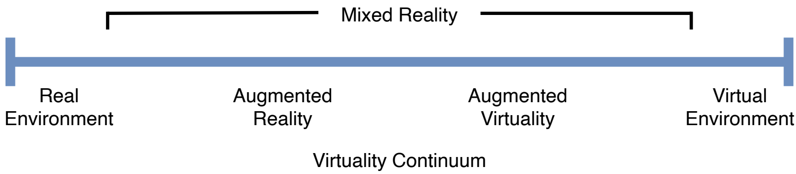


Figure 2.1: Representation of the reality–virtuality continuum [175].

coexistence through one display [175]. The proportion of virtual and real objects can vary between the extrema of the real and virtual environments.

> A **Mixed Reality** environment is one in which real world and virtual world objects are presented together within a single display; that is, anywhere between the extrema of the virtuality continuum.
>
> Milgram and Kishino [175]

The Mixed Reality definition by Milgram is tied to graphical displays that partially immerse the user. The derived taxonomy is composed of three dimensions: The extent of world knowledge defines the available details of the world being displayed; the reproduction fidelity describes the quality and realism of the displayed world; and the extent of presence describes the degree to which the user is immersed in the synthetic world [176]. Existing display systems drive the definition of Mixed Reality and, therefore, are restricted to the visual sense. In this thesis (cf. Chapter 9), we will describe an extended and broadened taxonomy mapping all human sensors in one unified space.

2.2 Presence and Immersion

The quality of augmented and virtual reality experiences is described utilizing the concepts of immersion and presence, where *immersion* [230] describes the capability of stimulating the human senses and creating a sense of presence.

> **Immersion** is a description of a technology, and the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant
>
> Slater and Wilbur [230]

In this sense, immersion is realized by overlaying and stimulating as many human senses as possible. The quality of the technology and the related degree of immersion can objectively be assessed as the characteristic of the used technologies. Immersion describes how well the illusion of the virtual environment is transferred. For the human visual sense, this could include the quality of the display, effects of stereopsis, Field of View (FOV), and accuracy of the facilitated tracking technology.

> **Presence** is both a subjective and objective description of a person's state with respect to an environment.
>
> Slater and Wilbur [230]

In contrast to immersion, *presence*, is the subjective response of being and acting within the immersive environment. The feeling of presence is the natural consequence of displaying virtual stimuli to the human sensory system. Research distinguishes between physical presence, social presence, and co-presence. In this thesis, we focus and measure physical presence as a consequence of visual, haptic, and auditive stimuli [261]. The recurring effort is to expose humans to highly immersive systems that convey a virtual environment in which their interaction is more engaging than in the real physical world.

2.3 The History of Mixed Realities

AR and VR have been worked on for several decades before achieving the capabilities that these systems are offering today. In this section, we give a quick technology-driven introduction to the history of augmented and virtual reality. This overview focuses on development milestones with a shift towards hard- and software achievements that were utilized in the course of the presented research. A non-comprehensive chronological map laying out these events is visualized in Figure 2.2. For an improved understanding and clarity, it is separated in VR in the top and AR at the bottom row.

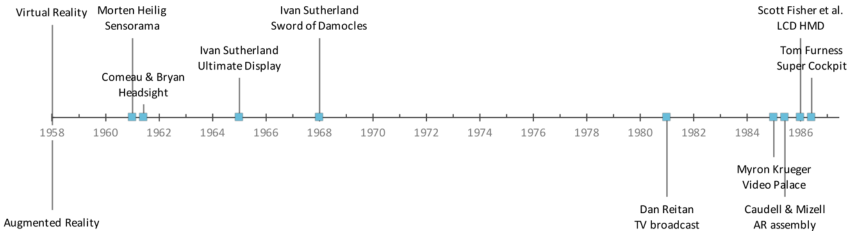


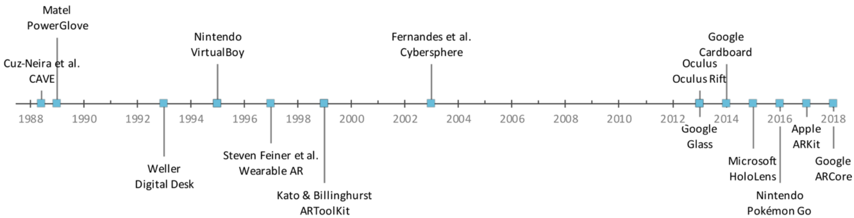
Figure 2.2: Key milestones in Virtual and Augmented Reality history.

2.3.1 Virtual Reality

The development of the first VR system was initiated more than half a century ago. In 1961, Morton Heilig patented a multi-sensory machine [99] that is now considered as one of the first immersive VR systems. A year later, he built the so-called *Sensorama* machine that provided auditory, visual, haptic, and olfactory stimulation that allowed users to immerse themselves in different scenes. Early VR development was further shaped by Ivan Sutherland with his vision of the Ultimate Display [240]. Sutherland paved the way to what is today known as VR, and his work in computer graphics at Harvard University on the *Sword of Damocles* is considered as the first interactive VR Head-Mounted Display (HMD) [241] that allows users to explore computer-generated virtual worlds. To generate a sense of depth and simulate a single three-dimensional object, Sutherland's HMD presents two side by side stereoscopic images to each eye of the user. These fundamental design principles of a stereoscope had already been researched by physicist Charles Wheatstone in his first article 1838 [258] and are still implemented in current HMDs.

The military was also exploring the potential of Virtual Reality at this time. Comeau and Bryan developed *Headsight*, a motion-tracked system with a stereo display that can be considered as a precursor to the HMD. This system allows the user to intuitively inspect a potentially dangerous location via remote cameras that are synchronized to the head movement.

The United States Air Force adopted this new technology and integrated virtual reality flight simulators in training to improve pilots' performance. Around 1986 Tom Furness directed the Air Force's Super Cockpit program [246]. This program aimed to create a more natural perceptual interface and reduce the complexity of a cockpit, and enabled Furness to raise issues of human-factors that would later confront the developers of today's virtual reality technologies.



Continuous research and consecutively more and more devices resulted in lighter and cheaper hardware. In the late eighties, specific peripheral devices for interactive VR environments were developed. The PowerGlove by Mattel is a glove that is tracked in space and enables gesture input. In the same decade, Scott Fisher integrated the first Liquid Crystal Display (LCD) into an HMD to significantly reduce its weight [65].

Parallel to the advances of HMDs alternative setups evolved. Cruz-Neira et al. [44] used several projectors that were directed towards walls, ceiling, and floor of a cubic room and surrounded a viewer with projected images to present a virtual environment. The users' head position was tracked within this room (*CAVE*) to support a viewer-based immersive virtual reality interface. They already recognized that VR interfaces must support more than one user within the same environment to become a successful tool. Later, this approach was further extended with the *Cybersphere* [62]. The projection system uses a large semi transparent sphere to project onto. The user is positioned within this sphere and can freely move around in a natural way. The users' motion is translated into the virtual environment creating the illusion of moving towards the scene.

From 1990 VR was expected to be ready for commercialization through video arcades. *Virtuality* by the Virtuality Group was one of those systems offering the first multiplayer VR gaming experiences [14]. However, most of the implementations failed to deliver the expected results. Nintendo, for example, released the *Virtual Boy*, a VR gaming HMD that commercially failed due to insufficient comfort and interactivity due to its static setup. The expectations faded and did not reemerge for two decades.

The second wave of VR started around 2012 with a successful crowdfunding campaign for the *Oculus Rift*. The Oculus Rift Developer Kit (DK)1 HMD was finally released in 2013 with a 90 degrees FOV and 3 Degrees of Freedom (DOF) head tracking. Subsequently, Oculus released its successor, the DK2, featuring improved displays and 6 DOF head tracking. In 2014, Oculus was purchased by Facebook.

Compared to the *Ultimate Display* presented almost 50 years ago, while the introduced devices were built on the same fundamental constructions, now the necessary computational power was available to render compelling interactive virtual environments in real-time. The DK1 and DK2 enabled researchers, developers, and game designers to explore the new advances at an early stage, and confirm the already existing fundamentals.

With the release of the Google Cardboard and more than 10 million shipped units, VR was becoming widely available for the general public. Google Cardboard is a very low-cost VR system made out of cardboard and simple lenses, and is powered by a smartphone. It was intended to encourage interest and development in VR applications. Nintendo also adapted the idea and built an HMD again, this time on the basis of the cardboard concept. By today, many major tech and entertainment companies have developed their specific VR HMDs tailored for diverse use-cases such as marketing, education, gaming and entertainment.

2.3.2 Augmented Reality

The early vision of what is today known and defined as Augmented Reality (AR) was also, to a considerable extent, shaped by Ivan Sutherland and his development of a head-mounted three-dimensional display [241]. Inspired by Sutherland's vision, Myron Krueger developed Videoplace: A projector-camera system that combines the user's live video feed with the augmented environment to create the illusion that the participants are in an interactive environment that reacts to pose and movement in real-time.

Early Krueger described the Videoplace as artificial reality and proposed to use the system for teaching and to support teachers to modify how and what was taught. Students would be able to explore the laws of cause and effect in scenarios previously designed by a programmer. Some years later, Krueger proposed to use Videoplace as an elementary telepresence system in which users could be placed at remote locations but share a joint video experience to interact and communicate [150].

The first time AR content was broadcasted on television was around 1981. Dan Reitan and his team were primarily responsible for the first weather forecast studio views. The early concepts only included several weather radar images and abstract symbols that were mixed for television weather broadcasts. Quickly stadiums were equipped with AR technology to augment sporting events and make important information like the first down line in Football visible to the viewer.

With the advances in tracking technologies, computational power and live processing sporting events in complex environments without geostationary reference points began to be augmented in real-time. The Americas Cup AC LiveLine System, for example, inserted graphical information into the live television feed of a sailing race that was transmitted from helicopters [111].

Increasingly industry was also investigating the potential of mixed realities to improve product quality and efficiency as well as reduce costs. To tackle the manufacturing challenges of sophisticated machinery, engineers at Boeing implemented a tracked see-through HMD to overlay diagrams in the real world at specific positions to support manual assembly [37]. Caudell also coined the term "Augmented Reality" with his work at Boeing.

Digital content that augments the environment can be displayed to the user by either instrumenting the users with projectors, glasses or HMDs or by instrumenting the environment. One of these environment-instrumenting technological approaches is to deploy projector-camera systems that observe and augment the environment. With the DigitalDesk, in 1993 Wellner presented such an interactive system that is built around an ordinary physical desk. By augmenting the tabletop and physical paper, this systeme lowered the burden of switching attention between screen and paper [257].

With this upcoming movement and the growing research direction of wearable computing, Steve Mann and his research group investigated new possibilities and developed several wearable AR prototypes. In separate research, Steven Feiner also analyzed how wearable mobile computing and AR might together support everyday interactions. At the end of the 20th century, Feiner et al. developed the Global Positioning System (GPS) supported AR prototype that allowed the wearer to explore an urban environment with the untethered freedom of mobile computing. They utilized a head-tracked, see-through, head-worn, 3D display to implement their application [60].

Anchoring virtual content in the real world was still a cumbersome and computationally expensive task. Hirokazu Kato streamlined this process with the release of the ARToolKit [124]. Later, this software was made available for mobile devices to capture real-world actions and combine these with interactive virtual objects. The highly specialized successor frameworks like ARCore [79] and ARKit [12] took advantage of these early works. These tools and frameworks are discussed in Section 2.4.3 in more depth. In the following years and continuing to this day, we have witnessed the shift from highly expensive fixed setups with heavy backpacks and separate displays to mobile and cheaper solutions integrated into average camera-equipped smartphones. Today, smartphones can render and display basic AR and VR experiences and provide widespread availability.

Augmented reality gained new momentum around 2008 with the availability of AR capable smartphones. Rather simple AR capable applications like travel guides targeted the consumer market but did not become widely used, but over time, technological advances replaced the bulky portable computers carried in backpacks. Smartphones and AR capable devices became lighter, smaller, and more powerful and are nowadays capable of displaying interactive AR content to the user. With the release of the smartphone Application Pokémon Go in June 2016, more than 45 million users globally were soon able to experience mobile handheld AR. The bright combination of AR and, for many users, well-known characters created a fascination around this AR experience that continues to today.

Initialized by Steve Mann, wearable computing matured, and the first consumer AR glasses were announced in 2012. However, the term AR glasses was misused for marketing purposes, even though the provided hardware was not entirely in line with established definitions of AR. In early 2012 first optical HMDs like Google Glass or Epson Moverio BT-100 were presented, starting the era of fast-paced evolution of head-worn displays. One of the next milestones in hardware development was the release of the Microsoft HoloLens. The augmented reality smartglasses combined tracking, display, battery, and processing unit in a self-containing device allowing the user to move around untethered. The HoloLens was utilized in multiple research projects to investigate novel AR concepts. Several reference devices were subsequently developed, building upon the design of the HoloLens to reduce the costs of these devices further. Recently the second generation of the HoloLens was announced improving interaction metaphors, the FOV, and comfort.

2.3.3 Summary

Virtual and augmented reality development has significantly progressed and is emerging out of its infancy. It is now being used in various domains, providing immersive experiences including art, architecture, commerce, education, entertainment, tourism, healthcare and military. The first hype of virtual and augmented reality allowed researchers to explore and understand the fundamental characteristics but did not reach out to the consumer market since the technology was expensive, bulky, and content creation was cumbersome. Now, VR and AR can be experienced on ubiquitous devices, namely our smartphones. Hence, these experiences are reaching out to the general public, immersing more and more users. Nevertheless, mixed reality is not omnipresent nor has acquired a state of everyday use. However, the fast-paced development accelerates the possibilities, and future AR and VR experiences will become readily available.

2.4 Technology and Modalities

The human sensory system is commonly recognized as vision, hearing, touch, taste, and olfaction [32]. In this section, we will keep to the five primary senses, nevertheless emphasizing that the human has even more subtle senses like proprioception, balance, temperature, pain, and hunger. Today, a multitude of technologies are available to stimulate any of these sensory systems and manipulate humans' perception of the world around them. In this section, we will briefly introduce various output modalities that stimulate any of the human senses. To better understand the different technologies, we will highlight the unique advantages and disadvantages to enable researchers and practitioners to choose the best technology for a given purpose (cf. Chapter 10).

Similar to the presented output technologies, we will continue with the different input modalities. In comparison to desktop computing, mobile mixed reality environments offer advanced input modalities through the increased variability of the systems and embodied input possibilities.

2.4.1 Output Modalities

In this section, we present and discuss different technologies to present computer-synthesized stimuli to the human sensory system, including vision, hearing, touch, taste, and olfaction.

Vision

While each sense takes an important role, humans usually tend to rely mostly on sight. Over the last decade, various technologies have matured and are today suitable to realize visual MR environments. Mixed reality displays can be categorized into stationary and mobile technologies, at which mobile technologies can be further divided into wearable or handheld technologies.

Displays - Stationary To present VR environments, large high-resolution displays or back projection is facilitated. The presented video is continually adjusted to the viewers' position to retain the virtual world's perspective and illusion. One of the most prominent setups using back projection is the cave automatic virtual environment (CAVE) [44]; a room whose walls, ceiling, and floor surround the viewer with perspectively corrected images. Today, these setups allow very controlled and high-quality presentation of content without necessarily instrumenting the user, which allows for a natural exploration of the space [97]. A particular issue with CAVE setups is the large physical footprint and the high costs for the setup and maintenance.

Displays - Mobile The latest smartphones and tablets contain high-resolution displays, fast processing units, and sensors that allow the rendering of mobile Mixed Reality Experiences (MREs). Spatial AR is realized by tracking the environment and presenting the augmented live video feed of the built-in camera. In conjunction with a head-mounted viewer for the smartphone, low-fidelity VR can be experienced. Mixed reality presented through smartphones stands to benefit in particular from its widespread availability and is successfully implemented in social media platforms, games [196], or integrated marketing programs [218]. The availability of specialized developer kits simplifies the creation of mobile AR experiences and makes them even more accessible (cf. Section 2.4.3).

Projection - Stationary Besides back-projection for CAVE setups, projectors are utilized to directly present perspectively corrected images into the real world, and hence directly augment the physical space. Usually, several calibrated projector-camera pairs are used to project the virtual information onto the physical world. Projection mapping is used to precisely align the information within the physical space. Therefore a digital representation for the physical space is needed, which can nowadays be created in real-time using depth sensors [94, 259].

Currently, projected augmented reality is one of the most promising technologies for stationary setups since it offers high-quality imaging, supports simultaneous use, and requires little setup since advanced calibration routines exist. Further, projected mixed reality systems circumvent the limitations of HMDs and offer a collaborative environment. Stationary projection-based systems were successfully deployed in various domains, including art, education, entertainment, and industry.

Projection - Mobile Due to their mobile character, portable projector-camera systems face additional challenges and limitations in contrast to stationary setups. Conceptually portable MR projectors do not differ from stationary setups other than the fact that the projection mapping needs to be adapted to the dynamic environment in real-time. Form factor and power supply naturally limit the luminous intensity, and virtual objects seem to appear transparent. Most projectors have a limited focal range and require constant refocusing if deployed in a dynamic setup. These limitations can be circumvented by laser projectors that are a reasonable alternative since they are focus-free by design [219]. Yet, not fully matured, mobile MR camera-projector systems are mounted on the helmets [43, 72] or backpacks [260] for research purposes.

Head-Mounted-Displays - Wearable Recently, Head-Mounted Displays (HMDs) gained much attention since several major companies released commercially available augmented and virtual HMDs. An overview is presented in Section 2.4.4. Modern HMDs are very specialized and adapt technology from the smartphone sector, including small high-resolution screens, gyroscopes, accelerometers, and other sensors to track hands, head, or body posture.

Virtual reality HMDs are worn on the head and provide separate images for each eye through a small opaque display and lenses. Hence, the user cannot visually perceive the real world around them but experiences the virtual world that adapts to the head's movements. In contrast, augmented reality HMDs use either video see-through or optical see-through display technology to permit perceiving the real world. Independent of the display technology, virtual content is superimposed onto the real world. While video see-through offers the best flexibility if augmenting the world to a large extent, optical see-through offers an unhindered view of the real world, accompanied by improved safety. However, this comes at the cost of a limited FOV for augmentations. Unfortunately, today's devices are still bulky and cumbersome to wear for extended periods.

Touch

The human somatosensory system can distinguish between kinesthetic and tactile feedback. Kinesthetic feedback is sensed through proprioception and is about the position of joints and muscles. In contrast, haptic feedback is sensed through receptors in layers of the skin that perceive pressure, vibration, temperature, and pain [233].

Haptic feedback through vibration can be generated with eccentric rotating mass or linear resonant actuators. In both cases, weight is moved or rotated to create a reciprocal effect. Actuators are installed in controllers, gloves, or vests to transfer feedback to dedicated body locations.

Other devices[4] use graspable proxies like pens or balls that are pivot-mounted and provide computer-controlled feedback when moved. Hence, users can explore virtual objects through the movements of the proxy. While these devices provide precise kinesthetic feedback, the interaction and exploration space is somewhat limited.

An array of ultrasonic transmitters can be used to emit focused ultrasound beams that create three-dimensional haptic shapes in mid-air [159]. This technology creates the sensation of pressure or vibration and gives the user the ability to feel virtual objects. Similar to graspable proxies, the sensation space is small, and the created pressure is limited. However, ultrasonic based haptic systems do not require users to be instrumented and can provide rendering accuracy up to millimeter scale.

Hearing

In a broader sense, two of the most prominent examples of augmented hearing are hearing aids and active noise-canceling headphones. Both combine the same core technology consisting of microphones, speaker drivers, and sophisticated signal processing units that modify the audio signal to the users' needs. Modifications could include the intelligent improvement of speech, feedback management, or removing unwanted ambient noise.

The human auditory system can distinguish between sounds emitted from the front and back or up and down. To create the illusion of sound emitted from a position in a virtual world, the timing, volume, and resonance must be carefully controlled. To further enhance this illusion and compensate for only using a headphone, a head-related transfer function can reproduce reflections caused by the shape of the torso, head, and ears. Current virtual and augmented reality engines already support the creation of spatial anchored auditory stimuli.

Taste

The human gustatory system can distinguish between the five basic tastes sweet, sour, salty, bitter, and savory. There are two established technologies for simulating the sense of taste; digitally controlled emission of chemical substances that stimulate the taste buds or electrical and thermal stimulation of the taste buds [172, 184, 203]. Despite these technologies' viability, simulating taste is still in its infancy and requires further analysis, research, and development.

---

[4] https://www.3dsystems.com/scanners-haptics#haptic-devices

Olfaction

The stimulation of the olfactory system in augmented and virtual reality has been largely neglected. Hence, technology to create olfactory stimulation is currently limited. A stimulus is produced by vaporizing liquid gels or waxes storing the odor, using heat or electrostatic methods. For research, computer-controlled scent palettes with several scent chambers are used [21]. Pressurized air is used to transfer the odors to the user. Similar to the presentation of synthesized taste, the latest research suggests the stimulation of the olfactory system by electrical stimulation [236].

2.4.2 Input Modalities

Novel new devices and sophisticated sensing technologies allow for an extended range of input modalities for explicit and implicit interaction while being immersed in MREs. Here, we group the distinct input modalities into peripheral and embodied interaction. This section comprehensively describes the different input methods with MR systems and discusses the advantages and disadvantages.

Peripheral Interaction

Dedicated peripheral devices give users precise input control and provide natural haptic feedback. Most handheld MR peripherals provide accurate motion tracking and responsive touch control. In the following, we will highlight and discuss selected peripheral devices for explicit input in MR.

Keyboard Well established for desktop computing, keyboards provide easy high-throughput generic text input for computing systems. Keyboards are also suited for productivity and tabletop interactions in augmented reality. However, in virtual reality scenarios, physical keyboards create unique challenges since they cannot be perceived visibly by default. We and others have addressed these challenges in several works [85, 137, 144]. From a user experience perspective, the keyboard is an excellent opportunity to blend familiar and new human-computer interaction interfaces into one MR experience.

Smartphone As outlined earlier, today's smartphones are capable of compelling MR experiences. At the same time, the multitude of sensors allow us to utilize them as versatile input devices. In handheld AR scenarios, direct multi-touch and device orientation can be used for implicit and explicit interaction. Alternatively, the smartphone can elegantly serve as a ubiquitous input controller in mobile MR settings. With dedicated applications, the smartphone connects to any output modality. Then, the different sensors and the screen can be repurposed and provide rich input possibilities (cf. Chapter 8) [19, 178].

Controller An input controller for MR can have endless variations in shape, utilization, and supported input functionality. The handheld remote controller can be as simple as a single button with accelerometer reporting only simple button presses and the controllers' orientation or can be a fully featured controller equipped with several analog and digital buttons, touch-pads, and full 6 DOF motion tracking.

In addition to generic controllers providing comprehensive input possibilities, there are also controllers, like the keyboard, serving very specific MR scenarios. The PlayStation 4 VR Aim Controller, for example, comprises motion tracking, used in weapon design with intuitively placed buttons to enhance VR shooter games. For precise movement in VR environments, specialized treadmills[62] and bikes exist that translate the physical movement into the virtual world thus overcoming any physical space limitations.

Embodied Interaction

Humans involve the abilities of their bodies when naturally interacting with the physical world. Embodied interaction adapts this concept and allows humans to interact with computing interfaces by naturally using their physical body. In the following, we present embodied interaction modalities that are suited for MR scenarios.

Head Pose Humans naturally orient the head towards the objects they want to interact with. This concept easily translates to MR when the user is wearing an HMD. Fusing gyroscope, compass, and accelerometer data, the head pose can be determined. Mixed Reality applications can react to changes of the head pose accordingly. In combination with a visual head pose pointer, users can comfortably navigate interfaces in MR environments [151].

Body Pose Due to the HMD and controllers, the position of head and hands is known in typical VR setups. Additional trackers attached to the feet make a full-body pose estimation possible. Alternatively, motion tracking or 3D-depth cameras could be used. Tracking the body posture in real-time enables MR to replace a person's real body by a life-sized virtual avatar along with the possibility for implicit interaction through specific body poses. Recently, Hoppe et al. [114] used a pressure sensor to determine the user's gait and changed audio feedback in a VR experience accordingly to increase the sense of presence and modify gait behavior.

Location The user's location in the real world is a piece of substantial contextual information that can be used as implicit input for any mobile MR application: On a small scale, with cameras and stationary tracking systems; on a large scale, with satellite-based navigation systems like GPS or Galileo. Knowing the user's location allows mobile MR applications to provide context-aware augmentations. Applications like Google Maps or Pokémon Go successfully integrate location-based augmentations to enhance the user experience.

Gestures Gestures have a long tradition in human-computer interaction, and several technologies have been developed to sense gestures. For AR systems, structured light depth sensors or Infrared Radiation (IR) stereo cameras are used to track the hands and detect gestures. Virtual reality systems additionally draw on sensor-equipped gloves to track hand movements. The human hand's flexibility and speed allow for endless uni- and bimanual gestures for implicit and explicit interaction. Unfortunately, direct interaction with virtual objects and non-instrumented hands lacks haptic feedback, and further, fatigue can occur when interaction over extended periods is required.

Eye Gaze Current camera-based eye-tracking technology for HMDs can assess different features of the users' gaze and sensitive biometric data. These data can provide essential information to enhance other input, track attention, and allow for implicit and explicit interaction in MR environments. Moreover, advanced cameras enable user authentication through biometric iris authentication. Thus shared MR systems can quickly adapt settings like interpupillary distance or account management to different users.

Unfortunately, accurate eye tracking still requires a calibration process for each user and is therefore not suitable for spontaneous interaction. Further, the Midas-Touch problem persists. Interactive systems need to distinguish between eye movement for exploration of the physical and virtual environment and explicit interaction. Therefore, eye tracking is often combined with other input modalities, e.g., gesture, controller, or voice.

Voice Nowadays, voice user interfaces are the primary form to interact with virtual assistants integrated into smartphones and smart speakers. In MR environments, voice interfaces provide an additional mode of input and can create a natural and familiar way of interaction. Still, voice interfaces are prone to certain limitations and offer only limited flexibility, speed, and usability [226]. When implementing voice interfaces, it is crucial to consider the user, social context, and environment.

2.4.3 Engines, Toolkits and Frameworks

The increased prevalence of MR devices requires tools to develop interactive and immersive experiences. Existing engines that handle complex calculations and simulations have been extended to serve the advanced requirements to render MR. Further, existing tracking solutions were adjusted, and new toolkits for mobile devices were developed. This section looks at different engines, tools, and frameworks that have accelerated MR development.

The two companies *Unity* Technologies[5], and *Unreal* Engine[6], are the leading game engine providers. The Unity game engine is, in particular, specialized in supplying games on mobile- and across-platforms. Unity Technologies extended the Unity game engine to support Android and iOS powered smartphones as well as emerging AR and VR devices. With a dedicated assets store, game developers can now quickly purchase or import digital assets to speed up the creation process. Unreal Engine provides with its engine the digital infrastructure for computer games and computer-powered AR and VR experiences. Recently they partnered with leading HMD manufacturers and optimized the engine for AR and VR gaming.

In a mobile context, stationary tracking systems are not feasible. Smart software solutions were invented to overcome the limitation of lacking physical space and computational power in this context. Kato and Billinghurst proposed to use computer vision techniques for fiducial marker tracking and thus realize an AR conferencing system [124]. Tracking the user's viewpoint allows displaying virtual information spatially anchored within the built-in camera feed on the mobile display. Their approach was directed to one of the first open-source AR Software Developer Kits (SDKs) for mobile devices, namely, the ARToolKit [123].

Using the same underlying principles of computer vision, PTC provides, with Vuforia[7], a powerful tool to create MR experiences. The refined algorithm, advanced processing capabilities, and sensor fusion can work alongside conventional marker tracking to track 3D objects, advanced viewpoint estimations, and interaction by occlusion detection within the camera image.

Today, the dominant mobile operating systems, Android (Google) and iOS (Apple), provide an Application Programming Interface (API) to overlay virtual content onto the real world seen through the smartphone built-in camera. Both

[5] `https://unity.com/`

[6] `https://www.unrealengine.com/`

[7] `https://developer.vuforia.com/`

ARKit [12] and ARCore [79] integrate marker-based and markerless motion tracking, in addition to other environmental properties like illumination. With today's variety of engines, toolkits, and frameworks, the barrier to create captivating MR experiences is lowered, yet still existent.

2.4.4 Commercial Mixed Reality HMDs

Augmented and virtual reality headsets have a long history starting with development and deployment in the context of research. With smartphone technology maturing and becoming widespread, the manufacturing costs for sensors and displays decreased. Consequently, MR HMD became smaller, more powerful, and finally entered the consumer market as affordable gadgets. In Figure 2.3, we present some low fidelity and high fidelity MR devices that our research probes are based on. In the following, we provide a selected overview of commercial MR HMD we examined during the course of this thesis. The subsequent Table 2.1 lays out the initially supported input and output modalities of these HMDs.

Augmented Reality

Microsoft HoloLens The HoloLens is an untethered optical-see-through HMD developed by Microsoft. The HMD comprises several sensors, including an Inertial mMeasurement Unit (IMU), microphone array, camera, ambient light, and low energy depth cameras to understand the environment and users' intentions. The wearable is equipped with a tinted visor surrounding transparent combiner lenses in which the rendered images are displayed. In 2019, three years after the first release, Microsoft announced the HoloLens 2 with improved FOV, comfort, processing capabilities, and eye-tracking. Further, the interaction concept was refined, supporting new gestures and direct manipulations.

Magic Leap One The capabilities of the MagicLeap One are very similar to the HoloLens HMD. The most significant differences are comfort, the provided controller, and the larger FOV. The HMD comes with a tiny wearable computer that powers the MR experience. Hence, the HMD itself is significantly lighter. In contrast to the HoloLens, the controller promotes 6 DOF interaction and contains several buttons and a touchpad.

Meta View Meta is one of many startups developing augmented or virtual reality headsets. The Meta 1 augmented reality glass is based on the reference implementation of the Epson Moverio MT-200 [56] with an integrated depth sensor and IMU for spatial understanding. The device requires a high-performance

Figure 2.3: High and low-fidelity AR and VR devices. Top: Microsoft HoloLens and Aryzon AR Headset. Bottom: Oculus Rift CV 1 and Google Cardboard.

computer for sensor processing and rendering. The follow-up model, the Meta 2, comprises a larger FOV, and refined tracking as well as interaction.

Google Glass In 2013 Google started selling the wearable HMD in the shape of eyeglasses to selected individuals. The mobile device features a touchpad on the side of the frame, a camera to take pictures, videos, and remote collaboration besides an adjustable display placed in the upper right corner of the user's FOV. Although the device does not fulfill all parameters of established AR definitions, it is often labeled as AR HMD in the media. Today, Google Glass is successfully deployed in industry scenarios, e.g. for order picking in warehouses [3].

Aryzon 3D AR The Aryzon 3D AR [2] is a low fidelity HMD made out of cardboard, plastic lenses, and semitransparent foil. It is powered by a smartphone that is attached to the headset, hence the quality of the experience also relies on the input and output capabilities of the smartphone. Due to the low price tag, this headset aims to make AR accessible to a large population. Equivalent to the

	Embodied				Peripheral			
	Headpose	Eye Gaze	Gesture	Voice	C Motion	C Buttons	C Touch	Smartphne
HoloLens	●	○	●	●	◐	◐	○	○
HoloLens 2	●	●	●	●	○	○	○	○
MagicLeap	●	●	●	●	●	●	◐	◐
Google Glass	○	○	○	●	○	◐	●	◐
Aryzon	●	○	○	○	○	○	○	○
Daydream	●	○	○	○	◐	●	◐	○
Vive (Pro)	○	○	○	○	●	●	◐	○
Vive Pro EYE	○	●	○	○	●	●	◐	○
Vive Focus Plus	○	○	○	○	●	●	◐	○
Cosmos	○	○	○	○	●	●	◐	○
Oculus	○	○	○	○	●	●	●	○
Oculus Quest	○	○	○	○	●	●	●	○
Oculus GO	○	○	○	○	◐	●	◐	○

Table 2.1: Overview of out-of-the-box supported explicit interaction modalities of selected AR and VR systems separated into embodied and peripheral (Controller (C)) interaction. ● fully supported, ◐ partially supported, ○ not supported.

Google Cardboard for VR experiences, the Aryzon 3D AR is suitable for short, low fidelity experiences and rapid prototyping [187].

Virtual Reality

Oculus Rift The Oculus Rift Development Kit 1 is one of the first fortunate crowd-founded VR HMDs developed by Oculus VR. The start-up Oculus VR aimed for the development of a consumer HMD with a gaming focus. The first development versions with low-resolution displays and only a few sensors were shipped in early 2013. Over time, Oculus VR refined the HMD, added handheld controllers, and 6 DOF support for both. Today they provide a wide range of tethered and untethered (Oculus GO) VR HMDs for professional gaming or mobile VR experiences.

HTC VIVE HTC developed in cooperation with Valve the HMD HTC Vive. From the user experience, the devices resemble the Oculus Rift. However, from a technical point of view, they differ considerably. Compared to the outside-in camera tracking of the Oculus Rift, HTC opts for laser-based inside-out tracking. Similar to Oculus, the latest devices still require a workstation for rendering the virtual environment, but the HMDs are wirelessly connected to the host to prevent the user from getting entangled. Also, these devices now support camera-based eye tracking.

Sony PlayStation VR Sony announced to develop an exclusive VR HMD for the PlayStation 4 in 2014 and released the device later in 2016. The setup of the device is similar to Oculus and HTC products except for details like a lower resolution but higher refresh rate, and the proprietary connectivity.

Google Daydream Google Daydream View and Google Cardboard are HMDs that require a compatible smartphone as a display and processing unit. Cardboard is the handheld low fidelity VR version contrary to the Aryzon AR headset. The Google Daydream comes with a head strap, comfortable cushioning, and a 3 DOF controller with buttons. Virtual reality environments can intuitively be explored, but due to the simplicity, implicit interaction is limited.

2.5 Current Challenges

In the following, we indicate some of the current challenges for the design of MREs. These challenges can be classified into technological challenges, interaction methods, social challenges, and content creation.

Parts of the subsequent sections are based on the following publications:

- P. Knierim, P. W. Woźniak, Y. Abdelrahman, and A. Schmidt. Exploring the Potential of Augmented Reality in Domestic Environments. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, USA, 2019. Association for Computing Machinery

- P. Knierim, T. Kosch, M. Hoppe, and A. Schmidt. Challenges and Opportunities of Mixed Reality Systems in Education. In *Mensch und Computer 2018 - Workshopband*, Bonn, 2018. Gesellschaft für Informatik e.V

2.5.1 Technological Challenge

Hardware has developed rapidly over the last years, and the first completely self-contained wearable MR devices became affordable for researchers and end-users. Still, most devices are in their infancy stage, and the next generation of MR devices overcoming some of their limitations are soon to be released. These technological limitations include short battery life, a small field of view, poor image quality, and registration of content in unknown environments. Without accurate object recognition, the real and virtual objects are not aligned to each other, and the illusion of the strong coexistence and feeling of presence breaks. Further, users complain about an uncomfortable feeling when wearing MR devices for several hours. Due to the fast technical advances, these challenges will most likely be solved in the upcoming years. Research suggests, for instance, new optical lenses to integrate AR with the human eye. Moon et al. [179] propose pancharatnam-berry phase lenses as a specialized combiner with a smaller form factor and large aperture. We can expect that these technological boundaries will be overcome in the future. Hence, the research question of how to develop meaningful MREs is of even greater interest.

2.5.2 Interaction Methods

Past research discusses several approaches to interact with augmented and virtual realities. Currently, most MR devices support interaction via physical remotes, hand gestures, or speech (see Table 2.1). However, to enable collaborations with peers and virtual elements, implicit hands-free interaction should be supported. Speech allows this kind of hands-free interaction but is most likely to be unsuited to situations with background noise or absence of rich feedback, and could be socially inappropriate. Long-lasting sessions of mid-air hand gestures are also undesirable as an input modality since they cause fatigue [29]. It still is an open research question for the research community to identify alternative multi-modal interaction concepts and design expressive input concepts for MREs.

2.5.3 Public Awareness

The introduction of novel technology always poses several social challenges, including public acceptance and retention, in addition to security and privacy concerns. This also holds for MR systems [145]. Furthermore, AR can be a severe risk on the physical safety of oneself or others by distracting the immersed user from safely navigating and interacting in the real world [224]. Solutions need to be implemented to prevent misuse of MREs on the one hand but also prevent abuse of sensor data or the users' privacy. However, it is as yet unknown how to prevent information overload. Stimulating several human senses with virtual information at once can create an overwhelming feeling. To conclude, there is a strong need to identify new ways to manage and prioritize information.

2.5.4 Content Creation

Even though MR devices are still in their infancy, a good number of SDKs, API, frameworks and engines exist to support the development of MREs. Nevertheless, technical standards are present and slow down the overall development of content since the reusability and generalisability of content is restricted.

Another challenge to content creation for MRE is the need for 3D models. However, the creation and storage of large amounts of content requires time and investments. Platforms like the Unity Asset Store are a partial solution, but cannot cover everything. Still, the biggest challenge of creating content is not the development or authoring of the content itself. It is the ideation of meaningful use-cases where the technology can deliver a quantifiable benefit. Except for entertainment, most of today's MREs are showcases of the technology rather than delivering an accessible and useful experience to the users, and are far from being desired to be used daily.

Chapter 3

Understanding the User

Mixed Reality technology proposes an exciting prospect of engaging directly with the lived environment and augmenting everyday spaces with digital artifacts. Augmented reality games like Pokémon Go have successfully enhanced social interaction and active learning [224]. Despite these appealing qualities, MR is yet to enter widespread use.

Historically, the development of interaction techniques and applications for MR was slowed down by high equipment costs and technical complexity. However, recent technological advances, like powerful smartphones and the continuous development of MR glasses, indicate that MR technologies' widespread availability is a highly probable technical future. Consequently, understanding how MR can become part of our everyday spaces and change everyday interactions with technology and the world emerges as a relevant user-centered research question **(RQ1-RQ3)**. In contrast to the majority of past research where researchers and designers proposed potential usage scenarios, we take an alternative approach to investigate how users perceive viable use-cases for MRE. In this chapter, we help to understand the design constraints and consequences of deploying MR systems in domestic environments. We investigate the users' visions and attitudes towards MR in domestic environments as a fundamental prerequisite to enhance interaction in MR and circumvent potential pitfalls when developing future MREs.

In this chapter, we first review related works that address MR technologies and technology probes in a domestic environment. We then describe the methodology of the survey and probe in detail, accompanied by the results of these studies. Finally, we discuss insights, challenges, and opportunities for MR applications in domestic environments and beyond.

This chapter is based on the following publications:

- P. Knierim, P. W. Woźniak, Y. Abdelrahman, and A. Schmidt. Exploring the Potential of Augmented Reality in Domestic Environments. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, USA, 2019. Association for Computing Machinery
- P. Knierim, T. Kosch, M. Hoppe, and A. Schmidt. Challenges and Opportunities of Mixed Reality Systems in Education. In *Mensch und Computer 2018 - Workshopband*, Bonn, 2018. Gesellschaft für Informatik e.V

3.1 Related Work

The work presented in this chapter builds on past advances in AR and the development of technology probes as a research methodology. Here, we review research that motivated our study of domestication in AR and informed our choice of methodology.

3.1.1 Applications of Augmented Reality

AR is a combination of the virtual and real world, where virtual objects are superimposed in the surrounding environment in real time to enhance reality and user experience [17]. Past research explored extensively where and how AR could be applied to improve user experience or task efficiency. Thomas et al. explored how AR can be used to create playful experiences [244]. They designed ARQuake; an outdoor mobile AR game. While the game was positively perceived by the users, many interaction issues specific to AR were revealed. These include effective item selection, tracking, and multi-person collaboration. AR has also been used for therapy and studies showed that AR could be effective in other medical circumstances, such as treating phobias by reducing people's fear of insects or animals [157].

Education is another field of opportunity for AR. Lucklin et al. showed that AR systems could be used to motivate and engage children in learning activities [162]. AR learning experiences can also foster improved knowledge sharing [210] and make the unseen visible in physics lab courses [238]. Industrial applications have also been explored. Funk et al. showed that projected AR could contribute to efficiently training assembly line workers [70]. Further, Liu et al. showed that handheld AR devices with real-time feedback outperformed paper and picture instruction [156] in providing contextual training queues. The works listed above were included in a recent review of AR user studies by Dey et al. [47]. The review identified that past application areas for AR which were subject to user studies were primarily professional environments or actions connected with entertainment. The review also highlighted that user knowledge in the AR field is primarily based on technology-driven within-subjects experimental research and more engagement with users in the field was required.

Our work is interestingly different from past efforts as it explores a frontier beyond the usual scope of AR applications — users at home. Further, instead of adapting a conservative experimental approach, we use a user-driven approach where we identify application scenarios through engaging with users in context.

3.1.2 AR systems with Potential for Home Use

Another topic addressed by AR research that could find its use at home was providing augmented senses. Fan et al. built SpiderVision [58]; a device that extended the field of view of the human eye. People using this device adapted to using a backward-facing camera as a 'third eye' attached to the back of their heads. Through AR technology, Jang and Bednarz enabled users to perceive and interact with real-time data provided by smart-home sensors [119]. They envision extending this concept to other domains like health. The AR health Application Mime [50] helps patients to analyze their blood at home. In this case study Djajadiningrat et al. illustrate the challenges of unassisted care at home and highlight that AR can act as an in-context manual for a novice user.

These design examples have shown that potential applications of AR may find use at home, helping users in everyday tasks and increasing their safety. In order to deploy these technologies in everyday settings, we must first know if and how they can be integrated into home environments. Our work aims at providing insights that could help translate existing knowledge about designing AR applications to design for the home.

3.1.3 Studying the Experience of Technology at Home

Our work uses a probe to study AR technology at home. Cultural Probes were first initiated by a group of designers under the lead of Bill Gaver. They wanted to explore new techniques to increase the engagement of the elderly in their local communities [28, 73]. The probes are inspirational objects designed specifically to prompt users to record their private life, ideas, and experiences [167]. The term 'cultural' indicates the type of the technique used. Thus, it can be replaced with other techniques such as empathy or technology [28]. Probe kits can include items such as disposable cameras, maps, stickers, lists of instructions, diaries, illustrated cards, and pre-stamped postcards accompanied by open-ended provocative tasks [167].

Gaver and his team describe the probes to the participants as a tool through which designers could understand users and vice versa. Probes create a bi-directional understanding between the designers and the participants [73, 74], allowing users to become active co-creators in the design process through giving the designers the chance to deeply understand their culture, aspirations, dreams, and needs [28, 225]. What is important to our work is that a probe is a practice-oriented alternative to social science approaches to understand a user's environment [28]. Additionally, a probe can overcome problems in traditional data collecting methods such as limiting the view into a specific area, by acquiring an extended view into the user's life style [167]. However, traditional methods like interviews can also be employed as an assistant factor for probes to acquire a deeper insight into the user's life.

Culture probes are mostly used in two scenarios: First, exploring the implications of a new technology before making them publicly available [225]; and second, identifying problem statements, and exploring novel and creative ideas inspired by the participants. Here, we are more concerned with the first usage. Although prototyping techniques seem efficient in such cases and can be used to simulate the interactions with a new technology, they would not guarantee the same deep understanding of active engagement of participants [225].

We utilize the concepts of cultural probes in terms of a technology probe to investigate AR in domestic environments. The extensive history of probes in Human-Computer Interaction (HCI) inspired us to explore AR at home using a probe.

3.2 Technology Probe

Next, we endeavoured to build a deeper understanding of the possible augmentations for everyday objects that we observed in the survey. As survey participants were eager to suggest AR solutions, we also explored what potential benefits of AR they identified and they anticipated their experience of everyday tasks to change when using AR. To that end, we used a hybrid methodology that combines a technology probe with photo elicitation.

We encouraged participants to generate AR user scenarios and interact with the technology. We combined a complete introduction to AR with a photo elicitation approach [80] and pre- and post-study semi-structured-interviews using the contextual laddering technique [249] to explore implicit insights on domestic AR usage. To facilitate the photo elicitation, we built an application inspired by Snapchat. The application enabled the participant to quickly take pictures and annotate them with stickers, emojis or text. All annotations were freely positioned, scaled, and rotated. After annotating, the original and annotated images were automatically saved to a cloud service. In contrast to previous works, where disposable cameras were deployed [73], storing images in the cloud allowed us to observe the study process remotely. Each participating household was given a Nexus 9 tablet with Android Nougat. The application including the user interface for annotating images is depicted in Figure 3.1. Thus, our solution combined the immediate access of a cultural probe with the illustrative benefits of photo elicitation. The study set-up enabled us to prompt users to imagine their desired experiences with AR and create them in a rapid. The visual qualities of an annotated photo helped communicate their visions effectively. This way, we opted for providing the participants with extensive means of expression rather than asking them to build prototypes.

3.2.1 Participants

We tried to invite a diverse set of people living in varying households. We managed to invite families, shared flat communities as well as couples who live together. Further, we attempted to ask people with different backgrounds. We conducted our technology probe with 13 participants in four households. All the households had two to four individuals. One household was a shared flat and one kept a cat. Participants (6 female) aged from 11 to 48 ($M = 26.8$, $SD = 10.8$) took part in the technology probe and the interviews. Their occupations included pupils and students with different majors, teachers, consultants, and lawyers. Six

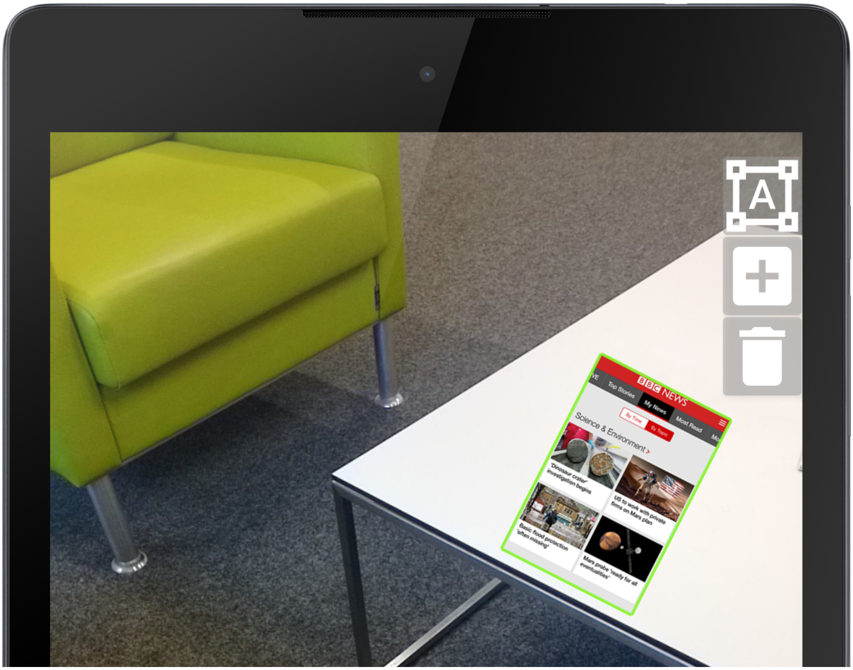


Figure 3.1: Tablet application to take pictures and annotate them. Top-right icons allow the user to add text and stickers or delete them. A news sticker is placed on the table.

were familiar with mixed reality applications, and one had used the Microsoft HoloLens before. Guardian consent was acquired for the participation of the minors in the study. Households ranged from 60 to 140 square meters ($M = 103$, $SD = 28.7$) in size with two to six rooms. Table 3.2.1 shows details about the participants of the probe.

3.2.2 Procedure

The probe lasted for 14 days and was divided to three stages. We outline the stages below.

ID	House	Occupation	Gender	Age	MRE
1	H1	student	male	24	●
2	H1	undergraduate	male	22	●
3	H1	student	male	22	●
4	H1	graduate	female	25	○
5	H2	student	female	27	○
6	H2	student	male	20	●
7	H2	teaching assistant	male	27	●
8	H3	teacher	male	48	○
9	H3	teacher	female	46	○
10	H3	pupil	female	11	○
11	H3	pupil	female	13	○
12	H4	lawyer	female	33	○
13	H4	consultant	male	31	●

Table 3.1: Demographics and previews experience of mixed reality applications of the technology probe participants.

Introduction to Augmented Reality

Before starting the study, we visited the participants' households. We collected their consent for taking part in the study and processing the images, and conducted a pre-interview to collect their demographics as well as their experience with AR. We then introduced the participants into AR, based on the definition of Azuma [17]. Afterwards, we explained the Microsoft HoloLens and gave each participant two small demonstrations. First, we showed them three holograms: a browser attached to a wall, a small city placed on a desk, and a globe set mid-space. Participants were instructed to walk around the holograms to understand that they are three-dimensional and fixed in the environment. Secondly, we demonstrated an interactive application which provides the ability to measure distances by placing two points in the environment. After setting the second point, a line connecting both points and the distance was shown. To prevent bias, we did not introduce speech and gesture interaction and used the provided clicker and gaze for interaction.

Start of the Probe

After the introduction to AR we set up the tablet and explained the photo-elicitation application to the participants. We explained in detail how to take pictures, annotate them and save them to the cloud. Participants were asked to place the tablet in a location in the home that would be accessible to all participants of the household. They were also asked to document their ideas using pictures and annotations.

Semi-structured Interviews

After 14 days, we revisited the participants' households to conduct the semi-structured group interview. In preparation for the interview, we copied all annotated images from the cloud to a separate tablet and preselected ten annotated images based on uniqueness and relevance to the online survey results for a detailed discussion. In addition to Figure 3.2 a more extensive selection of annotated images is archived online[8].

The first part of the interview focused on general AR usage in a domestic environment. We asked questions to understand how the families imagined using AR on a daily basis. We continued with the ten preselected images to get deeper insights about the favorite use cases and most useful situations. Participants had the chance to skim through all pictures afterward to recall any situation not mentioned before. The second part of the interview concentrated on social implications, concerns, and form factor.

We audio recorded the interviews and transcribed the interviews verbatim for post-hoc analysis. In an initial round, two coders used open coding to analyze 25% of the data gathered, then met to establish a coding tree. One researcher analyzed the remaining material. A final meeting was held to refine the coding tree and establish the emergent themes.

3.2.3 Findings

In this section, we present the findings of our probe. All study participants were positive towards the prospect of using AR every day. Most participants requested that AR capabilities be embedded in objects that they were already using on a regular basis, such as spectacles:

> *Since I need glasses anyway, I would definitely wear it all the time. –P2*

Further, we observed that some users welcomed features that required AR to be perpetually active. One wanted to use AR-based reminders because they considered themselves forgetful:

[8] `https://github.com/pknierim/Augmented-Reality-in-Domestic-Environments`

> *Well, I would definitely wear it all the time because I know I'm someone who needs to be reminded. –P3*

In contrast, some participants indicated that they envisioned that the usage of AR should be limited to the privacy of a home. They believed that AR devices could be problematic in social contexts:

> *I would rather use it in private situations when none is around. I don't think that's appropriate in a group. –P9*

We observed that, despite different views on the depth of AR adoption, all users contributed eagerly to possible users' scenarios and described their experience with AR extensively in the interviews. Our data analysis process revealed six themes of the experience of AR domestication: assistance, enhancing perception, social activity, device augmentation, and concerns. Next, we describe the themes in detail.

Assistance

We observed that participants often saw AR as a way to provide contextualized assistance. Users were eager to contribute new ideas for scenarios in which AR could overlay additional information. The context of specific actions was often explored. An often-mentioned use case was cooking:

> *First, we thought the recipe on the kitchen wall would be great, but then we figured out that we could have a real virtual coach who could stand next to you and assist you while cooking. –P8*

The family continued describing a fictional cooking scenario where a famous TV chef would provide cues on how to prepare the dish. They also reflected that they would have liked the cook to give them freedom and only appear when assistance was required. Participants also reported the desire to receive assistance based on location. Overlaying navigation cue using AR was mentioned by all households in our probe. One participant pointed to a scenario where they were riding their bike and both of their hands were holding the handlebars:

> *While riding my bike, I imagine navigation cues in front of my eyes. It could be so easy if it (AR) was always around. –P11*

Further, participants recognized that AR could be useful in scenarios that required the use of their entire body. An often-mentioned use case was sports. One user imagined a virtual trainer who could provide necessary exercise instruction:

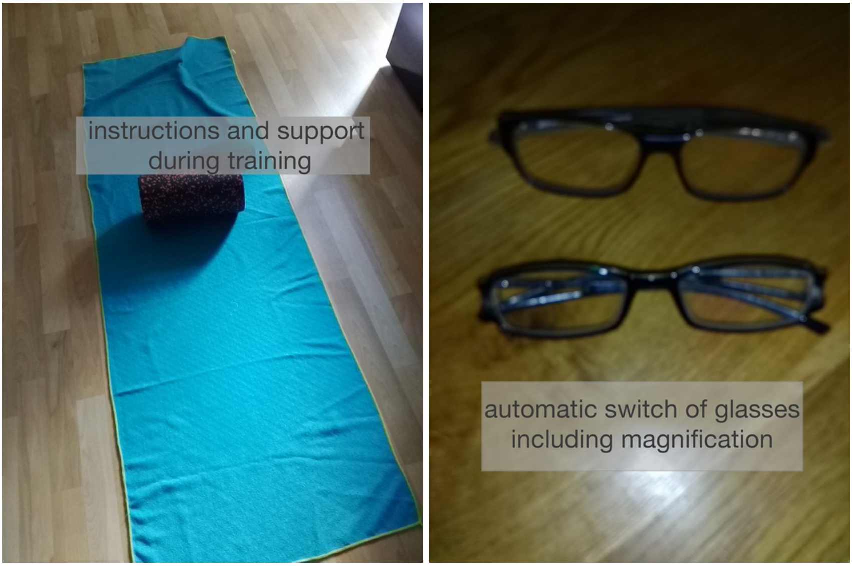


Figure 3.2: Four images created during the technology probe. From left to right: Assistance during exercise, enhanced perception due to magnifying glasses, an augmented stove with smart timer, and a door providing plenty of valuable information.

> *She (the virtual trainer) should see what I do and talk to me. I'd like it if someone was around to correct my mistakes to prevent injuries. –P12*

We also observed that users were eager to use AR in tasks that require an extended sense of spatiality, i.e. getting an overview of a larger space, dealing with distances between objects, spreading or aligning artefacts evenly. Home furnishing was one cited scenario:

> *For example, I could place the (virtual) shelf on the wall and check how it would look. Would it fit the new sofa that we are planning to order? –P9*

Finally, users would often cite AR as an opportunity to use their smartphone less in the context of assistance. Many participants remarked that information that they currently obtained using a mobile device could be displayed in the environment thus reducing the need for touch interaction:

You wouldn't have to pull out your phone and go to an app. Information would be shown to me directly in the real world, like the WiFi signal strength for example. –P5

Enhancing Perception

The participants in our probe also considered how the application of AR could offer them new possibilities to perceive the world. Users contributed insights on how they would like AR to augment their senses. One participant expressed that they would like AR glasses to replace conventional corrective spectacles:

It would be great if I did not have to change my reading glasses and varifocals all the time. –P2

Other users explored scenarios with more elaborate sensing capabilities that would give them new skills. For example, mood detection was discussed in one of the households:

[. . .] *if you could actually see a mood of a person. –P6*

Further, participants wondered if they could use AR to make themselves more aware of possible dangers or consequences of their actions:

What is bad for the environment OR what am I allergic to... I am wondering if there are any toxic or environmentally harmful ingredients in the shampoo. –P3

Obtaining additional non-sensory information about the environment was another often-cited case. Some participants wanted to interface with the history of the surroundings:

I would like to see additional information for a specific building or place of interest. For example, how did people live here back in the old days? –P8

We also noted that participants desired not only to be more aware of the properties of their environment but also about other people. A father expressed a wish for AR to allow him to monitor his children better:

If you have a baby and you are in the kitchen and the baby in the living room, I want to see the baby without leaving the kitchen. Like X-ray vision. –P13

Social Activity

Next, we show how participants imagined the role of AR in various social settings. All of the households recognized AR's potential to offer remote presence. One participant considered using AR to socialize instead of commuting to a sewing group meeting, which was troublesome. Interestingly, they suggested that using AR would enable them to participate in an organized group activity while still enjoying the comforts of their own home:

It is very time-consuming, and the sewing machine is heavy, but I like going to the meetings. Having a virtual meeting would be great. You could talk and listen to the people and the course instructor while working on your machine at home. –P2

Our participants also expressed that they envisioned AR helping with social coordination. Users built scenarios where a cue embedded in the environment provided a point of reference and helped build mutual understanding. One user reflected that the process of choosing a film to watch could be more effective if augmented with AR:

If we planned to go to the cinema, we'd look together at the program. It's great when we see the same information so you can point to it. –P10

Some Participants remarked that they were open to sustained presence through VR and willing to engage with a virtual character over longer periods of time. Engaging with virtual companions through AR was an interesting perspective:

> *Having a virtual pet, like a Tamagotchi 3.0, could prevent someone from feeling lonely, or ensure that the family is ready for a real pet. –P13*

Sharing content through AR and simultaneous media access was also seen as beneficial. Users remarked that AR offered opportunities for rapid media sharing with specific counterparts:

> *It would be wicked if I could share an article I was reading with a housemate just with a swipe. –P2*

Augmenting Devices

Another observation that emerged in our study was that users often wished to use AR to add new or better functionalities to devices they already possessed. As household artefacts were considered well integrated into the home environment, AR presented the opportunity to enhance the device without interfering with its structure or simply needing to buy a new better one. For example, one user reflected how AR could enhance the experience of weighing oneself:

> *AR glasses could display a scanned image of the body and visualize how much body lean and fat you have and a compare how it was four weeks ago. –P3*

Device augmentation was not only to be performed at home. Participants also remarked that they would use it for daily shopping:

> *First, it (shopping list) is in the kitchen, then it is attached to the shopping cart, and I can tick items. I could freely walk through the store and glance at the list without picking up my phone all the time. –P6*

Another user expressed the wish for AR to enhance their perception of quantities in the kitchen. They believed that AR could increase their cooking repertoire and enhance cooking skills simply through providing extra information and reducing the need for new equipment:

> *Sometimes I have this problem: I want to bake something. And actually, I've no idea how to bake, because the amount of butter or some ingredients are always so exact and I don't have some volume measure - I only have some random bowl or plate* [...]. *Since AR can measure distance and areas, I guess maybe it could help me to measure ingredients. –P7*

We also observed that users requested functionalities that many commercially available household objects already offer. In the following passage, the user requested an AR-based indicator for pan temperature, while pans with built-in temperature gauges are now easily available.

> *I would like something to tell me that the pan is relatively hot at the moment and I can start cooking. –P6*

Concerns

Here, we note some negative reactions that the perspective of using AR every day produced in users. These primarily fell into two categories: privacy and information overload.

Privacy The users in our study understood that extensive sensing was required to offer them an AR experience. As a consequence, they were worried that future AR devices would constantly record their actions and thus pose a threat to privacy:

> *The glasses sense all information around me. Everything that the camera, microphone and other sensors can capture. –P9*

In contrast, AR was also perceived as a way to embed confidential information in the environment. Users envisioned that an AR system would control access to parts of the AR environment and only allow specified users to see parts of the AR world:

> *Wi-Fi passwords, something only house mates can see* [...] *the benefit is that I am the only one seeing the information and no one else. That means I am not disturbed by others while sitting in public transport. –P8*

On a larger scale, one user proposed to adorn the house for a social gathering. Here, only invited or paying guests could enjoy the visually enhanced location:

> *It would also be interesting if you want to do decorations. It's good for a party, and then everyone who's invited can see it. –P2*

Information Overload Another concern was receiving too much information. Some users were afraid that excessive information embedded in the surroundings would provide too much stimulus and overwhelm them:

> *I am concerned that I get too much information all the time. Information should be presented only on request. –P9*

Further, participants recognized that a possible future proliferation of AR would imply the need for finding new ways to manage and prioritize information. One user commented that they would require a systematic way to access information in AR:

> *If information comes to you all at once, it is a little bit too annoying, but if there's a way that it can be systematically organized, and it is prioritized in such a way that the most important one is at a particular point, this is something I'll definitely always wear. –P2*

3.2.4 Discussion

We explored the design space of AR at home with two studies and found that users reported a large variety of possible usage scenarios. Our results show that users generally welcomed AR as part of their everyday experience. Entire families participated in our probe and AR-based activities fostered interaction and discussion in the families.

AR systems as personal technologies

Firstly, our studies showed that users envisioned using AR for personal means, in personal spaces, which is in stark contrast to what the majority of past research efforts explored [47]. The initial survey showed that AR was perceived as useful all around the home and our interviews showed that users envisioned employing AR even in simple tasks. This suggests that there is space for new exploration for HCI, widening the domestic frontier of AR applications. While past efforts explored AR's affordances for rendering complex tasks simpler or aiding in coordinating, our work shows that users expect AR to be deployed in more simplistic tasks. Next, we summarise more detailed findings and discuss challenges and opportunities for future work on domestic AR.

Users see AR primarily as a means of providing assistance

We observed that users were highly interested in AR providing ASSISTANCE throughout the day. While we do recognize that this may have many positive effects, such as fostering good habits or skill development, ubiquitous assistance may pose some problems. The threat of providing too much help and essentially rendering everyday life boring is a known issue in Ubicomp literature [207]. Our study shows that this potential issue is very relevant in the case of domestic AR. Thus, future designs of AR for the home should prioritize engaging experiences to avoid rendering everyday life facile. Further, the survey and the probe show that AR was often seen as a smartphone replacement or even a substitute for other information artefacts (e.g. replacing one's paper shopping list). This implies that past opinions in the literature about the blurring boundary between mobile interaction and AR [68] are also perceived by everyday users. While these uses may seem attractive to the users in the photos they contributed, it remains a challenge for HCI to explore how effectively AR could replace well-established interaction modalities.

Enhanced perception was an expected benefit of AR

Participants in both studies were eager to enhance their sensory perception through AR. We see that this may have many benefits such as an increased awareness of environmental dangers or a better understanding of their natural surroundings. Yet, increased awareness may come at a cost. As we observed in the CONCERNS theme, users are aware of their limited cognitive capacity. As enhanced senses generate vast amounts of information, cognitive overload becomes a threat. Further, as our senses are biologically limited, channelling additional sensing input through AR may interfere with regular vision thus essentially limiting perception. While augmented perception offers interesting opportunities, designers must be wary as the consequences of amplified perception have not been fully explored [214]. The users' willingness to accept augmented perception systems at home offers an exciting opportunity for new designs, but it also calls for making sure that augmented sense systems are safe and reliable.

AR-enhanced household goods are likely to be domesticated

We observed that users were eager to augment everyday objects with functionalities provided by AR. Participants in the survey listed many objects and probe participants provided many examples in the AUGMENTING DEVICES theme. This implies that future household goods could take simpler forms as some controls may be replaced by AR. Augmented household goods can not only lead to increased aesthetics and reduced production costs but also enable more customization. These findings resonate with our past work [88], where users proposed to replace or enhance the functionalities of household goods. The fact that users find turning everyday objects into interactive artefacts implies that many existing techniques for augmentation, e.g. WorldKit by Xiao et al. [266], can be deployed at home. We observed users envision interactions similar to annexing real objects as proposed by past work [1]. Our study shows that the domestication of AR may have implications for the design of household goods. As a consequence, designers of future interactive artefacts for the home should consider enabling AR-based functionalities. Further research is required on how existing AR techniques can be effectively applied to existing domestic artefacts.

Privacy and transparency are critical factors for AR domestication

We noted that some users expressed concern about whether AR could be used in all social contexts or if it could produce information overload. This highlights the need for further work on context awareness for domestic AR. Further research is required to understand how to design privacy protocols for AR at home. Further, our findings show a need for developing AR interactions that respect existing social structures [223] and support SOCIAL ACTIVITY. As AR is already used in social settings, most prominently through the game Pokémon Go [196], future AR applications will need to effectively navigate social structures, especially if they are to be used at home. In the survey, users envisioned that AR may enter all rooms in their homes. This poses the challenge of designing AR technologies that support social coordination and acceptability. New social contacts about AR at home may be required, as suggested by past work [46]. Consequently, designing AR systems that respect the users' privacy and communicate openly how the users are protected is necessary for AR to enter the domestic application area.

Limitations

While our probe and survey were designed to be comprehensive and address a wide user group, the studies are prone to certain limitations. Firstly, we recognize that our work surveyed only a Western European population. The domestication of AR is certainly affected by cultural factors that should be studied in future work.

Further, we see that, for financial and legal reasons, we were not able to let families use the HoloLens or any other advanced AR device over a longer period of time. While we are confident that our Snapchat-like app elicited rich feedback from the users, deploying a fully functional AR device in the household may have offered more ecological validity. To the best of our knowledge, none of the currently available devices can offer a lightweight head mounted-display experience that would enable spontaneous experience sharing. Consequently, our design emphasized the serendipity of idea generation over staying true to the form factor. We recognize that a different focus may have yielded alternative results.

Finally, we recognize that visions of AR are highly present in mainstream media and thus produce a certain hype effect in users. In our analysis, we tried to focus on recurring patterns and themes to reduce the impact of the novelty effect.

However, we recognize that users' attitudes towards the domestication of AR will evolve over time while new technology arise.

3.3 Beyond Domestic Environments

In the previous sections, we presented an in-depth analysis of the users' attitudes and understanding of augmented reality in a domestic environment. Through the course of this thesis, we further developed and investigated research probes for different contexts and scenarios. We directed special attention to MR applications in learning scenarios. Since the impact and long-term effects of MR applications are unknown, there is a need to investigate them with precaution. In the following, we present a brief overview of the usage scenarios and discuss the opportunities and constraints for new learning and teaching methods through novel MR applications targeting teaching environments.

Usually, teaching methods include the presentation of different disciplines in front of several students. A proficient person transfers knowledge in this area, such as teachers or scholars. Since the overall education quality depends on the teacher-to-student ratio, information is diversely perceived by students. Larger classes require a generalization of content and increase the overall workload of teachers when it comes to supporting students individually [125]. The understanding of new topics for students is, therefore, be negatively affected [76].

Modern MR systems became an integral part of conveying knowledge within educational institutes to foster individual learning skills and alleviate the workload of supervisors [197, 147]. Such systems do not depend on external instructors and can be used at suitable times for students to learn new skills. Cognitive assistance is also provided, where visualizations are displayed in-situ to parallel occurring lecture sessions (cf. Chapter 5). For example, Billinghurst [25, 26] surveyed and demonstrated the use of low-cost technology to provide immersive MR experiences during different learning scenarios. However, presented information remains static and often does not offer user interaction, making it not much different from a regular textbook. While the vision of ubiquitous computing is still an ongoing research challenge [255], current state-of-the-art teaching modalities do not blend educational information with their environment. In contrast to information displayed on a stationary screen, MR technologies offer new possibilities to engage directly with interactive digital content presented in a real environment [17].

Of course, a student can also teach themselves autonomously by e.g., reading a book. Students are autonomous regarding the time and length of lessons as they are not dependent on another person, such as a teacher. Nevertheless, this limits the way information can be conveyed as they are not included in the real world, and it is not possible to show information depending on the context or being able to adapt the information on the level of knowledge and skill of the reader.

However, with the enhanced hardware and software solutions available, applications have been explored beyond the lab by various researchers. We recently showed that using MR glasses do visualize the current state of an experiment in physics class and foster a greater understanding of the teaching material [237]. But a comprehensive analysis of 87 MR learning applications, including an in-depth analysis of seven applications, reveals that MR applications have diverse effects on the students' performance [211]. The impact of deploying MR applications in educational settings should be discussed critically since the design of MR applications in education opens new design spaces.

Suitable content and elaborated knowledge sharing, as well as interaction concepts, need to be incorporated to transfer the capabilities of MR to educational establishments. In particular, since previsions work has shown that MR applications can quickly harm the overall learning performance if not carefully thought through.

3.3.1 Mixed Reality in Learning Environments

Based on the developed prototypes, conducted workshops, and studies during the project Be-Greifen (cf. Section 1.4.1), we highlight four themes where we consider the usage of MR applications in education as beneficial. Primarily, we think of the enormous potential of MR environments when deployed in universities to support the comprehension of complex scenarios in applied science and lab courses.

Improve Learning though Amplification

When making efforts to improve learning ability, human memory, and recall research proposes to spread out the learning process and information presentation [48]. In a recent lab study, we found indications that complex experiments could benefit from augmentation using MR glasses [237]. This trend motivates enriching further learning material to foster a better understanding of complex relationships. It is promising to apply these new technologies within practical lab courses. We suggest developing MR applications that enable learners to see and understand the fundamental facts. Specialized sensors can measure environmental data or the current status of an experiment. Voltage and current could be directly displayed within the wires during electrical engineering classes [22] or heat propagation in metals within physic classes [237]. MR displays, in combination with sensors, allows us to extend the human vision and visualize in-depth details of learning material in place of occurrence.

Personalized Learning

According to the National Educational Technology Plan [245] personalized learning allows each learner to learn at an optimized personal pace and the instructional approach. Current interactive learning applications follow this idea and optimize learning material to be meaningful and relevant to the learner. In MR learning experiences, this concept can be adapted and even developed further. While consuming difficult material or conducting advanced experiments, learners could receive tailored and immediate feedback through AR overlays. While learning or running MR supported experiments, the system can video record the environment, experiment, and virtual overlay. Difficult learning material can later be revisited or even be played back in VR. Further, learners could create personal notebooks of AR experiences recordings on the fly.

Extension

Many educational establishments are limited regarding their financial resource. These have a direct impact on the quality of teaching and education. Once deployed, MR systems can overcome this limitation and enhance learning. Besides, interactive learning and exploring environments can be created that extend the current body of learning material. We envision interactive experiments that were not possible to realize before because of time, financial, or security constraints. For example, chemistry students could safely explore chemical reactions with hazardous elements or biology students can examine samples under an augmented microscope that are usually not accessible.

Ubiquitous Learning

Learning can be considered as an ongoing, voluntary, and self-motivated pursuit of knowledge [41]. MR systems can support continuous learning by presenting chunks of knowledge spatially and temporally distributed. Research showed that ubiquitous learning could be more productive and engaging [116]. While this may apply for language learning, it is an open research question of how ubiquitous MR learning systems should be integrated into everyday life.

3.4 Conclusion

This chapter contributes toward the understanding of how people imagine MREs in their daily life and presents answers to the user-centered research questions **(RQ1-RQ3)**. We explored how MR can blend into our homes and how users can benefit from it in a domestic environment and beyond.

We conducted an online survey and a technology probe to explore potential MR domestication. The survey helped us identify initial opportunities for exploring the design space of MR at home. Later, we explored these opportunities in detail in the technology probe. We developed a tailor-made mobile application to enable users to suggest possible MR applications at home. From the photos and semi-structured interviews, we identified and discussed five themes of domestic MR: assistance, enhancing perception, social activity, device augmentation, and concerns.

Our results of the survey and the technology probe showed that users engage with MR experiences when excessive cognitive load can be avoided. While concerns about privacy and transparency exist, participants were eager to experiment with AR and saw the potential for long-term use. We highlighted that future system designers should carefully choose the degree of assistance provided to avoid cognitive overload when designing for augmented perception, explore the design space of MR-enhanced domestic equipment and derive new privacy and transparency rules for domestic MR.

Besides the focus on domestic environments, we discussed how MR systems could open new opportunities in the education sector. In particular, we highlighted pedagogical and challenges as well as the broad options of personalized, improved, and ubiquitous learning.

Chapter 4

Enhancing Reality

The perception of reality can be enhanced through limitless analog or digital artifacts. Spectacles can revive human vision, while night vision equipment can enhance visual perception during darkness. When it comes to information access, we no longer rely on books but smartphones, whether we are at home or on the go.

With emerging Mixed Reality Experience (MRE), it is possible to display context-aware information directly in the real world. In this chapter, we explore the effects of enhancing reality by blending context-aware information into the real world. Following the theme of *Assistance* from the previous chapter, we developed a quadcopter-mounted projector by combining a remote-controlled quadcopter with a wireless programmable projector enabling the projection of in-situ information in the real world.

With this probe, we explore how we can enhance the perception of reality through in-situ projection. To that end, we facilitate the research probe to investigate the effect of extended reality by projecting contextual information directly within the vivid environment without instrumenting the users, thus extending the perceived reality around them. As an application scenario, we evaluate visual navigation systems that usually constrain the users to shift their attention to the navigation system, e.g. a printed map or navigational system, and then transfer the navigation cues to the real world.

In the remainder of this chapter, we first review related works addressing the challenges of pedestrian navigation, and recent quadcopter supported concepts.

We then present our research probe for extending reality with an in-situ projection quadcopter capable of projecting ahead and below the quadcopter. In a lab study, we show that extending reality through in-situ instructions for navigation using a quadcopter leads to a significantly higher ability to observe and recall real-world points of interest. Finally, we discuss the effect of enhancing reality by blending information in the real world **(RQ4)**.

This chapter is based on the following publication:

- P. Knierim, S. Maurer, K. Wolf, and M. Funk. Quadcopter-Projected In-Situ Navigation Cues for Improved Location Awareness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, New York, NY, USA, 2018. Association for Computing Machinery

4.1 Related Work

To improve the sightseeing experience, it has been suggested to use visual landmarks or POIs as a navigation aid [169]. This has been evaluated, and has been found to improve the navigation experience [105]. Recently, Wakamiya et al. [250] suggested using this fact for generating memorable routes based on useful visual landmarks. However, using a smartphone while walking not only limits the sightseeing experience; it can also be dangerous. According to a study [186], the number of pedestrian accidents in the United States which involved a pedestrian talking or texting on a mobile phone while walking has increased over the last years. This is because even more pedestrians are immersed in using their phone and do not pay attention to their surroundings anymore.

In the following sections, we present previous work focusing on multimodal navigational systems and quadcopter supported navigational systems, which inspired the presented work.

4.1.1 Multimodal Navigation Systems

Although the traditional smartphone-based navigation is one of the most used, different navigation systems have been suggested. These navigation systems are not limited to visual navigation cues; in fact, over the last decades, a number of navigation systems have begun or have proposed using haptic, auditory, or visual navigation cues. Considering haptic navigation systems, Heuten et al. [104] use a belt that provides tactile navigation cues. Similarly, Pfeiffer et al. [198] use electronic muscle stimulation to remote control a pedestrian's route. In the area of auditory navigation, Baus et al. [20] are investigating auditory landmarks as navigation aids. Further, Lokki and Grohn [158] found that using auditory navigation in addition to visual cues is beneficial for navigation in virtual environments. Finally, regarding visual navigation cues, the state-of-the-art is traditional turn-by-turn navigation instructions, which are presented on the screen of a smartphone. These can be found on the pre-installed Maps applications in Android or iOS. Research also suggested using in-situ projection [200]. E.g. Winkler et al. [260] are presenting a body-worn system which can present in-situ navigation instructions. Their system is further capable of presenting public and private projected content. Moreover, for receiving projected navigation instructions while riding a bike, Dancu et al. [45] proposed mounting a projector and a smartphone on a bike. All presented systems require a shift of attention from the environment to the display. AR solutions overcome this limitation by presenting the information directly in the environment. Narzt et al. [185] describes a visualization paradigm for in-car AR navigation systems. Rehrl et al. [206] conducted a study comparing navigation performance using voice, digital map and AR cues and report that AR presented on a smartphone causes significantly worse navigation performance. AR solutions still lack comfort and require the user to carry a separate, often bulky device [206, 260].

4.1.2 Quadcopter as Display

We are observing a trend of quadcopters becoming more widely used in navigation-related applications. In a sports context, Müller & Muirhead [181] have proposed using a quadcopter as a jogging companion. They were using a fully autonomous flying quadcopter which can follow a previously defined route. For providing a clean environment, Obaid et al. [194] are using quadcopters to locate, then encourage people to clean up trash. Schneegass et al. [216] are suggesting to use free-floating displays mounted on a quadcopter to create temporary navigation signs to control crowd movements in emergency situations. Scheible, Funk, and Nozaki [212, 193] present two similar concepts using a projector and a canvas attached to a single quadcopter to provide information to people. Matrosov et al. [166] additionally added a depth camera to the down-facing projector to facilitate a tangible interaction projection surface. Moreover, Avila et al. [15] suggested using small nano quadcopters to navigate blind and visually impaired travelers using the sound which a quadcopter naturally emits. Recently, Kim et al. [127] proposed to use a quadcopter as a navigation aid to get home safely when walking alone in the dark and Colley et al. [42] explored the potential to communicate navigation cues just with the quadcopters' movements. In conclusion, quadcopters are currently used to control crowds in emergency situations [216], display information during sports [181, 216], and display tourist information [216]. However, we want to explore how to extend these scenarios by combining quadcopters with additional technology.

4.2 Quadcopter-Mounted Projector

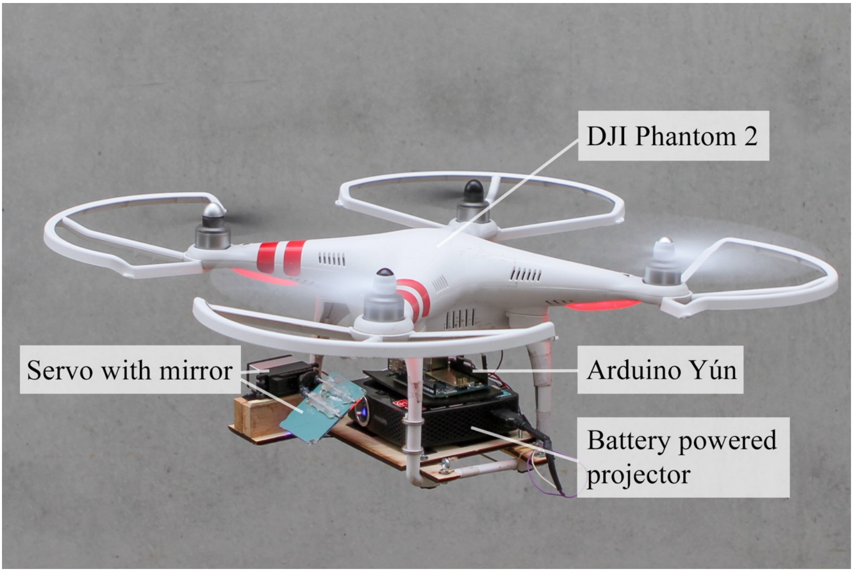


Figure 4.1: The research probe of the steerable levitating projector.

Our research probe is a quadcopter-mounted projector. The levitating projector can present wirelessly streamed content onto surfaces underneath or in front of the projector. Inspired by Wilson et al. [259], we use a steerable mirror to change the angle of projection and therefore the surface which is projected on. All components of the prototype are depicted in Figure 4.1.

The starting point of our levitating projector is a DJI Phantom 2 quadcopter. This quadcopter was chosen because of the additional take-off weight of 1300 g with a flight time of up to 25 minutes. Additionally, it offers advanced GPS-supported flight assistance systems. We use a Phillips PicoPix PPX3610/EU projector. It weighs only 284 g and is battery powered with a runtime of 90 minutes at 60 lumen. Further, the projector has built-in WiFi and is running Android 2.3.1. A small mirror was glued to a light servo and placed next to the projector to deflect the projectors light cone. The servo is connected to an Arduino Yún which includes a separate WiFi module. The Arduino Yún is USB powered by the projector and is running a server application receiving commands to adjust the angle of the mirror. Depending on this angle the projected image is in front of the quadcopter, directly underneath it or slightly behind it.

We developed two applications which we connected via WiFi. One runs on the projector and controls the projected navigation instructions. Based on the mirror position the projected image is mirrored horizontally to compensate for the effect of the mirror. The second application runs on a mobile device and sends the navigation instructions to the projector as well as the required angle to the Arduino Yún. During our study, this application was operated by a Wizard of Oz.

As a control condition, participants had to use simple graphical navigation instructions presented on a smartphone. In this condition, we showed the same visualization that was presented using the in-situ projection quadcopter. In fact, the smartphone was running the same application as the projector, displaying the navigation instructions controlled by the Wizard of Oz with the subtle difference that the content never got mirrored. To facilitate future replications of this work a detailed description of the assembled hardware and developed software is available online[9].

4.3 Method

To evaluate the effect of enhanced reality through in-situ navigation instructions, we conducted a user study. Using the previously introduced quadcopter-mounted projector allows for blending information in the real world. As a baseline, we choose the presentation of visual navigation cues presented on a smartphone.

We designed the study following a repeated measures design with the used instruction method (in-situ instructions by levitating projector and smartphone

[9] https://github.com/pknierim/QuadcopterProjectedNavigationCues

Figure 4.2: We conducted a user study comparing (left) projected in-situ navigation instructions to (right) traditional navigation instructions presented on a smartphone.

navigation) as the only independent variable. As dependent variables, we measured the Task Completion Time (TCT), the NASA-TLX [95], and the error distance between the actual location of the Point of Interests (POIs) and where the participants remembered them ($\text{DISTANCE}_{\text{POI}}$). We treat the ORDER of the routes that were used in the study as a between-subjects variable.

The apparatus used in the study was the two prototypes that were presented previously; i.e. we were using the *in-situ* projection quadcopter and the *smartphone* navigation as a control condition. We defined two different routes starting and ending at distinct points (e.g. at a crossing or a landmark). The routes had approximately the same length (510 m and 530 m). Both routes are depicted in Figure 4.3. The walking time at a slow paced walking speed is approximately 6 minutes for each of the routes. Each participant walked both routes once – one route with the quadcopter navigation system and one route with the smartphone navigation system. We counterbalanced the order of the navigation systems and the order of the routes in a way that each route and each navigation system was used equally during the study. To ensure a good visibility of the projected navigation information, the study was conducted at dusk (between 7 pm and 9 pm). We only conducted the study on days with good weather and clear sky.

As we are interested in analyzing the memorability of the surroundings, we introduced points of interest (POIs) located at buildings or crossings along the route. During each route we presented three POIs, which were visualized using a geometric shape (blue circle, red rectangle, and orange triangle). The geometric shapes were presented on the smartphone or by the levitating projector when the participant reached the position of the POI with an accuracy of 1 m (also depicted in Figure 4.3). The POIs in the first route were a distinctive set of stairs, a small alley, and a big sign. For the second route the POIs were another distinctive set of stairs, a noticeable balcony, and a container. All POIs were not directly on the path of the route; rather on the sides. When choosing the POIs for the study, we made sure that they were distinctive points along the way which could be used for memorizing the way afterwards.

4.3.1 Procedure

After explaining the course of the study, the participants were asked to sign a consent form, which informed the participant about potential risks of participating in the study and explained which data would be collected. We further instructed the participants to always be aware of their surroundings. Further, we instructed the participants to walk at a comfortable pace. As a first priority we defined following the navigation instructions, and as a second priority we instructed the participants to memorize the location of the POIs. After briefing the participants, we made them familiar with the first navigation system. Both the levitating projector and the smartphone navigation system were presented as fully automated systems. In the in-situ condition, two study assistants acted as a Wizard of Oz. The first wizard was operating the quadcopter at a distance of 5 m in front of the participant. The second was controlling the projected navigation instructions using a tablet computer. In the smartphone condition, only one wizard was needed to control the presented navigation instructions. Once the participants were familiar with the task and the navigation system, we started the study. The two wizards were walking behind the participants at a distance of approximately 2 m. After each route, we asked the participants to fill out an NASA-TLX questionnaire and to recall the position of the POIs by walking back along the route and telling the experimenter the position of the POIs.

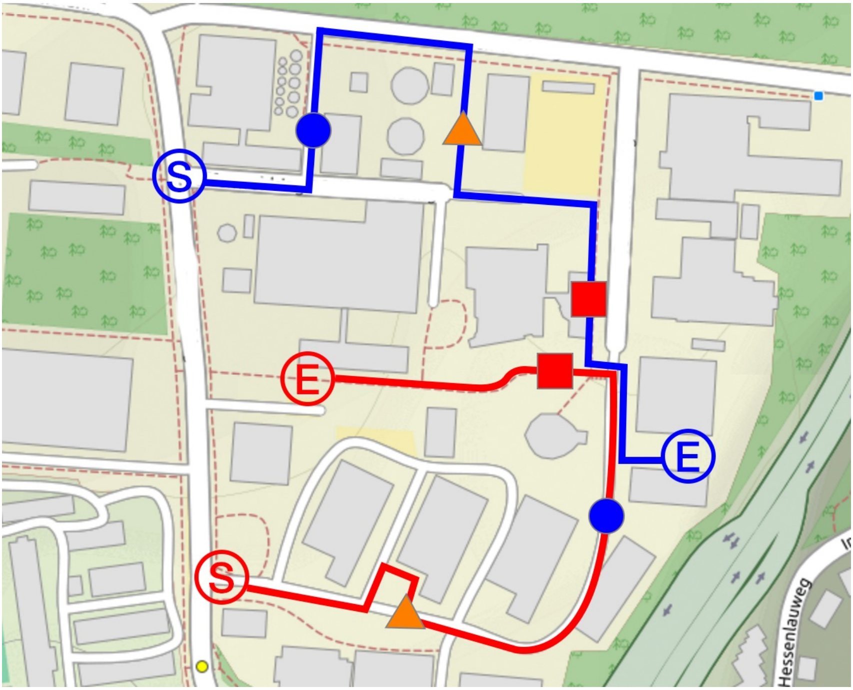


Figure 4.3: The routes that were walked by the participants in our user study. Both routes are approximately 500m long and consist of 8 turns. The S indicates the start of the route and the E indicates the end of the route. The location of the POIs are marked with the geometric shapes (circle, rectangle, and triangle) as they were used in the study.

The experimenter logged the position of the recalled POI using a smartphone application, which determines the recalled POI using GPS. Afterwards, the real POI GPS location is compared with the perceived POI GPS location to calculate the distance between perceived location and real location using the Haversine formula. Afterwards, we repeated the procedure with the other route for the second condition.

4.3.2 Participants

We recruited 16 participants (7 female, 9 male) via our university's mailing list and personal contacts of one of the authors. The participants were aged from 18 to 62 years (M = 33.7 years, SD = 13.9 years). They were mostly students with various majors or persons working in a variety of industry jobs. All participants owned a smartphone. The participants were rewarded with candies for participating in our study. As we conducted the study at a remote part of the campus, none of the participants were familiar with it.

4.4 Results

We statistically compared the TCT, the NASA-TLX, and the $\text{DISTANCE}_{\text{POI}}$ using a one-way repeated measures ANOVA with the order of the routes (ORDER) as a between-subjects variable. A graphical representation of the results can be seen in Figure 4.4.

First, we analyzed the average time the participants needed to walk the routes (TCT). The *smartphone* navigation condition ($M = 314.62$ s, $SD = 43.55$ s) was faster than the *in-situ* projection condition ($M = 368.69$ s, $SD = 77.88$ s). A one-way repeated measures ANOVA revealed a significant main effect between the approaches, $F(1,15) = 7.33$, $p = .017$. The effect size estimate shows a large effect ($\eta^2 = .301$). There was no interaction effect for TCT × ORDER.

Considering the RTLX, representing the perceived subjective workload the participants had while consuming navigation instructions, the *in-situ* projection condition led to a lower perceived workload ($M = 17.14$, $SD = 8.11$) compared to the *smartphone* condition ($M = 21.82$, $SD = 12.35$). A one-way repeated measures ANOVA could not reveal a significant difference between the two conditions ($p > 0.05$). Also there was no significant interaction effect for RTLX × ORDER.

Finally, when comparing the $\text{DISTANCE}_{\text{POI}}$, the *in-situ* condition led to a smaller error in the distance between the POIs and the recalled position of the POIs ($M = 15.16$ m, $SD = 6.33$ m) compared to the *smartphone* condition ($M = 30.25$ m, $SD = 22.46$ m). A one-way repeated measures ANOVA revealed a significant difference between the approaches, $F(1,13) = 6.384$, $p = .027$. The effect size estimate shows a large effect ($\eta^2 = .347$). There was no interaction effect for $\text{DISTANCE}_{\text{POI}}$ × ORDER of the routes. It has to be mentioned that two participants

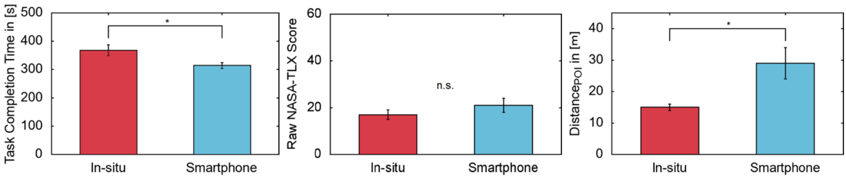


Figure 4.4: The results of the user study. (left) The Task Completion Time in seconds needed for each condition. (center) The perceived cognitive workload represented by a NASA-TLX score. (right) The error distance the participants made when trying to recall points of interest. All error bars depict the standard error. The * indicates a statistically significant difference.

were not able to remember one of the presented POIs. One could not remember a POI in the in-situ condition and another could not remember a POI in the smartphone condition.

After conducting the user study, we asked the participants to provide additional qualitative feedback through a semi-structured interview. Considering the quadcopter condition, three participants (P2, P3, P16) were *"a little scared of the quadcopter"* and therefore told us that *"[they] reduced [their] walking speed not to come too close to the quadcopter"*. Further, some participants felt that our prototype of the projector quadcopter is *"too loud for using it in an everyday setting"* (P6, P7, P12). When we asked them if they followed the quadcopter instead of the in-situ navigation instructions, P4 stated that the quadcopter was *"flying too high to just follow the quadcopter. Therefore, the projected navigation instructions were very useful"*. Similarly, considering the POIs, P1 stated that he *"like[s] that the drone is projecting the points of interest directly into the real world. This helps to not miss any of them."* (P1) In the smartphone condition, P1 had problems noticing the POIs as he was concentrating on his walking and not continuously looking at the screen of the smartphone.

4.5 Discussion

Considering the TCT, we found that the smartphone-based navigation led to a significantly faster TCT. Qualitative analysis revealed that this was mainly due to participants being careful about walking too close to the quadcopter. As we did not want to rush the participants, we instructed the two Wizards-of-Oz to always retain a fixed distance between the quadcopter and the participant. Defining a fixed speed for the quadcopter might have yielded different results considering the TCT. We could not find a significant effect between the two navigation systems considering the Raw NASA Taskload Index (RTLX) score. However, when comparing the $\text{DISTANCE}_{\text{POI}}$ between the two conditions, we found that the in-situ navigation instructions provided by the levitating quadcopter were leading to a significantly shorter $\text{DISTANCE}_{\text{POI}}$ compared to the smartphone-based navigation instructions. Additionally, during the smartphone control condition of our user study, we had to intervene twice verbally. Once because P7 almost ran into a wall and P13 respectively falling over a concrete barrier. Both participants were entirely focused on the phone in their hands, trying to identify POI that they could not perceive their surroundings anymore.

4.5.1 Limitations

Our implementation of the proposed quadcopter-mounted projector for visually extending reality comes with certain limitations.

During the study, some participants mentioned that the quadcopter we were using was loud. The participants thought that in an everyday scenario, this would distract passersby. However, while conducting the study, passersby were very interested in the prototype and the reality extending capabilities. This opens up exciting possibilities for investigating whether personal companion quadcopters could foster social interaction. Despite the in-situ navigation instructions, our projector quadcopter additionally presents two more navigation aids: the quadcopter itself and the sound that it emits naturally [15]. We asked the participants whether they followed the quadcopter or the instructions, and all except one stated that they were following the projected instructions. Further, our system was only tested at dusk and night. With the current prototype, projected navigation instructions are barely visible in daylight due to the low luminous intensity of the projector. However, we believe that using an advanced monochrome laser projector could solve this issue.

4.6 Conclusion

In this chapter, we investigated the effect of enhancing reality by blending information directly into the real world **(RQ4)**. The developed research-probe comprises a quadcopter-mounted projector for extending reality with projected information. As an application scenario, we chose pedestrian navigation in urban environments. Navigation instructions were displayed in-situ as an alternative to traditional smartphone navigation instructions. In a user study, we compared these in-situ navigation instructions to state-of-the-art smartphone navigation. The results reveal that although participants required considerably more time to complete a route using in-situ navigation instructions with a levitating projector, the participants could memorize points of interest significantly more accurately.

We conclude that extending reality through in-situ navigation instructions leads to a higher memorability and awareness of the surroundings. Beyond this thesis [170], we envision quadcopters as a versatile instrument and smart personal companion to enhance the perception of reality in sports or outdoor sightseeing scenarios.

Chapter 5

Perceiving Reality

The fundamental characteristic of MR systems is blending the virtual and physical worlds into a coherent environment. In addition to displaying context-aware information like navigation cues, MREs can also be implemented to present the human senses with supplementary real-time information that is usually not perceivable by the human sensory system. Outperforming sensors can be employed to overcome the limited and restricted human senses and either extend or amplify the perception of reality. Mixed reality applications offer a potential solution to visualize more detailed information or abstract quantities that are not directly perceivable by humans.

The previous chapter's results indicated that spatially anchored visual cues help to better retrieve and remember information. In the next step, we further investigate the effect of enhanced perception of reality by blending real-time information into the environment. Therefore, we choose a physics lab experiment as an application scenario, and examine the effect of substituting the physical components of the environment with tangible or virtual replicas.

In this chapter, we present the development of an MR research probe enabling students to observe the heat flux through a metallic rod. A thermal camera observes the experiment and provides real-time information about the heat gradient and physical effects taking place. A virtual graph floats above the setup, augmenting it and visualizing the real-time temperature values. Based on this physics lab experiment, we developed two additional abstraction levels by substituting the physical pieces of the experiment with non-functional replicas or entirely virtual representations, allowing us to study the quality knowledge transfer.

This chapter is based on the following publications:

- P. Knierim, F. Kiss, M. Rauh, and A. Schmidt. Tangibility is overrated: Comparing learning experiences of physical setups and their virtual equivalent in augmented reality. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia (MUM)*, 2020
- P. Knierim, F. Kiss, and A. Schmidt. Look Inside: Understanding Thermal Flux Through Augmented Reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 170–171, Oct. 2018

5.1 Related Work

In the following sections, we will give a brief overview of the concept of augmentation in simulation-based learning environments.

AR learning environments provide a unique set of features and affordances that are often adopted from other domains as ubiquitous and mobile computing [265]. New learning opportunities include individual learning pace to alleviate the overall workload of supervisors and students [147]. Such AR systems offer new possibilities to manipulate and engage directly with the interactive content presented in the real environment [17] and do not rely on external supervision.

Simulations enable further abstraction from the real learning material. Students that used computer-supported experiences and learned with AR performed better on conceptual questions and developed a greater facility at manipulating a real object. Computers further promote student learning and skill development in reasoning and manipulating if well designed [64].

Prior this work, we already pointed out, that teaching is improved using AR to display invisible properties [238]. With the same augmented thermodynamics experiment [143] we evaluated the effects of real-time AR overlays and highlight the positive effect of AR in physic lab courses [237]. These results are supported by research that shows how students benefit from AR when learning about circuits [22]. In contrast, Radu et al. explored how more abstract material, such as fractions and numbers, can be taught to elementary school students [202].

In traditional teaching scenarios, teachers usually educate using a whiteboard, writing, and handheld teaching aids. These teacher-based methods allow a structured and guided learning experience. However, students may lack the chance to develop learning autonomy. DiSessa [49] argues that computers can be fundamental for new literacy that can change how students think. Further, he critically discusses how new immersive technologies can be integrated into learning to keep them exciting and intellectually generative. Without a clear focus, learning material may not foster knowledge. In other studies [55], researchers highlight the unique strength of AR to foster engaging and interactive well situated and collaborative problem solving. However, new technological, managerial, and cognitive challenges arise.

Learning also includes collaboration and dialogue with peers, but integrating these characteristics into an AR environment can be particularly challenging. Having a heterogeneous group with non-immersed collaborators adds an additional level of complexity [112]. Researchers have found that concepts like face-to-face communication [26], shared spaces and objects, as well as new forms of user interaction, can enhance collaboration, and several AR and VR experiences have been developed to investigate how to collaboratively solve tasks as a group [180].

There is still an open discussion to what extent AR and VR environments can support learning. An extensive analysis of 87 research articles showed a small adverse effect to a significant effect [210]. AR is emerging in the field of education and literature already identified affordances and limitations related to teaching, learning. Dunleavy identified three instructive design principals namely challenge, fantasy, and curiosity [55]. With this work, we focus on curiosity and make the unseens visible. We further investigate this area of research and focus on the effects of substitution of functional items in real physical experiments.

5.2 Realizing the Thermal Experiment

We tailored our augmented reality application to teach heat conduction in metals for an introductory laboratory course in thermodynamics. Previously, students had been required to take snapshots with a handheld thermal camera to acquire data and do an offline analysis. With our application, students get real-time feedback and enhanced data visualization of the experiment and can observe the thermal flux. Fist results of a user study using the prototype showed a small positive effect of augmented reality on students' performance with regards to acquired knowledge in thermodynamics [238].

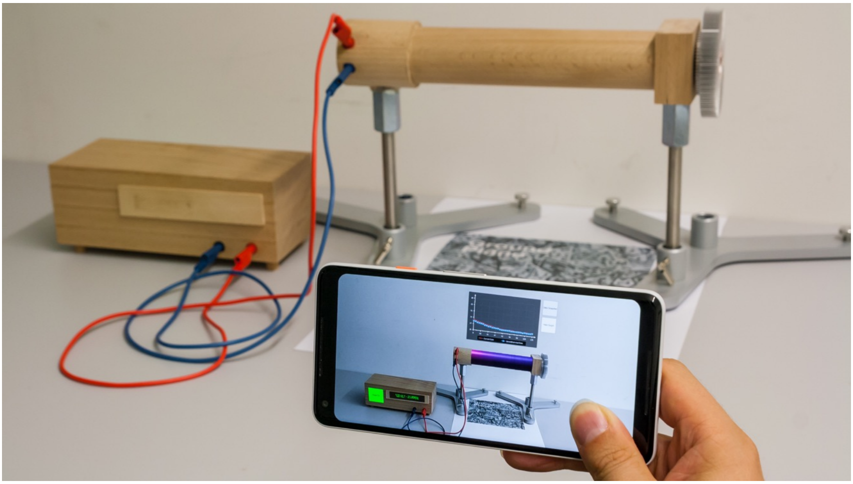


Figure 5.1: Thermal flux lab experiment with components substituted by non-functional replicas. Augmentations and simulation on smartphone enable functionality and interactivity of the experiment.

System Architecture Our system comprises the thermal experiment itself, an infrared camera attached to a server and an augmented reality display acting as a client. Our simple server-client architecture supports multiple users to enable collaborative experiment execution. An overview of all components is depicted in Figure 5.2.

The thermal experiment itself consists of several metallic rods made of aluminum, copper or brass. A power adapter supplying 12 volts is used to control the temperature of one end of the metallic rod. Further, there are insulated rods to generate different thermal flux properties.

The thermal camera is centered in front of the metallic rod to capture real-time temperature values. These are forwarded to the server for further processing. The infrared imager is connected via USB to a computer running a server application and image processing pipeline. The captured infrared video feed is analyzed and based on the temperature signature the metallic rod is registered within the data. The sampled temperature data is recorded and sent wirelessly via a simple communication protocol to the client on request.

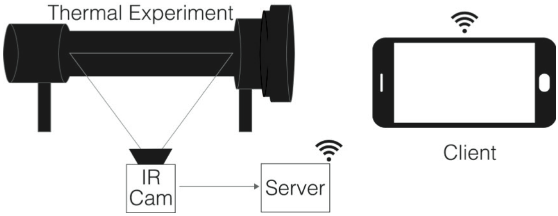


Figure 5.2: Architecture of the research probe comprising a server that streams real-time experiment data to the HoloLens client.

AR Application The augmented reality display is in charge of rendering the false-color representation on-top of the metallic rod as well as plotting the floating graph above the experiment setup representing the real-time data. To enable correct augmentation of the experiment, the AR display needs to register the setup in space. For our research-probe, we specifically utilize a smartphone that supports advanced AR capabilities. To register the experiment in space, we use the Vuforia framework[10], the back facing camera of the smartphone and printed a marker to identify the experiment's location in space. Finally, the augmented reality display gives in-situ hints and additional information to guide the students through the experiment.

5.3 Method

To investigate the effect of the different levels of virtual abstraction of the experiment, we conducted a user study. The Independent Variable (IV) APPARATUS consisted of three levels of abstraction: *real*, *replica* and *virtual*. Since none of the participants should conduct the physical experiment twice APPARATUS was used as the between-subject variable.

[10] https://library.vuforia.com/

Figure 5.3: Entities for experimental assembly for all three versions. left) *real*: Power supply, metal probe, thermal camera, wires, smartphone and stopwatch. center) *replica*: wooden power supply replica, wooden probe replica, wires, smartphone and stopwatch. right) *virtual*: smartphone and stopwatch.

5.3.1 Conditions

Each participant was invited to conduct the thermal experiment in one of the three conditions described below. An overview of all three set-ups, including the AR overlays, is illustrated in Figure 5.4.

Condition 1: Real Setup

The *real* setup condition comprised the original thermal flux experiment that is currently conducted by students in the Physics lab. Students could observe the experiment through the augmented reality display and get live data captured from the thermal camera. In the *real* setup condition, participants were asked to execute the full experimental set up which includes probe set up, camera calibration, wiring and operating the power supply. All necessary hardware components are depicted in Figure 5.3. The setup including AR overlay is depicted in Figure 5.4.

Condition 2: Replica Setup

In the second condition, we replaced the original brass probe with a wooden replica of the same size and shape. Thus, the replica could be mounted on the

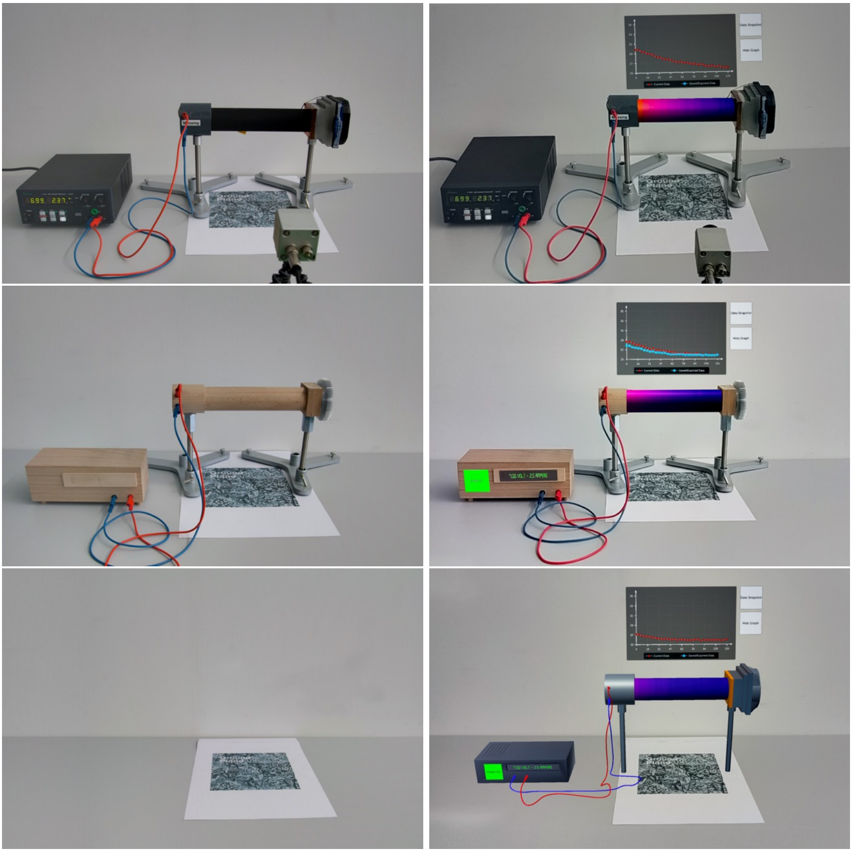


Figure 5.4: Three experiment condition setups from top to bottom: *real*, *replica*, and *virtual*. Each without (left) and with (right) augmentation of the experimental area. Augmented images are screen captures recorded during the experiment.

same tripods. The power supply was also replaced by a wooden replica including wooden elements indicating a display and plugs. Both the probe and the power supply replica had color-coded drilled holes to allow to plug in regular laboratory-style banana plug cables. The smartphone application was extended by the ability to augment the power supply and visualize a virtual power button and display to present the current values for electric potential and electric current. In contrast to the *real* setup, in this condition participants were not required to set up and calibrate the camera. Instead, previously recorded data was streamed from the server and visualized respectively to the status of the power supply.

Condition 3: Virtual Setup

In the *virtual* setup condition, no physical components except of the smartphone and stopwatch were involved. The experimental probe, thermal camera, power supply, and wires were replaced by their virtual counterparts and rendered within the AR display. Again, the power supply had a virtual power switch to turn on the supply and start the experiment. Thermal data was provided from the previously recorded data set and was overlaid in exactly the same way as in the other conditions.

5.3.2 Apparatus

Our server application ran on a RasperyPi 3. This application provides a user interface and live stream for the thermal camera for calibration for the *real* condition as well as data streaming of recorded or real-time thermal data. In the *real* condition, we utilized an Optris PiConnect 160 infrared camera with an optical resolution of 160 x 120 pixels at 120 Hz and a spectral range of 7.5 to 13 μm. We used the Google Pixel XL with 64 GB as AR display showing our previously outlined overlays.

5.3.3 Task

In this study the participants had to conduct a simple thermal flux experiment that is widely used in laboratory classes. They were asked to set up the thermal flux experiment according to the printed manual. Depending on the APPARATUS this task involved different steps. For the *real* or *replica* condition, the metallic rod sample, camera, and power supply or the wooden replicas needed to be placed in the experiment area; and the power supply (real or wooden) needed to be connected to the heating and cooling element. For the *real* setup, the thermal camera needed to be aligned and calibrated. Since the *replica* and *virtual* condition do not rely on live data, the camera was not involved in this setups. For the *virtual* setup, only the smartphone was used and no other components had to be set up. After successfully setting up the experiment, participants had to switch the power supply and start the stopwatch. During the heating process, the participants' task was to record the minimal and maximal temperature at intervals of 2 minutes for 10 minutes.

5.3.4 Procedure

After welcoming the participants, we asked them to sign the consent form and take a seat next to the dedicated experimental setup area. We gave a brief introduction into the thermal flux experiment and explained the study. We assigned each participant randomly to one of the conditions and showed them to the appropriate table next to the setup area that contained all the components. The different components for each condition are shown in Figure 5.3. We explained how to start and use the AR application and handed the participants the assembly instruction and task description. They set up and conducted their experiment, observing the thermal flux and recording the temperature values. Throughout the study, we gave advice on request and manually logged time and errors made. After finishing the experiment and answering the knowledge questions, the participants filled out the RTLX [95] and System Usability Scale (SUS) questionnaire [31]. In the last step, we collected demographic data and the participants filled in the compensation form. Including debriefing, the participants completed the study in 25 to 40 minutes.

5.3.5 Participants

We recruited 30 participants (8 female, 22 male) aged from 18 to 55 years ($M = 28,45y, SD = 7,77$) via our university mailing list and social media. All were undergraduate students with mostly technical background. However, non of the participants had additional physics knowledge extending secondary school knowledge. All participants had normal or corrected to normal vision, and 23 of them had previously experienced AR. Participants received a small gratuity and either course credits for the computer science lecture or 5 EUR as compensation for their participation.

5.3.6 Measures

In this user study, we measured the setup time, perceived task load, system usability, and the quality of acquired thermal values. We measured the setup time

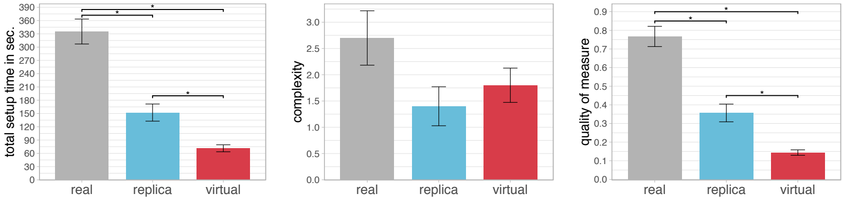


Figure 5.5: Mean values for set up time in seconds (left), complexity as sum of errors and given assistance (center) and quality of measures (right) for each condition. Error bars show standard error of the mean (SE). Asterisk indicate statistically significant differences between conditions.

starting with handing over the printed manual to the participant until switching physically or virtually the power supply to heat up the probe. We assessed SUS [31] and RTLX [95] through questionnaires presented in a browser. We evaluated the understanding of the experiment itself through free text questions which we ranked by quality. Questions were adapted from the Heat and Temperature Conceptual Evaluation (HTCE) [247] questionnaire and included the identification of the temperature distribution curve out of six possible graphs and a non-fit option as well as questions targeted on temporal and spatial properties of the probe. We also manually recorded the necessary help or errors during the setup or execution of the experiment.

5.4 Results

To analyze our data, we conducted multiple one-way independent-measure ANOVA with the between-subjects variable APPARATUS. The homogeneity of variances was tested using Levene's test. All significance levels are at $\alpha = .05$. The results are visualized in Figure 5.5.

5.4.1 Task Load Index (RAW NASA-TLX)

To assess the users' perceived task load while conducting the thermal flux experiment, we used the TLX score of the NASA-TLX questionnaire. All recorded NASA-TLX scores are very similar with *real* ($M = 29.80, SD = 14.28$), *replica*

($M = 29.80, SD = 12.73$) and *virtual* condition ($M = 30.20, SD = 15.53$). Hence, a one-way independent-measures ANOVA could not reveal a significant effect of APPARATUS on one of the three conditions ($p > 0.05$).

5.4.2 System Usability Scale

Considering the SUS, representing the subjective usability of the system, the *real* setup condition led to a higher subjective usability ($M = 89.25, SD = 6.877$) compared to the *replica* ($M = 86.75, SD = 6.877$) and *virtual* ($M = 84.00, SD = 8.991$) condition. A one-way independent-measures ANOVA could not reveal a significant difference between the three conditions ($p > 0.05$).

5.4.3 Setup time

Setting up and calibrating all components of a physical experiment consumes time. After handing out the experiment instructions to the participant, we recorded the time (in sec.) it took them to prepare and initiate the thermal flux experiment (see Figure 5.5 left). The *virtual* setup condition led to the lowest setup time ($M = 71.40 sec, SD = 24.99$) followed by the *replica* ($M = 152.10 sec, SD = 61.25$) and *real* ($M = 335.20 sec, SD = 89.10$) A one-way independent-measures ANOVA revealed a significant difference between the conditions $F(2,27) = 44.51, p < .001, \eta^2 = .767$.

Post hoc analysis was performed using Bonferroni corrected pairwise t-tests to determine statistically significant differences between all conditions. Post hoc comparisons of the average set up time until experiment start revealed significant differences between all three conditions. (all with $p < .05$). Further, Cohen's effect size value (all $d > 1.7$) suggested a large practical significance.

5.4.4 Quality of aquired Thermal Values

To determine the quality of the acquired thermal values we calculated the absolute difference in Celsius from of measurement to the average as quality of measure (*qom*) as follows:

$$qom = \left| M_{min(t)}(i) - \sum_{n=0}^{N} \frac{M_{min(t)}(n)}{N} \right| \tag{5.1}$$

Here, M is the minimum or maximum measure at time t (interval of 2 minues) for participant i. n is the the amount of samples per condition ($n = 10$). Given the formula, a smaller *qom* indicates less variation and fluctuation in the measurement. To determine if there was significant difference between the absolute difference of the three conditions, we performed a one-way independent measure ANOVA (see Figure 5.5 right). Results show a significant main effect of APPARATUS on the quality of measurement ($F(2,291) = 55.29, p < .001, \eta^2 = .275$)

Bonferroni corrected pairwise comparisons revealed a significant difference between the *real* ($M = .767, SD = .570$) and the *replica* ($M = .357, SD = .447$), ($p < .001$) condition, *real* and *virtual* ($M = .144, SD = .143$), ($p < .001$) as well as *replica* and *virtual* ($p = .003$).

5.4.5 Knowledge Transfer & Complexity

Questions were ranked by the laboratory assistant from wrong (1), neutral (2) to correct (3) allowing a total score of nine points. *Replica* led to the best results ($M = 5.6, SD = 1.71$) followed by *virtual* ($M = 5.5, SD = 2.21$) and *real* ($M = 4.6, SD = 2.63$). A one-way independent-measures ANOVA revealed no significant difference between the conditions (all $p > .05$). We summed up all errors and required assistance during the implementation of the experiment as a measurement of complexity (see Figure 5.5 center). In the *replica* condition, participants required least assistance ($M = 1.4, SD = 1.17$) followed by *virtual* ($M = 1.8, SD = 1.03$) and *real* ($M = 2.7, SD = 1.63$). Though statistical analysis of the results revealed no significant difference between the conditions, participants in particular struggled during calibration and setting up the thermal camera and required assistance on the AR application start.

5.5 Discussion

The proposed system enables students to perceive physical phenomena in a novel and more relatable way. Our results show that the setup time using the abstracted more virtual versions of the experiment was significantly reduced compared to the original lab experiment. Spending less time on setup and calibration allows investing more time in complex tasks like knowledge transfer or understanding the experiments and underlying concepts. It remains an open question if and to what extent experiment preparation contributes to the knowledge fostering process.

However, several works evaluating augmented learning environments show that students perform significantly better when relying on AR and simulations [22, 64, 81].

As we found no effect between the conditions for Task Load Index (TLX) and SUS, we cannot confirm that reducing functionality by substituting real objects with their virtual equivalent reduces workload or increases usability on a large scale although the data indicate lower complexity for the abstracted experimental setups. Therefore, we expect higher possibilities to run lab experiments at home without external guidance correctly.

Our results imply a significantly higher quality of measures for the more virtual experiments. We are confident that higher errors are a combination of camera orientation, calibration, distraction and error in measurement. For an adequate offline analysis of the experiment, exact measurements are crucial to derive correct conclusions. Hence, data acquisition based on simulated data could reduce the frustration of students caused by noisy data recordings.

Setting the invested time, quality of measurement, knowledge transfer, and complexity in contrast, our results suggest that it is possible to run augmented experiments of abstract concepts with reduced complexity. We recommend lower complexity, in particular, if no technical assistant is available to assist students.

For future systems that support the implementation of an augmented experiment to foster learning, our findings recommend running the real experiment once and repeatedly running slightly adapted versions with nonfunctional augmented replicas. We assume students will further benefit from the opportunity to retake an experiment based on simulated or recorded data.

5.5.1 Limitations

Participants were mostly students with a technical background; however, we did not invite students from the field of Physics to the user-study. Consequently, the findings may not fully apply to all teaching scenarios. Nevertheless, considering more heterogeneous groups as usually such as those we used our findings are very applicable.

Measuring knowledge transfer is challenging, and we only scratched the surface to investigate overall knowledge gain. In particular, we did not investigate the long-term effects of substitution and augmentation. We believe that a long-term study will provide more insights into the effects that augmentation and substitution can produce on learning.

We argue that substituting experiments with carefully considered mixed reality environments can outperform real-world experiments. Students can particularly benefit from simulations, enhanced perception of reality, and a more extensive variety of experiments since virtual adaptations of the experiments in size, shape, material, or the like will be feasible.

5.6 Conclusion

In this chapter, we further investigated the effect of enhanced perception of reality through MREs in the context of education. We based our research-probe on an existing physics lab course experiment since many physical concepts are based on abstract non-visible quantities. Our probe enabled students to observe the thermal flux in situ. Moreover, we studied the effect of substituting the real pieces with non-functional replicas and virtual representations. The results of the conducted user study indicate improved setup times and increased quality of measurements. Nevertheless, we could not show improved fostering of knowledge or comprehension using any of the simplified lab experiments.

Based on our findings, we argue for combining real experiments enriched by variations using MR enhanced experiments. Additionally, the use of entirely virtual experiments could strengthen existing knowledge by intrinsically motivated repetitions.

Chapter 6

Enhancing Virtuality

Present HMDs deliver rich and high immersion due to the latest technology advances. They deliberately limit the connection to the real world to create a high level of immersion and strong sensation of presence. Unfortunately, visual immersion not only substitutes real world distractions but also limits the connection to the real world and considerably decreases the possibility of interacting within the real environment. To overcome this limitation, we investigate the feasibility and effect of enhancing VR environments by blending physical objects into the virtual space.

We introduce typewriting in VR as an application scenario for our investigation. To enable users to input text or work as efficiently in a virtual environment as in a real office, they require high performance input devices. This is especially true for users who are not fluent touch typists, where text input quickly becomes tiresome if they cannot see their hands or the keyboard [173]. VR systems with efficient keyboard based text input can offer great potential to create pleasant working or study environments. Commuters in trains and cars [174], or employees working from home could wear an HMD to sit virtually in their familiar working environment or attend business meetings far away. External visual and auditory distractions can be blocked entirely, which would aid productive and focused work. Furthermore, VR allows the creation of entirely new environments with vast three-dimensional display space in any direction. VR user interfaces are no longer bound to rectangular two-dimensional displays limited by the size of our desks.

To that end, different solutions have been proposed for text input while immersed in VR. They embrace point and click solutions with tracked controllers, handwriting with a pen on a tablet, and speech. Others overlay the virtual environment with a cropped video stream of the real world. However, none of these solutions can facilitate high-performance text input known from real world typing. Of course, the user can take off the HMD every time a text input needs to be made, but this quickly becomes inconvenient and destroys the immersion.

In this chapter, we pave the way for efficient work due to new techniques for generic text input in VR. We first review related solutions enabling text input in VR and the effects of different avatar representations on presence. We then present our developed research probe that enables efficient typing in VR. Our prototype visually represents the user's hands and the physical keyboard of a desktop workspace in VR. The keyboard and fingers are tracked and visualized in real time to support the user visually to interact with the peripheral. In the accompanied user-study, we evaluate typing speed and accuracy in contrast to real life typing. Based on these results, we further developed a portable implementation for text input in VR to research the unique challenges and opportunities when blending real and virtual in a portable configuration.

This chapter is based on the following publications:

- P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–9, New York, NY, USA, 2018. Association for Computing Machinery[1]
- P. Knierim, T. Kosch, J. Groschopp, and A. Schmidt. Opportunities and Challenges of Text Input in Portable Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, New York, NY, USA, 2020. Association for Computing Machinery

[1] Parts of this paper are also included in the PhD thesis of Valentin Schwind.

6.1 Related Work

The first Virtual Reality HMD system was created in 1968 by computer scientist Ivan Sutherland [241]. With new technology advances companies like HTC, Oculus and Sony recently catapulted virtual reality into the living rooms. However, the potential which lies within education and business and beyond gaming is largely under-explored. We review text input solutions tailored for VR and works exploring the effect of avatar hands on VR environments.

6.1.1 Text Input

Current game-controllers or gesture interfaces are suitable for micro inputs and games. Unfortunately, they fail to support high bandwidth generic input. Researchers, developers, and stakeholders have proposed a wide variety of different text input solutions for VR. A comparison of early works using pen, keyboard and gesture-based techniques showed that the achieved text input rates of less than 14 Words per Minute (WPM) stay far behind real world typing speeds on a physical keyboard [30]. Gonzales et al. [78] confirmed these results analyzing a different set of input devices developed for text input in VR.

All commercially available controllers enclosed with the HMD of HTC, Oculus and Sony support text input on a virtual keyboard. Users point a virtual ray with the controller onto a character and confirm the selection with a button press. Alternatively, a built-in touch pad can be used to move the pointer around. Similarly, R. Kim and J. Kim [128] use the touch screen with hovering capabilities of a smartphone for selecting characters on a virtual keyboard. With their input technique they achieve up to 9.2 WPM.

The Microsoft HoloLens supports text input in augmented reality applications through a holographic keyboard and a pointer which is controlled using head rotation. Yu et al. [268] also studied head-based text entry for VR and combined the concept with gesture-word recognition whereby experienced users perform up to 25 WPM. The VR system FaceTouch [87] leverages a touch-sensitive front cover of the HMD and the sense of proprioception to enable text input with up to ten WPM on a virtual keyboard.

None of these approaches can keep up high input speed and usability known from typing on a physical keyboard. Recently, researchers focused on augmenting VR by incorporating a video stream of reality into the virtual environment to compensate typing performance decrease [173]. Lin et al. [152] extend this approach by utilizing a depth camera to display a point cloud of a user's hands beside a rendered virtual representation of the physical keyboard. To compensate for the increased error rate introduced while typing in VR, Walker et al. propose to use decoders known from text entry on touchscreens to correct errors [251, 252].

Overall, it remains an open challenge how to build a VR system that supports accurate and fast text input that can compete with typing on a regular desktop setup.

6.1.2 Avatar Hands

The effectiveness of virtual environments has been linked to the subjective experience of being and acting at one place while physically situated at another [261]. New sensors can easily determine the hands pose and position to render them in VR accordingly. Displaying them increases the immersion and presence and further enables natural user interaction within the virtual environment [16]. Schwind et al. [222] investigated the effect of different hand renderings on presence. Results highlight the importance of users' diversity when designing virtual reality experiences.

We take the current body of related work and investigate how hand representation regarding model, texture, and transparency affect typing performance, workload and felt presence. We restrict the physical environment to a seated setup while the user feeds text into the system via a physical keyboard.

6.2 Realizing Typing in Virtual Reality

For any physical keyboard based text input users execute, they need to localize and reach out to the keyboard in a first step. Localizing could either happen visually or haptically using the surface features of the keyboard. VR HMDs prevent the user from visually localizing any physical peripherals. A system realizing effortless typing in virtual realities should support the user with an easy to understand representation of the keyboard's location in relation to their fingers. According to Feit et al. [61], non-touch typists' gaze switches up to 1.2 times between the display and the keyboard within a sentence. They spend up to 41% of their attention looking at the keyboard. Hence, an accurate representation of the keyboard and hands seems necessary particularly for this group of typists.

Figure 6.1: Side by side illustration of the real environment (left) and the virtual reality replica (right).

6.3 Implementation

To investigate the different aspects of typing in a virtual environment, we implemented our VR apparatus using an Oculus Rift CV 1. The Oculus camera tracks the headsets position. We incorporated a motion tracking system comprising eight OptiTrack 13W cameras and the Motive 1.10 motion capture software for very accurate finger and keyboard tracking. Twenty-three 4 mm retroreflective markers are affixed to anatomical landmarks of each hand ensure precise tracking of each joint and bone of the hand. During the application startup, markers are seamlessly analyzed and automatically mapped to the virtual skeleton. In case of losing track of a marker during typing due to occlusion, our software automatically reassigns it, when it reappears, to untracked joints following a nearest neighbor approach. The layout of the markers is depicted in Figure 6.2.

A second generation Apple wireless keyboard is used for text input. Four retroreflective markers are attached to the top of the keyboard to enable repositioning of the keyboard during runtime to allow comfortable typing. The precise and interactive virtual replica of the keyboard is rendered according to physical position and keypresses in the virtual environment.

Our apparatus uses the OptiTrack NetNat SDK for streaming position data of bones, joints, and keyboard in real time. Our application and the virtual environment are implemented using the Unity game engine 5.4.0.

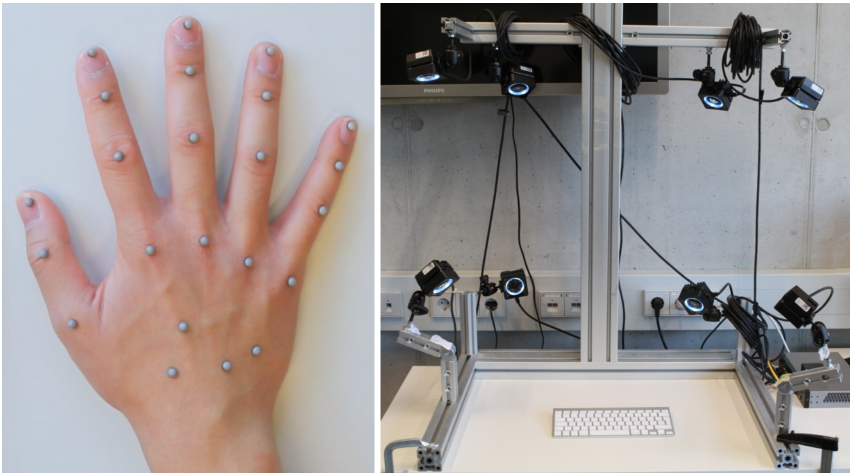


Figure 6.2: Hand with 23 retroreflective markers (left) and the hardware setup for finger and keyboard tracking (right).

6.4 Method

Our apparatus enables users to see the virtual representation of a physical keyboard and their own hands. The goal of this study is to evaluate the effect of virtual hand representation and hand transparency on typing performance of experienced and inexperienced typists in VR. Further, we investigate the overall typing experience by measuring the task load and sense of presence. We used a mixed nested factorial design with the nested within-subject variable HAND and TRANSPARENCY and the between-subject variable TYPING EXPERIENCE. For HAND we had three different levels. All hands were presented with 0% and 50% TRANSPARENCY. In addition, we use 100% TRANSPARENCY resulting in no hand visualization and the real world scenario. An overview of all eight conditions is shown in Figure 6.3. Typing performance was measured while

Figure 6.3: Pictures of the eight hand visualizations used in the study. Realistic, abstract, fingertips with no transparency and real hands (1st row) as well as 50% transparency and no hands (2nd row).

participants typed outside of VR on the real world apparatus, or inside of VR seeing different hands with varying transparency levels.

6.4.1 Participants

In a first step, we asked 80 (5 female) participants to conduct a simple online typing test.[11] Based on their results ($M = 53.3$ WPM, $SD = 18.8$), we invited a random sample of 16 participants with more and 16 participants with less than 53.3 WPM to shape groups of inexperienced and experienced typists. The 32 participants (three female) were aged from 18 to 27 ($M = 21.9$, $SD = 2.3$). Thirteen participants had previous experience with VR. Fourteen of them were wearing corrective lenses during the study. Participants received a small gratuity and either 10 EUR or course credits as compensation for their participation.

6.4.2 Apparatus

The apparatus for this study comprised two individual setups. One facilitated the real world typing task, the other allowed users to type on a physical keyboard while immersed in VR.

Real World Apparatus

The real world setup served as a baseline and consisted of a sixth generation 27 inch Apple iMac with Intel Core i5 and a second generation Apple wireless keyboard. The computer was running a full-screen typing application showing one stimulus after another at the display. It was developed in Unity game engine 5.4.0.

Virtual Reality Apparatus

For the virtual reality setup, we used our developed VR apparatus. We designed an alike looking virtual environment representation of our laboratory including the real world study apparatus comprising the iMac. The real world apparatus next to the virtual replica is shown in Figure 6.1.

[11] `https://10fastfingers.com`

Our experiment was running on a Windows PC with an Intel i7-6700, 16GB RAM, and a Nvidia GTX980. The target frame rate was set to 90 Frames per Second (FPS) to match the refresh rate of the Oculus Rift CV 1. Of course, there is a latency between a user's finger movement and photons hitting the user's retina. We used the provided performance toolboxes to monitor the latency. The summed up calculated latency caused by motion tracking, rendering pipeline, and HMD never exceeded 30 ms during the study.

6.4.3 Task

In this study participants had to accomplish a simple text input task on a physical keyboard. Participants were asked to place their hands left and right next to the keyboard to mimic aperiodic typing. Being in this resting pose, a 3-second countdown, displayed on the (virtual) iMac, started. After it elapsed, a random phrase from the MacKenzie and Soukoreff [164] phrase set was displayed. Participants were asked to enter the phrase as accurately and fast as possible. Phrases were presented at the top of the (virtual) display while participants' input was shown underneath. Participants were allowed to correct errors but also to confirm inaccurate or incomplete phrases. Pressing the enter key confirmed the input and the next phrase was displayed. For each condition, participants performed three sets of ten phrases. In between each set participants had to place their hands in the resting position again and wait for the countdown to elapse. The task was the same for all conditions inside and outside of the VR.

6.4.4 Procedure

After welcoming the participants, we asked them to sign the consent form and take a seat next to the apparatus. While attaching the 23 self-adhesive markers to each hand, we explained all devices and the course of the study to the participants. Afterward, the participant placed his hands within the tracking volume, and we defined the four markers at the dorsum of the hand as rigid bodies. In the last preparation step, we adjusted the HMD to the participant's head and calibrated it to the participant's inter pupil distance for best visual results. Then participants started with the typing task. After each task (three sets of 10 phrases), they had to fill out the NASA-TLX [95] and presence questionnaire (PQ) [261]. Subsequently, they repeated the procedure using the next hand representation. The first set of ten phrases at the start of each condition was a practice set to familiarize the participant with the different appearances. We did not include this set in our analysis. For the baseline outside of VIRTUAL REALITY, participants had to take off the HMD and move to the real setup to continue with the text input task. HANDS and VIRTUAL REALITY were presented in a counterbalanced order using a full latin square to prevent sequence effects. Throughout the study, we logged every keystroke including the timestamp for offline analyses. After all eight iterations, we asked for comments about their experience, typing performance, and which hand representation they finally preferred. Including debriefing and detaching the self-adhesive markers, participants completed the study in 70 to 110 minutes.

6.5 Results

We conducted multiple four-way repeated measure analyses of variance (RM-ANOVA) with the within-subjects variables VIRTUAL REALITY, HAND, TRANSPARENCY, and the between-subjects variable TYPING EXPERIENCE. As previously mentioned, the within-subjects factor HAND is a nested factor of the VIRTUAL REALITY condition. TRANSPARENCY is nested into HANDS, which means that conditions of a nested factor cannot be compared with levels of factors above (*e.g.*, there is no transparency in the *Real World* condition). All significance levels are at $\alpha = .05$.

6.5.1 Objective Measures

One participant was removed from the analysis of the objective measures due to missing correct inputs (error rate: 100%) in multiple conditions. Hence, we invited one more participant from the same group of typists to compensate for the deficit. In total participants wrote 7680 phrases and we analyzed 5120 phrases since the first ten phrases of each condition were assigned for training. The results of the objective measures are shown in Figure 6.4. The mean values of all metrics are listed in the Appendix.

Words per Minute (WPM)

The average typing performance is calculated in WPM where one word is defined to be five characters long [231]. Based on the logged keystrokes, we divided the

	VIRTUAL REALITY														REAL WORLD	
HAND	**No Hand**		**Finger Tips**				**Abstract**				**Realistic**					
TRANSPARENCY	**100%**		**0%**		**50%**		**0%**		**50%**		**0%**		**50%**		**0%**	
INEXPERIENCED TYPIST	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
WPM	31.848	11.338	36.189	8.669	37.648	1.576	38.809	7.307	39.808	6.919	37.581	5.611	38.430	5.839	45.398	4.909
Error Rate (in %)	.740	.538	1.244	1.528	1.304	1.127	.973	.707	1.121	.842	1.140	1.041	1.001	1.097	.713	.635
Corrected Error Rate (in %)	14.015	7.549	9.486	4.660	7.683	4.521	7.726	3.422	7.749	3.480	7.681	5.031	7.712	4.313	4.904	3.388
1st correct Keypress (in s)	4.386	2.813	2.200	1.236	1.971	.632	1.986	.793	2.129	.842	2.456	1.290	1.864	.528	1.769	1.054
EXPERIENCED TYPIST	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
WPM	61.830	17.047	68.547	15.810	37.648	1.576	67.018	16.134	69.172	16.370	66.566	16.569	67.165	16.589	67.223	14.837
Error Rate (in %)	.540	.505	.757	.922	.846	.974	1.003	1.135	.687	.613	.449	.415	.362	.316	.597	.528
Corrected Error Rate (in %)	7.383	6.116	4.354	1.858	4.766	2.297	5.118	2.307	4.889	2.909	5.034	3.361	5.025	2.578	5.467	2.899
1st correct Keypress (in s)	3.638	2.026	2.108	1.332	1.831	.670	1.791	.568	1.821	.561	1.953	.754	1.980	.692	1.370	.423

Table 6.1: Means and Standard Deviations (SD) of the typing performance indices of inexperienced and experienced typists: words per minute (WPM), error rate, corrected characters per phrase, and the time for the 1st correct keypress.

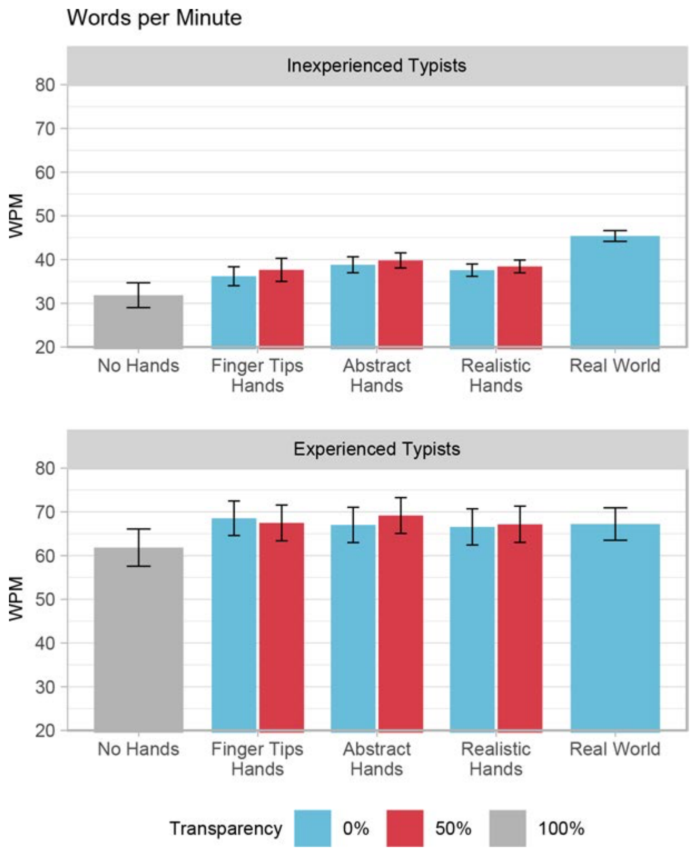


Figure 6.4: Mean values of words per minute for each condition. Error bars show standard error of the mean (SE). Exact values are also listed in the Appendix.

length of the final input by the time the participant took to enter the phrase. We measured the time from the first to the confirm keypress to calculate the WPM.

We found a significant effect of VIRTUAL REALITY, $F(1,30) = 22.97, p < .001$, and an interaction effect of VIRTUAL REALITY × TYPING EXPERIENCE, $F(1,30) = 22.97, p < .001$. Furthermore, we found a significant effect of HAND, $F(3,90) = 8.336, p < .001$, but no interaction effect of HAND × TYPING EXPERIENCE, $F(3,90) = .439, p < .726$. And we found no significant effects of

TRANSPARENCY, $F(3,90) = 1.596, p = .196$, and no interaction of TRANSPARENCY $\times$ TYPING EXPERIENCE, $F(3,90) = 1.022, p = .387$.

Post-hoc analysis was performed using Bonferroni corrected pairwise t-tests to determine statistically significant differences between the conditions. Due to the significant effects of TYPING EXPERIENCE, we compared the measures between *experienced* and *inexperienced* users separately. Due to no statistically significant effects of TRANSPARENCY, the data were aggregated across the transparency levels. For *inexperienced* users and the HAND factor we found significant differences between *No Hands* and the *Real World* condition ($p < .001$), *No Hands* and *Abstract Hands* ($p = .024$), between *Finger Tips* and *Real World* ($p = .006$), between *Abstract Hands* and the *Real World* condition ($p < .001$), and between *Real World* and the *Realistic Hands* ($p = .041$). No significant differences were found by comparing the other hand pairs (all with $p > .05$). Furthermore, we found no significant differences between the hand conditions only considering *experienced* typing users in VR (all with $p = 1$).

We summarize that the rendering of hands in VR has a significant effect on the typing performance measured using the WPM for *inexperienced* users in VR. The actual appearance of hands had no significant effect on the WPM measure of *experienced* users in typing.

Error Rate

One measure as an indicator of the users' typing performance alongside the WPM is the number of errors in the transcribed string. The *Error Rate* is given by the minimum string distance (*MSD*) between the transcribed string (T) and the presented phrase (P). The *Error Rate* in percent is: $ErrorRate = \frac{MSD(P,T)}{max(|P|,|T|} \times 100$. It captures the minimum number of insertions, deletions, or substitutions we have to perform to change one phrase into another [231].

We found a significant effect of VIRTUAL REALITY, $F(1,30) = 6.463, p = .016$, but no interaction effect of VIRTUAL REALITY $\times$ TYPING EXPERIENCE, $F(1,30) = 3.086, p = .089$ on the correction measure. There was no significant effect of HAND, $F(3,90) = 2.389, p < .073$ and no significant interaction of HAND $\times$ TYPING EXPERIENCE, $F(3,90) = 1.034, p = .381$. Both TRANSPARENCY, $F(3,90) = .158, p = .924$, as well as the interaction of TRANSPARENCY $\times$ TYPING EXPERIENCE, $F(3,90) = .337, p = .799$, were not significant. Pairwise post-hoc comparisons of the corrections showed no differences between the conditions of experienced and inexperienced users (all with $p > .05$)

Corrections

Neither WPM nor *Error Rate* captures the number of corrections and edits made during text input. The *Corrected Error Rate* [231] represents the effort put into correcting errors. We calculated the *Corrected Error Rate* by offline analysis of the keystroke log file. Therefore, we analyzed the log file and sought characters appearing in the keystroke log file, but not in the final transcribed text.

We found a significant effect of VIRTUAL REALITY, $F(1,30) = 14.4, p < .001$, and an interaction effect of VIRTUAL REALITY $\times$ TYPING EXPERIENCE, $F(1,30) = 18.4, p < .001$ on the corrected error rate. There was a significant effect of HAND, $F(3,90) = 9.933, p < .001$, however, not interaction of HAND $\times$ TYPING EXPERIENCE, $F(3,90) = 2.03, p = .115$. Both TRANSPARENCY, $F(3,90) = 1.006, p = .393$, as well as the interaction of TRANSPARENCY $\times$ TYPING EXPERIENCE, $F(3,90) = 2.527, p = .062$, were not significant.

Pairwise post-hoc comparisons of the ratio between corrected and overall inputs considering *inexperienced* users in typing showed significant differences between all hands and the *No Hands* condition (all with $p < .05$). Further pairwise comparisons considering other pairs and pairwise comparisons of experienced typists were not significant (all with $p > .05$).

Response Time Until the 1st Correct Keypress

For several applications, the time to react on a specific event using keyboard input is a critical measure of typing performance. After the expiration of the countdown, we recorded the time (in *s*) a user needed for the first correct keyboard input.

VIRTUAL REALITY had a significant effect on the reaction time, $F(1,30) = 22.85, p < .001$, however, there was no interaction of VIRTUAL REALITY $\times$ TYPING EXPERIENCE, $F(1,30) = .19, p = .666$. We found a significant effect of HAND, $F(3,90) = 17.947, p < .001$, however, not on HAND $\times$ TYPING EXPERIENCE, $F(3,90) = .374, p = .772$. There were no effects of TRANSPARENCY, $F(3,87) = 1.324, p = .271$, or TRANSPARENCY $\times$ TYPING EXPERIENCE, $F(3,90) = .872, p = .459$).

Pairwise post-hoc comparisons of the average response times until the first correct keyboard input revealed significant differences between all hands and the *No Hands* condition for inexperienced as well as experienced users in typing (all with $p < .001$). Other pairwise comparisons of the reaction time measure were not significant (all with $p > .05$). Thus, particularly to have *No Hands* in VR affected the initial response time for the first keyboard event negatively for both *inexperienced* and *experienced* users in typing.

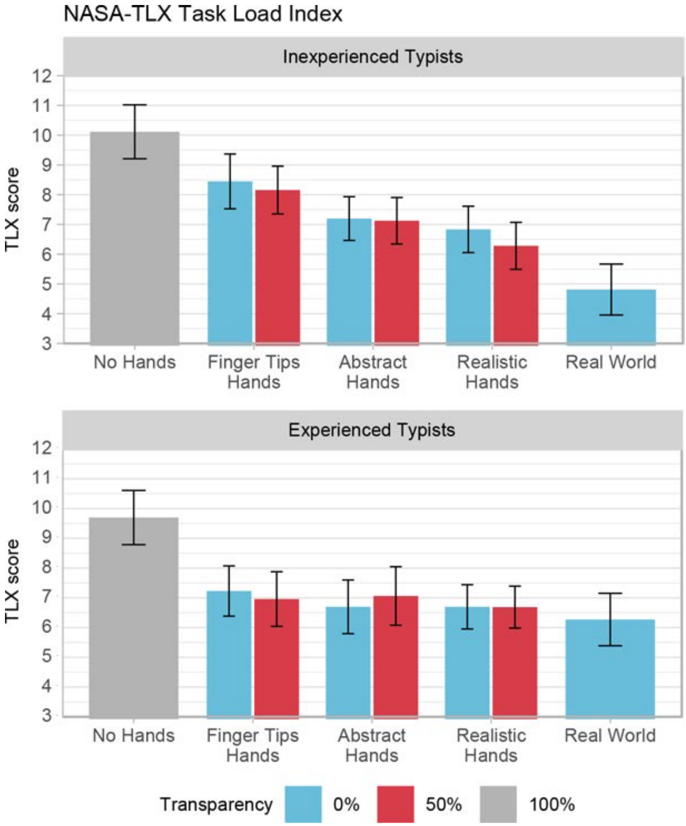


Figure 6.5: Subjective assessments of task load. Error bars show standard error of the mean (SE).

6.5.2 Subjective Measures

Further analyses were conducted to assess how the participants subjectively perceived the virtual hands. We asked for perceived work load and presence. All measures are shown in Figure 6.5 and 6.6.

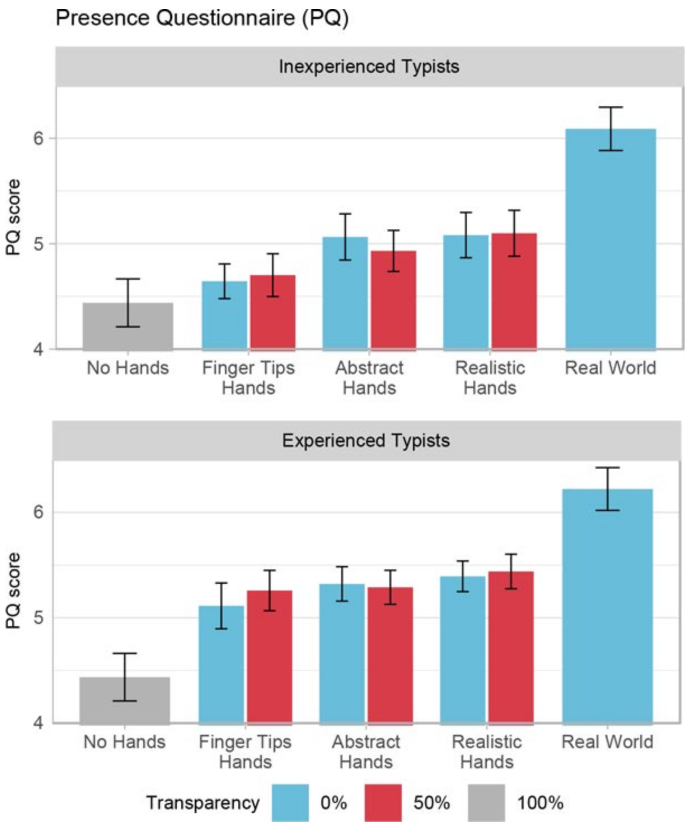


Figure 6.6: Subjective assessments of the sense of presence. Error bars show standard error of the mean (SE).

Task Load Index (NASA-TLX)

To assess the users' perceived task load of each hand we used the TLX score of the NASA-TLX questionnaire. We found significant main effects of VIRTUAL REALITY, $F(1,30) = 17.514, p < .001$, and HAND, $F(3,90) = 13.735, p < .001$, but no effect of TRANSPARENCY, $F(3,90) = 0.676, p = .569$. There were no interaction effects and none of the TLX measures was significantly affected by TYPING EXPERIENCE (all with $p > .05$).

Pairwise post-hoc comparisons of typing accuracy between the conditions considering the aggregated TLX measures across TRANSPARENCY and TYPING EXPERIENCE show statistically significant differences between *No Hands* ($M = 9.906, SD = 3.583$) and *Finger Tips Hands* ($M = 7.698, SD = 3.463$, with $p = .025$), between *No Hands* and *Abstract Hands* ($M = 7.021, SD = 3.348$, with $p < .001$), between *No Hands* and *Real Hands* ($M = 5.542, SD = 3.500$, with $p < .001$), between *No Hands* and *Realistic Hands* ($M = 6.625, SD = 2.954$, with $p < .001$), and between *Finger Tips Hands* and *Real Hands* ($p = .030$).

We summarize that having *No Hands* caused a significantly higher workload than the other conditions for both *experienced* as well as *inexperienced* users in typing. The lowest TLX score within the conditions of VIRTUAL REALITY was achieved by using *Realistic Hands*.

Presence

The Presence Questionnaire (PQ) was primarily designed to compare experiences in VR [261]. For the sake of completeness and to avoid potential biases, we asked for presence in the *Real World* condition as well. The overall score was averaged. Subscales are not considered in the following analysis. We found a significant effect of VIRTUAL REALITY, $F(1,30) = 99.62, p < .001$, and HAND, $F(3,90) = 13.269, p < .001$, but no effect of TRANSPARENCY, $F(3,90) = .549, p = .650$. There were no interaction effects and none of the PQ measures was significantly affected by TYPING EXPERIENCE (all with $p > .05$).

Pairwise post-hoc comparisons of the measures between the conditions considering the aggregated PQ scores across TRANSPARENCY and TYPING EXPERIENCE show statistically significant differences between *No Hands* ($M = 4.438, SD = .890$) and *Abstract Hands* ($M = 5.151, SD = .744$, with $p = .002$), between *No Hands* and *Real Hands* ($M = 6.155, SD = .806$, with $p = .001$), between *No Hands* and *Finger Tips Hands* ($M = 4.929, SD = .806$, with $p < .010$), between *No Hands* and *Realistic Hands* ($M = 5.253, SD = .753$, with $p < .001$), between *Finger Tips Hands* ($M = 9.906, SD = 3.583$) and *Real Hands* ($M = 7.698, SD = 3.463$, with $p < .001$), between *Abstract* and *Real Hands* (with $p < .001$), and between *Real World* and *Realistic Hands* ($p < .001$). Other pairwise comparisons (*Finger Tips* and *Abstract Hands*, *Finger Tips* and *Realistic Hands*, *Abstract* and *Realistic Hands*) were not significant (all with $p = 1.000$).

We summarize that the perceived presence was significantly affected by VIRTUAL REALITY and HANDS. The highest presence score was achieved using *Realistic Hands* while *No Hands* and *Finger Tips Hands* received the lowest presence scores.

6.6 Discussion

Our results show that the typing performance of mainly inexperienced users using a physical keyboard in VR was significantly decreased compared to real world text input. This is confirmed by several works evaluating typing in VR [152, 173, 252]. Experienced typists' text input performances were not significantly affected by missing hands or the different hand visualizations. However, rendering virtual avatar hands significantly increases the typing performance, response time, and typing accuracy of inexperienced users. Renderings of each virtual hand pair brought their typing performance back to a level that did not significantly differ from measurements in the real world.

Our results neither confirm an effect of appearance nor of transparency. Related to the degree of realism or human likeness of a virtual avatar, previous work suggests an effect of the Uncanny Valley. As we found no effects between abstract and very realistic hands, we cannot confirm an effect of the Uncanny Valley on the typing performance in VR.

Since the mental workload increases while typing in the virtual world, we assume that users are rather focused on the typing task than on the appearance of their hands. This finding is supported by two studies by Schwind et al. [220, 222] which reported that participants were highly focused while performing a typing task using virtual hands and non-physical keyboards in VR. In the present study, we confirm these observations even using a physical keyboard in VR.

Our results show that the workload is statistically higher for all typists when no hands are visible. However, experienced typists' workload is not affected by typing in VR as long as hands are rendered. This leads to the assumption that hand rendering has less impact on typing performance since experienced typists do not rely as much on the visual cues. Further, Realistic Hands caused the lowest workload for all, while maintaining the highest presence scores for typing in VR. Abstract or the absence of hands causes lower presences and a higher workload. We assume that the possible negative effect of latency, tracking errors as well as limited headset resolution and field of view contribute to the increased workload for inexperienced typists since they rely on seeing the own hands while typing [61]. Video see-through solutions [173] could minimize some of this factors like tracking errors or latency, though at the expense of full control over the hand and keyboard rendering as well as higher levels of immersion.

Setting typing performance, workload, and measured presence into contrast, our results suggest a correlation in particular for inexperienced typists, who seem to struggle more with abstract hand representations. We assume they need more visual guidance and abstract hands look less familiar to them. For future systems that enable typing in VR, our findings imply rendering realistic looking hands for best typing performance as well as high presence.

6.6.1 Lessons Learned

To achieve precise tracking and visual accuracy, our research probe relays on a high-quality motion capturing system. Hence, our setup is not mobile and self-adhesive retroreflective markers need to be attached to each hand. Large occlusion of markers or palm up-facing hand poses cause the tracking to fail. We evaluated the Leap Motion[12], a small sensor specific for hand tracking, to build a mobile version of our apparatus. Positional tracking is almost accurate enough, however, cannot match the precision of a professional motion capturing system. In the next section, we present our approach to overcome this limitation with current of-the-shelf hardware and present the development of a truly mobile setup.

6.7 Beyond Stationary Setups

As already outline, the fundamental requirement for realizing effortless typing on a physical keyboard in mixed reality is to enable the user to localize and reach out for the keyboard and understand the keyboard's location in relation to their fingers.

To investigate the effect of enhanced virtuality in a low-fidelity portable mixed reality environment, we implemented an additional research-probe using a Google Pixel 2 XL as the main component. We incorporated the smartphone with the Google Daydream VR viewer to create an HMD. To enable 6 DOF tracking, and to capture the environment, the smartphone's inertial measurement unit and camera are used. Since the heat sink of the VR viewer blocks the camera, we drilled a notch into it.

A wireless physical keyboard is used for text input. A printed visual marker is attached above the keyboard to enable visual tracking of the keyboard. Following the approach of Feiner et al. [59], we use the smartphone's camera during runtime

[12] https://www.leapmotion.com/

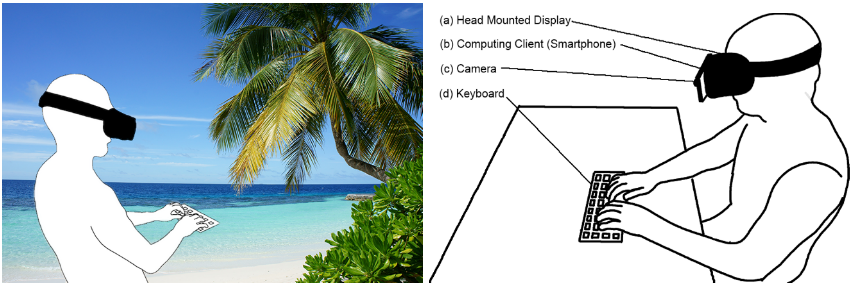


Figure 6.7: Left: Our vision of a user copy editing text in a relaxing virtual world provided by a portable HMD setup. Right: Our mixed reality apparatus for text input comprises a Google Daydream HMD [a], a Google Pixel 2 XL [b], and a wireless keyboard [d].

to create a cropped video texture of the keyboard, which is dynamically anchored to the physical position of the keyboard within the virtual environment. Similar to the previously presented stationary research-probe, this one allows for enhancing virtuality by blending the video texture of reality. In contrast, this setup does not require any instrumentation of the environment or user, yet it is fully portable. All components of our apparatus are shown in Figure 6.7. The virtual environment, including the cropped and arranged video, is demonstrated in Figure 6.8.

6.7.1 Method

Our mobile apparatus enables users to visually perceive the physical keyboard and their own hands while being immersed in a virtual environment. The objective of the following study is to evaluate the text input and editing performance using a mobile low-fidelity setup in contrast to today's smartphone input. We investigate the overall user experience by assessing system usability scale [31], NASA-TLX [95], and AttrakDiff [96]. We used a 3×1 factorial design with the within-subject variable SETUP. We employed three different levels for SETUP: *Mixed Reality*, *Smartphone + Keyboard*, and *Smartphone*. Both conditions that include the keyboard are shown in Figure 6.9. The typing performance was measured while employing a physical keyboard using the MR apparatus, the smartphone display, or direct typing using the smartphone soft keyboard.

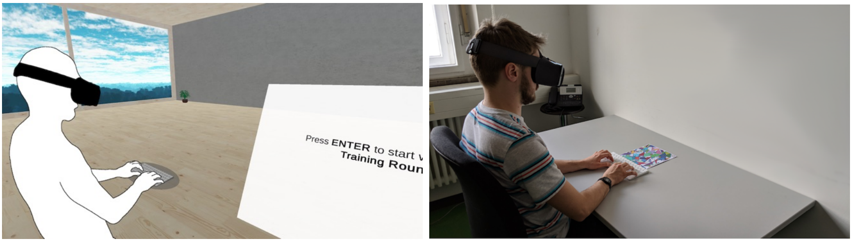


Figure 6.8: Left: The mixed reality environment with keyboard video texture and large floating display. Right: Untethered user typing with our mixed reality apparatus.

Subjects

In total, we recruited 24 participants via social media and our university's mailing-list to participate in our user study. The participants (six female) were aged from 19 to 38 (M = 27, SD = 4.66). Five participants were wearing corrective lenses during the study. Participants received either 5 EUR or course credits as compensation for their participation.

Apparatus

The apparatus for this study comprised thee individual setups sharing the same three, but individual combination of components: smartphone, keyboard, and MR HMD. The latter was only facilitated for the *Mixed Reality* condition.

Smartphone The smartphone setup served as a baseline and consisted only of a Google Pixel 2 XL running Android Pie. The smartphone was running our application in portrait mode showing the stimulus and text edit field at the top of the screen and below the google stock soft keyboard.

Smartphone + Keyboard For the second setup, we facilitated an Apple Magic Keyboard, which pairs wireless with the smartphone. This time the smartphone is placed above the keyboard in landscape mode serving as a portable display showing only the stimulus.

Mixed Reality For the mixed reality setup, we used our developed MR apparatus comprising the modified Google Daydream View 2, smartphone, and keyboard. We designed a virtual environment showing a room with a large screen displaying the stimulus. The cropped video of the physical keyboard and hands is displayed within the virtual environment at the corresponding physical location.

All smartphone applications were developed with the Unity game engine 2018.3. For head and keyboard tracking, we employed the Vuforia Engine 7.5.

Figure 6.9: Participants typing text during the user study. Left: *Mixed Reality* condition; Right: *Smartphone + Keyboard* condition.

Task

In the user study, participants had to accomplish two simple tasks. First, a simple text input task and second a copy editing task requiring to remove spelling error and adding or removing words. Participants started in a resting position with their hands placed next to the keyboard or smartphone. While being in this pose, a 3-second countdown elapsed on the smartphone or virtual display, indicating the start of either the text input or copy editing task.

Text Input For the text input task, a random sentence from the MacKenzie and Soukoreff [164] phrase set was displayed. Participants were asked to enter the phrase as fast and accurately as possible. Participants could correct errors during input but were also allowed to confirm inaccurate or incomplete phrases. With the enter key, participants confirmed the input, and the next phrase was displayed. For each condition, participants performed three sets of ten phrases. The task was the same for all conditions.

Copy Editing For the copy editing task, the participants had to review and correct three different texts. Each text consisted of 12 modified sentences from the MacKenzie and Soukoreff [164] phrase set. The required corrections were indicated between the lines highlighted in green. Participants were asked to edit as fast as possible all corrections. Except for the Mixed Reality condition, the edit cursor could be placed by touching the screen or with the arrow keys of the keyboard. We compensate for potential complexity differences by counterbalancing the prepared texts across all conditions.

Procedure

After welcoming the participants, we asked them to sign the consent form and explained the apparatus as well as the course of the study. Afterward, we asked participants to put on the HMD to adjust it to the head for the best visual results. Before starting with the typing task, participants were asked to get familiar with the virtual environment and get used to the tracking and visualization of the keyboard. After finishing both tasks (input and copy edit), participants had to fill out the RAW NASA-TLX [95], the AttrakDiff, and the System Usability Scale (SUS) [31] questionnaire. This procedure was subsequently repeated for all conditions. The first set of ten phrases at the start of each condition was a practice set to familiarize the participant with the apparatus. We did not include this set in our analysis. SETUP was presented in a counterbalanced order using a full Latin square to prevent sequence effects. After finishing the third iteration, we conducted a short semi-structured interview and asked for comments about their performance, user experience, and personal preference. Including the debriefing, participants completed the study between 60 to 90 minutes.

6.7.2 Results

We conducted multiple one-way repeated measure analyses of variance (RM-ANOVA) in order to reveal statistically significant effects of the within-subjects variables SETUP. All significance levels are set to $\alpha = .05$.

Words Per Minute (WPM)

For the text input task, participants entered a total of 2160 sentences. Since we discarded the first ten sentences of each participant, only 1440 sentences were used for analyses. We used the logged keystrokes to calculate the WPM by dividing the length of the final input by the time required to input the presented phrase [231]. The calculated WPM provides a measure for the average typing performance. We found a significant effect of SETUP on the typing speed, $F(1.23, 28.28) = 31.22$, $p < .001$. Furthermore, post hoc tests revealed a significant difference between the conditions *Smartphone + Keyboard* and *Smartphone* (M = 17.97, SE = 2.18, with $p < .001$), between *Smartphone + Keyboard* and *Mixed Reality* (M = 10.14, SE = 1.37, with $p<.001$) and between *Smartphone* and *Mixed Reality* (M = -7.82, SE = 2.99, with $p = .046$).

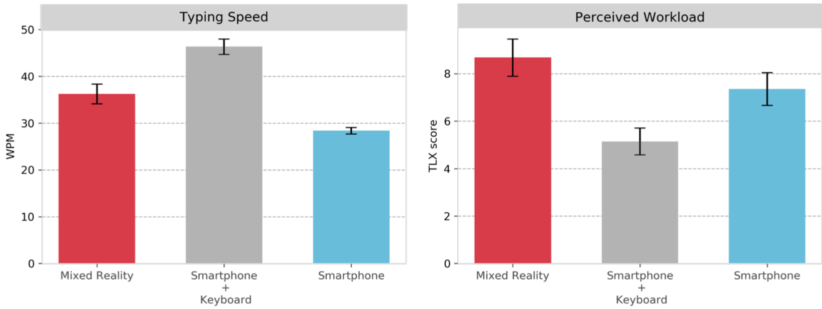


Figure 6.10: Mean values of words per minute (text input task), and NASA-TLX score (both tasks) for each condition. Error bars show the standard error of the mean (SE).

Error Rate

Besides the WPM, the typing and editing performance can also be expressed through the *Error Rate*. We calculated the ratio of the length of the input and the minimum number of insertions, deletions, or substitutions that are needed to transform the presented text into the transcribed on [164]. The results neither show a significant effect of SETUP on the *Error Rate* for the typing task, $F(1.81, 41.68) = 1.109$, $p = .339$ nor for the copy editing task $F(1.93, 44.40) = .702$, $p = .496$. Besides, we calculated the *Corrected Error Rate* [231], which represents the effort put into correcting errors. We found no significant effect of SETUP regarding the number of corrections, $F(1.37, 31.41) = 0.301$, $p = .658$.

Task Completion Time (TCT)

For the copy editing task, we measured the TCT as a performance indicator. We measured from the very first keypress to the confirmation keypress for each text. We found a significant main effect of SETUP on the TCTs of the copy editing task, $F(2, 46) = 25.86$, $p < .001$. A post-hoc tests revealed significant differences between *Smartphone + Keyboard* and *Mixed Reality* (M = -102.41, E = 21.93, with $p < .001$), between *Smartphone* and *Mixed Reality* (M = -140.60, SE = 20.92, with $p < .001$), but no significant effect between *Smartphone + Keyboard* and *Smartphone* ($p = .120$).

Task Load Index

We assessed the raw score of the NASA-TLX [95], representing the perceived subjective workload the participants had while inputting or copy editing text. We found a significant main effect of SETUP on the perceived workload, $F(1.61, 36.93) = 13.83$, $p < .001$. Post-hoc tests revealed a significant difference between *Smartphone + Keyboard* and *Smartphone* (M = -2.22, SE = 0.52, with $p < .001$), between *Smartphone + Keyboard* and *Mixed Reality* (M = -3.54, SE = 0.63, with $p < .001$), but no significant effect between *Smartphone* and *Mixed Reality* ($p = .393$).

System Usability Scale (SUS)

To receive an indication of the overall usability of our apparatus, we assessed the SUS [31]. We found a significant effect of SETUP, $F(1.74, 39.97) = 32.70$, $p < .001$. Post-hoc tests revealed a significant difference between the conditions *Smartphone + Keyboard* and *Mixed Reality* (M = 23.16, SE = 3.29, with $p < .001$), between *Smartphone* and *Mixed Reality* (M = 19.27, SE = 3.51, with $p < .001$), but no significant difference between *Smartphone + Keyboard* and *Smartphone* ($p = .295$).

AttrakDiff

To gain further insights into the perceived user experience, we used the AttrakDiff questionnaire, which accesses the user experience divided into pragmatic and hedonic quality. Participants rated the system by ranking word pairs of different dimensions. The results are shown in Figure 6.11. The top diagram classifies the apparatus into character areas (i.e., *self-oriented* or *action-oriented*). The bottom diagram shows the mean values of the dimensions of AttrakDiff. The results show that the *Mixed Reality* setup has the highest hedonic quality, but the lowest pragmatic quality. According to the diagram, the characteristics of the apparatus is not unambiguous and lies between the areas *neutral* and *self-oriented*. The other two setups, *Smartphone + Keyboard* and *Smartphone* lie in the characteristics area of *action-oriented*, thus were rated more practical.

Personal Preferences and Qualitative Results

After conducting the user study, we asked the participants regarding their preferred SETUP and to provide additional qualitative feedback. Participants ranked *Smartphone + Keyboard* as the best solution for portable text input and editing, followed by the *Mixed Reality*, which is directly followed by the *Smartphone*

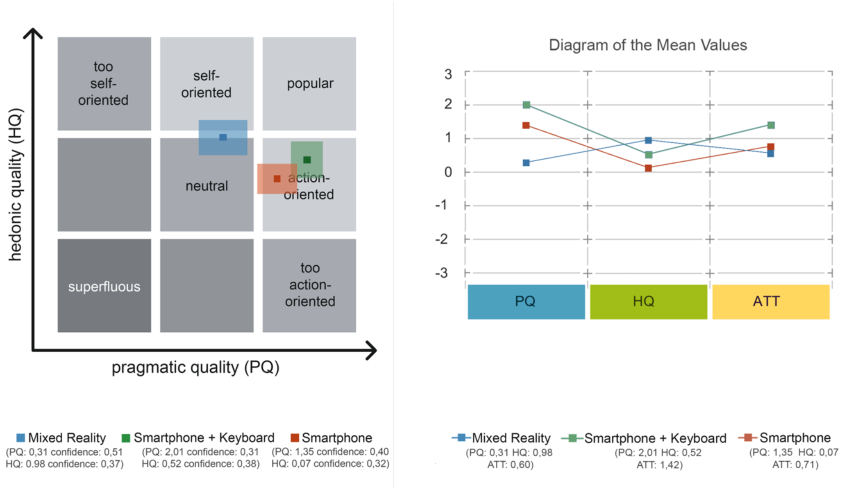


Figure 6.11: Diagrams from the AttrakDiff questionnaire revealing the characteristics including the pragmatic quality (PQ) and hedonic quality (HQ) (top diagram) and the mean values of the dimensions (bottom diagram).

setup. Participants endorsed the great display-space and privacy in MR, however, complained about occasional orientation problems due to the limited field of view of the HMD.

6.7.3 Discussion

Considering text input, we found that our mixed reality apparatus led to significant higher words per minute compared to soft keyboard input. Results did not show significant changes in the error rates of the typed text. Further analysis revealed that the slightly higher workload and lower usability caused by the HMD was mainly compensated through the support of the physical keyboard. For copy editing texts, the mixed reality led to a significantly higher task completion time (TCT) compared to both the smartphones soft keyboard and the smartphone and keyboard combination. Further, the analysis revealed that participants benefit from the large virtual display space but got thwarted by the lacking opportunity to quickly navigating the text (e.g., touch or mouse). Adding mouse support or alternative methods to place the cursor quickly might have yielded different results considering the TCT. The analysis of additional qualitative feedback unfolded that participants overall enjoyed our apparatus. They envisaged working in enhanced virtuality and highlighted the larger display area and the possibility to collaborate in future scenarios. We argue that optimizing the setup and further improve the interaction modalities is necessary. Improved positioning of the keyboard visualizations and multimodal input for copy editing are relevant parameters to improve portable mixed reality text entry.

6.8 Conclusion

In this chapter, we presented our work on exploring the effects and potential of enhanced virtuality through blending reality back into the virtual realm. More specifically, we have shown VR's potential for a wide variety of use cases by enabling natural generic text input on a physical keyboard while being immersed in a virtual environment. We present our research-probe that comprises calibration-free, low latency, and accurate finger tracking with an HMD. Thus we can create virtual environments allowing for effortless typing in VR.

The results of the conducted user-study indicate no significant difference in typing speed for the experienced typists while being immersed in the enhanced virtuality. Further, results show that all typists benefit from seeing a representation of their hands during non-contiguous typing.

Based on these results, we further investigated a portable low-fidelity solution for text input in mixed realities. To achieve this we developed a second research-probe t comprising a smartphone, a virtual reality viewer, and a wireless keyboard. We compared state-of-the-art smartphone soft keyboards to physical keyboard input and our mixed reality approach.

The study results indicate that participants have significantly higher input speeds when immersed in mixed reality compared to regular smartphone input, while error rates remain low. Furthermore, our participants enjoyed interacting with the large virtual display, even though copy editing required considerably more time to complete.

Chapter 7

Touching Virtuality

Current VR consumer devices comprise visual and auditory displays capable of providing very high degrees of realism and thus allow the sensation of high levels of presence. However, VR technologies are lacking in presenting the virtual environment's appearance to the other human senses. In particular, there has been less emphasis on stimulating the different components of the haptic system even though it can increase immersion [110] and users' performance [242].

Recently, researchers have addressed the challenges of providing appropriate stimuli for simulating pressure, vibration, touch, temperature, or pain. Current solutions include vibrotactile gloves [27], belts [248], or wests [154]. However, these approaches have specific limitations, such as carrying additional hardware, insufficient expressiveness, or limited feedback areas. Since the users' sense of presence in VR is a crucial factor for the overall user experience and haptic feedback is an essential component for interactions in VR [148] we explore how levitating tangible elements can further enhance the experience of VR environments.

In this chapter, we investigate the effect of enhancing virtual reality by incorporating levitating tangible objects that provide flexible haptic input and output. First, we present previous works that influenced our design decisions. Then, we introduce the development and implementation of the research probe; a system that utilizes autonomous flying quadcopters as levitating haptic feedback proxy in VR. Quadcopters are dynamically positioned in the physical interaction space to provide haptic feedback according to the VR environment. The user can freely explore and interact with the VR environment and sense haptic stimuli while not being required to wear any additional devices. In a user study, we explore how the haptic feedback provided by the quadcopters affects the users' sense of presence compared to vibrotactile controllers. Finally, we present possible application scenarios and discuss guidelines for building convincing haptic experiences using levitating haptic proxies.

Parts of this chapter are based on the following publications:

- P. Knierim, T. Kosch, V. Schwind, M. Funk, F. Kiss, S. Schneegass, and N. Henze. Tactile Drones - Providing Immersive Tactile Feedback in Virtual Reality through Quadcopters. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 433–436, New York, NY, USA, 2017. Association for Computing Machinery
- M. Hoppe, P. Knierim, T. Kosch, M. Funk, L. Futami, S. Schneegass, N. Henze, A. Schmidt, and T. Machulla. VRHapticDrones: Providing Haptics in Virtual Reality through Quadcopters. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pages 7–18, New York, New York, USA, Nov. 2018. ACM

7.1 Related Work

In this section, we introduce the two streams of research that inspired our development. Specifically, haptic feedback in VR and human-quadcopter interaction.

7.1.1 Haptic Feedback in VR

Most commercial VR devices provide haptic feedback through vibrotactile actuators, integrated into handheld controllers. However, there are other wearable devices to provide haptic feedback in VR, e.g. using electrical muscle stimulation [160, 161], vibrotactile vests [155], head-worn motors [89] and gloves [35], and exoskeletons [183]. Previous work also developed approaches for providing realistic feedback for handheld devices. For example, [103] presented a new handheld haptic device which uses propeller propulsion to generate 3-DOF force feedback that is generated by six motors that are attached to a handheld frame.

Benko et al. [24] proposed to augment a handheld controller with a device that can convey the shape of an object in VR and its texture using a 4×4 matrix actuated pins. To overcome the fact that the users still have to hold a controller, Gu et al. [86] presented an exoskeleton which is mounted on the user's hand to provide haptic feedback in VR. As the exoskeleton prevents fingers from moving when a virtual object is in reach, the user has the feeling of haptic resistance from the virtual object. To provide haptic feedback at different body locations, Signer and Curtin [228] developed a body-worn construction, which is overlaid by holograms for providing a tangible and haptic AR experience. Systems using this technique will soon become commercially available. Further, Schmidt et al. [215] created a user-mounted device for simulating steps in VR.

To provide highly realistic haptic feedback, Simeone et al. [229] proposed to repurpose objects that are already in the environment of the user. The authors arrange a virtual environment according to the physical environment to use existing objects for their – already existing – haptic capabilities. By scaling this down to an object granularity, Hettiarachchi and Wigdor [1] use the physical properties of objects to spontaneously create haptic experiences, while Sun et al [239] scale up this approach to a world level. Furthermore, Cheng et al. [38] recognized the capabilities of using humans for their ability to spontaneously create haptic experiences, which they could scale up to providing haptic walls [39].

Another trend is to build systems for providing haptic feedback that is scalable, programmable, and can be placed in the environment. For example, Araujo et al. [13] use a robotic arm and a cube with different surfaces for providing different haptic experiences to a user wearing an HMD. Depending on where the user touches a virtual object, the robotic arm rotates the cube in a way such that always the correct surface is being touched. Furthermore, He et al. [98] suggest using small mobile robots as a haptic proxy for VR tabletop applications, while Jeong et al. [120] suggests creating haptic experiences using movable wires. Also, regular objects, e.g. furniture can be augmented to create a haptic experience [90]. Another system that augments a tabletop has been presented by Follmer et al. [66]. The authors purpose a dynamic shape display for displaying forms and shapes according to the digital input. This can be used to dynamically provide haptic feedback for VR scenarios at a fixed position. Conversely, instead of making the environment scalable, other research focuses on making the user believe that the haptics of the environment is matching the virtual scene. Azmandian et al. [16] propose a technique called haptic retargeting for physical feedback in VR. Thereby, the user's hand is redirected to touch a single object that is in the user's proximity, while the user believes that multiple objects are present.

Figure 7.1: *TactileDrone* supports three different feedback modes: (left) Passive: The object is levitating and the user can touch it. (center) Active: The object is proactively bumping into the user. (right) Proxy: The object can be grasp and moved by the user.

7.1.2 Human-Quadcopter Interaction

Since the proliferation of small quadcopters in the research domain of Human-Computer Interaction [69], they were mostly used as a flying camera [213] or for navigation purposes [15, 42, 141] as outlined in Chapter 4. This changed when Gomes et al. [77] proposed BitDrones, quadcopters that can be tracked and controlled. The quadcopters are equipped with LEDs, screens, and a cage to make them graspable. The BitDrones project is one of the first approaches to use quadcopters as flying input devices. In contrary, Kosch et al. [146] show how a remote control can be used as input for quadcopters. Abtahi et al. [10] investigated the social interaction properties of quadcopters in a cage and quadcopters without a cage. Additionally, Yamaguchi et al. [267] proposed using a quadcopter that is carrying a canvas as a haptic target for a sword fight. In their prototype, they use the drone as a resistor that the user feels to have hit the enemy. Recently, Knierim et al. [136, 140] showcased using autonomous drones as haptic agents that make contact with the users to provide feedback that is passively received by the user. Abdullah et al. [9] use a quadcopter and hand tracking for providing 1D haptic feedback.

Overall, related work recognized the need for haptic feedback to make VR experiences more immersive. Other related work used quadcopters as an input device and a haptic target. To combine these two aspects a scalable platform for managing quadcopters to stimulate the user at the right body positions is required. In this chapter, we extend previous work by using quadcopters to provide active haptic feedback in VR, where the user is actively reaching out to make contact with quadcopters that are used as haptic proxies in order to simulate the surfaces of virtual objects.

Figure 7.2: Quadcopter prototypes to deliver different types of feedback. (left) passive feedback, (center) active feedback, (right) haptic proxy feedback.

7.2 Realizing Haptic Feedback through Quadcopters

With the *TactileDrone* research probe we can dynamically provide haptic feedback in VR thorough levitating tangibles. *TactileDrone* adds haptics to virtual objects and allows the user to sense a haptic stimulus while immersed in a VR environment. To flexible enhance virtuality, we envision to provide three different types of haptic feedback in VR: passive, active, and positioning haptic proxies. We showcase these three types of haptic feedback using the underwater world scenario depicted in Figure 7.1.

7.2.1 Passive Feedback

While users are immersed in our underwater world, they begin in a dim surrounding with only one glowing sphere floating in front of them. Users can explore the dark space by walking and looking around. Not being limited to looking and walking users can also haptically explore the surroundings. When reaching out with their hands to touch the glowing sphere, they can feel the resistance of it. After touching the sphere, the scene gets illuminated. An anglerfish (Figure 7.1 left) becomes visible and swims away.

The haptic sensation is provided by an encased quadcopter with a passive surface by levitating at the virtual position of the sphere. For any virtual object which should provide a haptic stimulus when touched the *TactileDrone* system dynamically aligns a touchable surface of a quadcopter with the virtual objects. This is not only limited for objects but can also be facilitated for user interface elements like buttons. Figure 7.2 showcases the modified quadcopters.

7.2.2 Active Feedback

As a second feedback category, our underwater scenario contains elements that actively engage with the user. A shark is appearing in the users' vicinity and directly swims towards them. Active feedback is provided to amplify the impact of the shark nudging the user. In such a scenario, quadcopters are controlled to actively contact the user's body-parts according to the location of the virtual object. The shark nudging the user is illustrated in Figure 7.1 (center). Figure 7.2 (center) shows how we implemented the concept using a quadcopter.

7.2.3 Haptic Proxy Feedback

For more complex haptic feedback, which goes beyond active and passive exploration, *TactileDrone* can provide haptic proxies. Haptic proxies are small and lightweight tokens the user can touch and interact with. Quadcopters place these proxies at the required position to enable seamless interaction. Figure 7.1 (right) shows a worm attached to a fish hook, which is lowered to the seabed. Users can grab the worm and take it as a trophy. A rubber worm is attached to the quadcopter as a haptic proxy to enable this kind of interaction. The quadcopter is hovering, whereby the proxy's physical location matches the virtual one. Figure 7.2 (right) shows a quadcopter capable of providing haptic proxy feedback.

7.3 Implementation

TactileDrone, comprises a high-speed motion tracking system, quadcopters as haptic feedback appliance, a VR HMD and a software backend. The motion tracking System captures the position of the user's HMD and the quadcopter and streams it to the *TactileDrone* backend core. The core processes all data and sends updates regarding the quadcopter's position to the PID-Controller and scene events to the VR Renderer. The PID-Controller takes care of maneuvering the quadcopters, while the VR renderer processes updates from the core and displays the VR scene on the HMD. An overview of all components and connections is shown in Figure 7.3.

Tracking System The system tracks the HMD, quadcopters and defined body parts. All data is streamed to the *TactileDrone* backend. We set up a Motive OptiTrack motion capturing system with 12 Flex 3 cameras covering an interaction space of 4 m × 4 m × 3 m. It samples with 100 Hz at a millimeter accuracy. Quadcopters can hover anywhere around the user inside this volume. A Leap Motion sensor is mounted at the front of the HMD and is used for tracking the user's hands, which enables to include them into the *TactileDrone* system.

Haptic Drone Each quadcopter is used as a haptic feedback interface. Different lightweight haptic extentions can be attached to the quadcopters (see figure 7.2). Our implementation is based on the commercially available Parrot Rolling Spider quadcopter. They are powered by a 550 mAh battery, providing approximately 6 min of flight time depending on the attached haptic proxy. We removed all the unnecessary panels, such as casings, to increase the payload capacity. The maximum weight of the haptic proxy including the markers for the tracking system is 10 g. The quadcopter connects via Bluetooth low energy to our *TactileDrone* backend. The underlying Linux OS processes steering commands with 20 Hz.

***TactileDrone* Backend** The *TactileDrone* Backend interconnects the VR rendering engine, the quadcopter control, and the motion tracking system. Our system runs on a workstation with an Intel i7-6700 processor, 16 GB of RAM, and an NVIDIA GeForce GTX 970 running Windows 10.

Core Our software backend core processes the streamed location data and controls the quadcopter. Furthermore, trajectory planning and synchronization with the virtual world renderer are processed inside this component. Further, it maintains quadcopter, users, and interaction states and manages the application behavior.

VR Renderer The VR scenarios are rendered by the Unity3D game engine and are displayed on an Oculus Rift HMD. The VR Headset equipped with a Leap Motion for displaying the user's hands within VR. As proposed by Schwind et al. [222], we used a neutral hand style representation to avoid potential biases of our participants. A set of reflective markers are attached to the HMD. Positional data is forwarded from the *TactileDrone* core to the rendering engine and is rendered accordingly. The game engine further calculates collisions between virtual objects represented by either quadcopters or human body parts and reports back to the *TactileDrone* core.

PID-Controller The PID-Controller component wirelessly sends control signals to the quadcopter over Bluetooth LE to direct the quadcopter to a particular

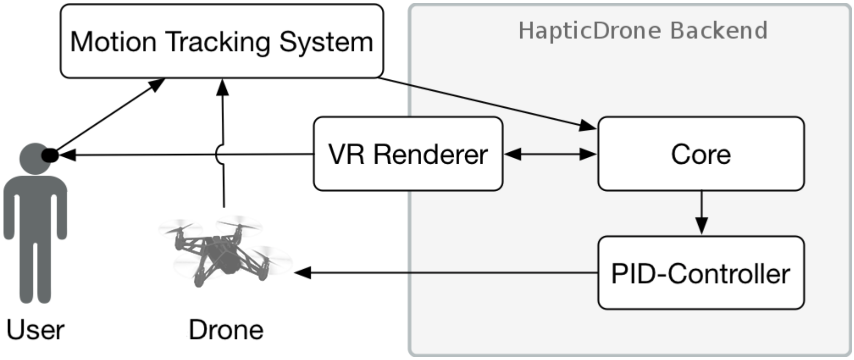


Figure 7.3: All components of the *TactileDrone* system.

location. Data transmission is exposed through a local nodeJS server application. A set of four Proportional, Integral and Differential (PID) loops to control the movement of the quadcopter towards the positions managed by the *TactileDrone* core. During hovering, the quadcopter relies on its own IMU, ultrasonic sensor and down facing the camera to stay at a fixed position.

7.4 Method

In this study, we focused on providing a physical surface that delivers haptic feedback when the user explores virtual objects. We assume that interactions that involve the user's hands result in a higher presence when hand-held devices are not needed for interaction or haptic feedback. Therefore, we conducted a user study to examine the increase of presence which typically accompanies by increased immersion.

We used Unity3D to create a scene that resembles a birthday party. Participants interacted playfully with a balloon in VR. We hypothesized that adding haptic feedback via a quadcopter-positioned surface to hand tracking would result in a higher presence than providing no haptic feedback while using hand tracking or providing state-of-the-art feedback through a controller with vibration.

We defined the feedback modality as the only independent variable with three levels: *No Haptic Feedback* (1), *Vibrotactile Feedback* (2), and *Quadcopter Feedback* (3) delivered by *TactileDrone*. In each of the experimental conditions, participants saw a virtual representation of their hands, as such visualizations are essential parts of interactions. Haptic feedback is only triggered when the balloon in the virtual scene is touched by the participant.

7.4.1 Participants

We recruited 12 participants (6 female; 6 male) aged from 17 to 28 ($M = 21.6$, $SD = 1.84$) via the local university mailing lists. All participants had normal or corrected-to-normal vision. Participants who were in need of vision correction had to wear contact lenses to avoid wearing glasses under the HMD.

Five participants had no VR experience, seven had minor experience with VR, i.e. five minutes up to three hours. Except for one participant, none had experience with the Leap Motion sensor.

7.4.2 Apparatus

The apparatus consists of an Oculus Rift and noise-canceling headphones to negate the buzzing produced by the drone. For the *Vibrotactile Feedback* condition (2), we used the Oculus Touch controller as well as the Oculus tracking system. For the *No Haptic Feedback* (1) and *Quadcopter Feedback* (3) conditions we used a Leap Motion sensor attached to the front of the HMD for hand tracking. The positional tracking of the participant's hands was of equal quality in each condition. The hand models were adjusted to look like the Oculus Touch hand models used in the *Vibrotactile Feedback* condition (2).

We used the OptiTrack system for positional tracking during the *Vibrotactile Feedback* (2) and *Quadcopter Feedback* (3) condition. For the *Quadcopter Feedback* condition (3), we used a Parrot Rolling Spider quadcopter, including attached wheels with tulle-textile covers as the touchable surface (as depicted in Figure 7.4).

The physical interaction with the balloon was only affecting the back and forward movement of the balloon i.e. moving away from the participant in the room. The up and down and sideways movement was animated via unity physics to ensure a balloon-like behavior. This restriction ensured comparable balloon behaviors between the three conditions. As participants only interacted with the front-side of the objects, the touch surface did not only provided feedback but also served as a protection to not get in contact with the quadcopter's rotors.

Figure 7.4: Participant exploring the virtual balloon during the first user study through touching and pushing.

7.4.3 Procedure

After being introduced to the system and task the participant filled out the consent form and the demographic questionnaire. When the participant put on the HMD, he or she was immersed by the virtual birthday party scene. The experimenter asked the participant to interact with the balloon. The balloons responded naturally to interactions such as touching and pushing.

We used a within-subject design, hence each participant performed all three conditions. The order of the conditions was counter-balanced across participants using a Balanced Latin Square design. Participants interacted with the balloon for 3-5 minutes and were encouraged to start with an extended index finger and to try out different hand postures in each condition. After each condition, participants filled out a Presence Questionnaire (PQ) [261]. While being interviewed by the experimenter, the participants were encouraged to provide suggestions and concerns about the system. After completing each condition, the participants answered questions regarding the comparison and liking of each condition.

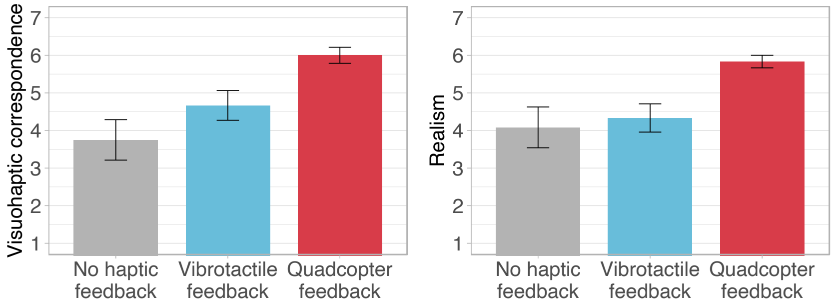


Figure 7.5: Mean values of visuohaptic correspondence and realism for each condition. Error bars show standard error of the mean (SE).

7.5 Results

The results of the full-scale PQ for each condition are: *No Haptic Feedback* 142.5 points (SD = 20.2), *Vibrotactile Feedback* 142.0 points (SD = 16.1), *Quadcopter Feedback* 164.5 points (SD = 15.8) (see Figure 7.6). Presence ratings differed significantly between the three conditions (Friedman test, $\chi^2(2) = 18.681$, $p < .05$). Wilcoxon tests were used to follow-up on this finding (Holm-Bonferroni corrections were applied to control the family-wise error rate). We found that using *Quadcopter Feedback* (3) significantly improved presence ratings compared to using *No Haptic Feedback* (1) (p = .004) or using *Vibrotactile Feedback* (2) (p = .003). Using a controller does not increase presence when compared to the hands only condition (p = .326).

A closer look at the subscales of the PQ reveals that the difference in presence can be traced back to an increase in the *Haptic Visual Fidelity*-subscale ratings. The quadcopter condition is rated higher than the controller (p = .002) and the hands-only condition (p = .006). Furthermore, the *Adaptation/Immersion*-subscale rating of the quadcopter condition is rated higher than the controller (p = .004) and the hands-only condition (p = .019).

To further evaluate the experience provided by *TactileDrone*, we presented three Likert-items to the participants, that specifically targeted the haptic feedback. In particular, we asked participants to rate on a scale from 1 to 7 how much they agree with each of the following statements: (1) “The visuals match what I feel.“, (2) “How realistic did it feel? “ and (3) “I was able to pass through the object.“ We used Wilcoxon tests to perform pair-wise comparisons between the ratings for the

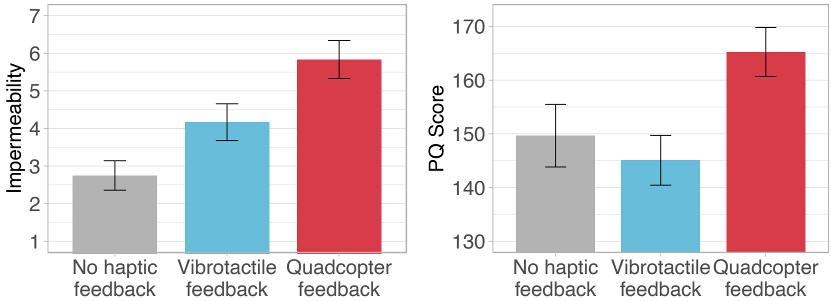


Figure 7.6: Mean values of impermeability and presence score (PQ) for each condition. Error bars show standard error of the mean (SE).

conditions. The results demonstrate that our system improves the correspondence between haptic and visual feedback. The quadcopter condition is rated higher than the controller (p = .003) and the hands-only condition (p = .003). It also increases the realism. The *Quadcopter Feedback* condition is rated higher than either the *Vibrotactile Feedback* (p = .005) or *No Haptic Feedback* condition (p = .008). Furthermore, it improves the perceived solidity of objects. The *Quadcopter Feedback* condition is rated higher than either the *Vibrotactile Feedback* (p = .041) or *No Haptic Feedback* (p = .012). Overall, the results show that providing haptic feedback via a quadcopter creates a higher level of immersion compared to state-of-the-art controllers. Therefore, this induces a stronger feeling of presence in the participants as this kind of feedback represents an exploration that is closer to the real world experience of touching an object.

Subjective Comments

One participant mentioned that using the controller (*Vibrotactile Feedback*) felt like holding an object that is used to push the balloon. Another participant noted the same experience and added that

> "[...]it did not feel like a balloon because you can feel the controller and the vibration was only in the palm and not on the fingertips." *(P1)*

These comment show that interacting with a controller lacks direct and unintrusive feedback and interaction and is hence inappropriate / not mature enough for a having a realistic VR experience.

Two participants commented that they preferred the hand tracking via the Leap Motion sensor.

> “[. . .]the controller was not able to detect that [hand posture] type of detail and did not reflect it visually” *(P6)*

The state-of-the-art Oculus Touch controllers do not allow accurate hand tracking and are only able to provide a fixed approximation of hands divided into discrete hands postures.

Nine participants explicitly valued that the quadcopter combined both a mature hand tracking while still being able to provide haptic feedback. Six participants remarked that the quadcopter feedback felt real and natural.

> “It is really exciting! I didn’t expect it to work that well already. Even the “Hands-Only“ condition was working really well, despite the lack of feedback. Controllers are annoying because you have to hold something in your hands. The “quadcopter“ was cool because there even was something there where I touched the balloon.” *(P2)*

After experiencing the quadcopter’s haptic feedback, participants already imagined further use cases. The quadcopter could be used for providing improved feedback in 3D drawing apps. These do not have any feedback other than vibration while drawing in the air or when two brush strokes collide. A further use case was using the quadcopter as and flying inventory to select usable items.

7.5.1 Discussion

The results of the user study indicate that our research probe outperforms on state-of-the-art interactions in VR environments concerning haptically exploring and interacting with virtual objects. The feedback is perceived as more immersive, realistic, and better suited to the visuals of the virtual object. Haptic feedback that is provided by controllers, which are currently the most readily available technology, lacks in comparison behind. When quadcopter-mediated haptic feedback is provided, participants report a higher sense of presence and find the interaction method less intrusive.

While haptic feedback that is mediated via quadcopter-positioned surfaces has various advantages, the variety of objects that can be presented is still restricted by the force that the quadcopter can generate. Objects with a higher mass or stationary objects do not easily yield to pressure when touched or prodded. In contrast, the quadcopter-positioned surface will not remain stationary if the force applied by the user is larger than the counterforce provided by the quadcopter. On the other hand, this allows for providing more dynamic feedback. This can result in objects exhibiting haptic properties that do not comply with the physical expectation, such as a solid wall feeling unbalanced. It also can be used intentionally since other feedback methods are not able to provide such a feeling. While humans tolerate some discrepancy between visual and haptic experiences, the cohesion of the multisensory percept will be diminished if the discrepancy becomes too large. In other words, the haptic experience will no longer appear realistic.

Limitations

The current implementation of the research probe and the presented study is still prone to certain limitations. The general limitations of human-drone interaction apply for our research probe. Theses include limited payload and flight time of the quadcopters and audible noise. However, in our setup, we can compensate for these challenges with multiple alternating quadcopters and active noise-canceling headphones.

During the study, we used only one quadcopter at a time. This choice reduces the complexity of our flight control component in the first place. However, the used Bluetooth stack supports a simultaneous connection to up to five quadcopters. Using several quadcopters at the same time may increase complexity in trajectory planning, collision prevention, and noise level, but it allows higher frequencies of multiple feedback interfaces at a time. Further, we only explored the appearance of one virtual object. Further investigations of the influence of the expected mass and behavior of different objects are presented in our full publication.

7.5.2 Application Scenarios

The ability to provide passive, active, and proxy-based feedback makes *TactileDrone* very versatile. We envision dynamic haptic feedback thought quadcopters in several use cases, which we present in the following.

Gaming and Entertainment

Besides the described underwater world, *TactileDrone* offers the potential to further enhance games and entertainment. Passive feedback for almost any virtual

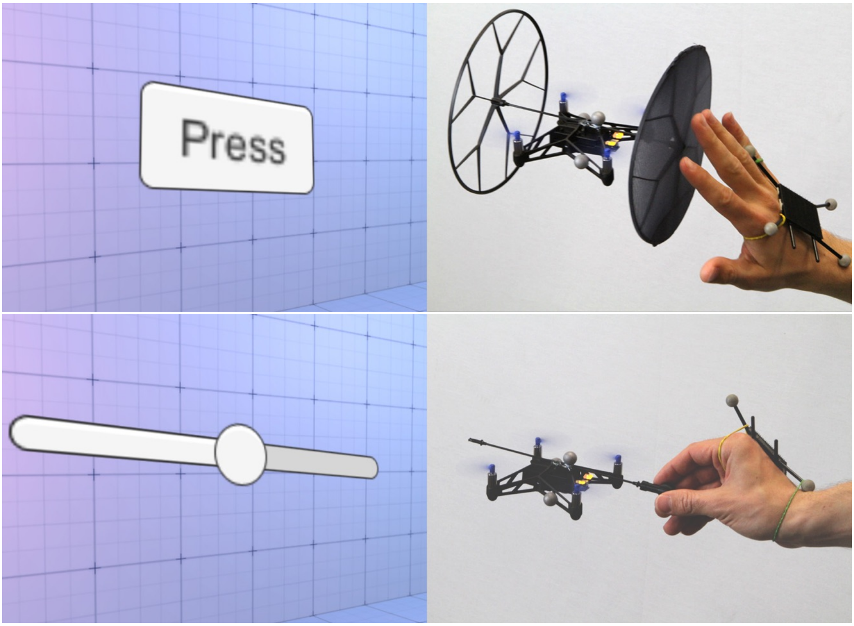


Figure 7.7: Showcase of virtual UI elements with haptic feedback provided by quadcopter. Top: button element with passive feedback quadcopter , Bottom: slider element with haptic proxy quadcopter.

object can be provided. We esteem feedback to be especially valuable when users interact with the environment. In a game, players could be asked to open a door (passive feedback). When the door swings open arrows are shot at the player from behind the door (active feedback). To stop getting shot by arrows the player must pull out a spring from a device to disarm the arrow trap.

Construction and Design

We also envision application scenarios in the construction and design domains. Car designers can benefit from *TactileDrone* during the design process while potential customers can virtually touch (passive feedback) their new car before ordering. Further, customers could go for a virtual ride supported by several haptic proxies surrounding the user, such as the knob to turn on the air condition.

User Interface Elements

While using cameras that track the user's hands in VR, such as the Leap Motion, no haptic feedback is provided when interacting with User Interface (UI) elements. Our System can represent haptic virtual buttons (passive feedback) while maintaining hands-free interaction. Further, elements or notifications not in the field of view of the user could gain attention by nudging the users' left or right shoulder (active feedback). Finally, virtual sliders could have a haptic proxy anywhere in space to facilitate intuitive usage (haptic proxy). Virtual UI elements and their haptic counterpart are showcased in Figure 7.7.

7.6 Conclusion

We explored with this research probe how to provide flexible haptic input and output in VR and how VR environments can be enhanced through levitating tangible objects. The developed *TactileDrone* apparatus supports three different types of haptic feedback; namely, passive, active, and proxy-based feedback. Further, the apparatus can provide different types of feedback with various positions and various intensities depending on the current VR environment.

We conducted a user study showing that haptic feedback provided by quadcopters significantly increases the user's sense of presence in the VR environment compared to traditional vibrotactile feedback or no haptic feedback. While we used only a single quadcopter at a time, this research can lead to scalable and unintrusive haptic feedback for large VR environments. Therefore, we already present different application scenarios that could benefit from haptic feedback.

Chapter 8

Interacting in Mixed Realities

Current MR devices offer only a limited set of interaction capabilities, that lack expressiveness. Voice commands, for example, are primarily supported but can be unreliable under excessively loud conditions. Further, using voice commands to interact with an MR system is not widely prevalent. Mid-air hand gestures are an alternative input method to interact with the MR system. Besides the gesture itself users need to learn about the approximate boundaries of the gesture-sensing space in which they can interact not to break the gesture tacking space. Despite mid-air gestures being known to cause fatigue, many MR glasses rely heavily on gestures as the primary interaction method (cf. Section 2.4.4).

In order to enhance the interaction within MREs, past research proposed various systems combining HMD and novel controllers. Smartphones were facilitated as secondary output screens or as input devices using touch or other built-in sensors. However, the potential of smartphones as an MR controller has not yet been fully exploited.

In this chapter, we present and evaluate a multimodal interaction concept by pairing a smart-phone as an input controller with AR glasses. In a user study, we investigate the effects on interaction speed, accuracy, and workload in different tasks. We can show that our smartphone-based controller results in significantly faster and more accurate interaction and reduced cognitive workload compared to state-of-the-art mid-air gestures. Concluding the chapter, we discuss how future AR systems can benefit from touchscreens as an additional and complementary interaction modality.

The chapter is structured as follows. First, we review relevant related works that address the challenges of interaction in MR. We then describe our approach of facilitating dynamic MR interfaces that allow for higher bandwidth of input and output. Following this, we present the methodology of two user studies, accompanied by the results. Finally, we discuss insights, challenges, and opportunities for smartphones as a ubiquitous controller for MR experiences.

Parts of this chapter are based on the following publication:

- P. Knierim, D. Hein, A. Schmidt, and T. Kosch. The SmARtphone Controller: Leveraging Smartphones as Input and Output Modality for Improved Interaction within Mobile Augmented Reality Environments. *i-com Journal of Interactive Media*, 20(1):49–61, 2021

8.1 Related Work

The developed research-probe builds on past research in AR and the development of novel interaction concepts for AR, VR, and mobile devices. Here, we review research that motivated our development and work that explores the interaction space for mobile AR. Afterward, we summarize and discuss the interaction capabilities of selected commercially available AR and VR devices.

8.1.1 Interaction Concepts in AR and VR

Today, we use smartphones as a ubiquitous computing device to interact with our environment [19]. We control our home appliances, buy tickets, navigate or engage with location-based games. Current smartphones are equipped with numerous sensors that facilitate a good understanding of the environment. Further, devices are becoming more connected than ever and act a remote interface for current cameras[13], speakers[14], or toys[15].

There is a large body of work, and various input techniques have been proposed to interact with virtual elements presented by an HMD in a mobile augmented

[13] GoPro gopro.com/en/us/shop/softwareandapp

[14] Spotify - spotify.com/us/connect/

[15] Parrot - parrot.com/us/freeflight-6

environment. From mid-air hand gestures to foot-tapping [182] viability solution for interacting with virtual elements exist. In smartphone-enabled handheld mobile AR experiences, direct selection and manipulation of objects give a natural and convenient user experience. However, maintaining visual tracking while holding the smartphone in one hand and interaction via touch with other is challenging [18, 121]. With the availability of optical see-through smartglasses novel interaction techniques where investigated. Wearable input is often facilitated through touch surfaces encircling users' clothes [51, 217] or even fully garment-integrated sensors [130] that are enriched by AR. Others used sensor-enabled smartwatches to provide natural interaction with virtual objects [108, 126].

With hybrid systems that are comprising AR, VR or large displays as primary screen technology and an additional handled secondary display or smartphone as a controller enable seamless advanced interaction in mobile context [33, 92]. Similarly, bring-your-own-device approaches have been proposed before to support spontaneous interaction with large public displays via the user's smartphone [199] and were also deployed for multiplayer gaming [171].

Prominent interaction metaphors for object selection and manipulation are image plane based or ray casting techniques. Motions of the controller are translated to a spatial ray or are projected onto a plane to support interaction. The IMU integrated in tangible interfaces [40] or smartphones [92] senses the orientation that is directly translated into the interaction space [107]. Sophisticated and highly specialized handheld controller were build. Incorporated with a touchscreen, 6 DOF tracking, and tactile buttons, interaction with immersive applications for VR and AR are viable [177, 209]. Recently, Mohr et al. developed an application, that turns a regular smartphone into a 6 DOF pointing and selection device to retrofit AR or VR HMDs. They confirmed the feasibility of repurposing smartphones as an input controller without any hardware modifications.

8.1.2 Current Interaction Concepts

The development of new AR and VR HMDs has ramped up over the last years. Mayor technology and entertainment companies have released the second to the third iteration of HMDs to the mass market. Comfort, weight, FOV, visual, and audio quality have been continuously improved. Nevertheless, interaction concepts have not changed significantly. In the domain of VR, interaction is mainly controller driven, allowing to intuitively grab, throw or precisely modify the virtual environment. Typically controllers require to be calibrated and are therefore unsuitable for mobile setups. New approaches using ultrasonic and magnetic sensors fused with gyroscope and accelerometer are promising but are still in its infancy. Unlike, commercial AR solutions offerer a more fragmented interaction space. Hand gestures, in combination with head pose are the most prominent ones. Lately, eye gaze was added to support more advanced and meaningfully interaction.

In a nutshell, VR environment interaction is controller based while AR mostly relates to embodied interactions. Integrating smartphones as a ubiquitous input and output device is still underrepresented and underestimated given the past research and possibilities that emerge by advances in smartphones.

8.2 Smartphone Supported Interaction in Mixed Realities

Figure 8.1: We conducted a user study comparing traditional mid-air hand gestures (left) to hybrid interaction with a smartphone (right) as in and output controller in Augmented Reality environments.

Current interaction with AR applications can be divided into pointing or selection and point manipulations in space. Interaction with free-floating Graphical User-interfaces (GUIs) or menus can be reduced as a combination of pointing and selection. The combination of these two interactions is the fundamental requirement for any interaction with AR environments. Our approach targets an easy to understand system, that utilizes known gestures and paradigms. For object manipulation, we focus on an eyes-free operation to do not distract the user from the AR elements and keep the cognitive load low, while keeping the flexibility and functionality high. Users can manipulate elements with high fidelity via swipes and touch gestures, buttons, and more sophisticated user interfaces can be displayed on the smartphones' multi-touch screen for fast and intuitive touch interactions. In both cases, kinesthetic as well as tactile feedback is provided through the smartphone. Auditive feedback is either provided via the smartphone or the HMDs speakers to guide the user's attention.

8.2.1 Object Manipulation

Object manipulation can be separated into translation, rotation, and scaling along the three different axes (i.e., the x-, y-, and z-axis). For object *translation*, users can move the focused object horizontally (along the x- and z-axes) by touching the screen and swiping horizontally or vertically. Users can adjust the height (y-axis) of an object by double-tapping with and subsequent swipe. For an intuitive translation, a reference coordinate system is created every time the user begins a translation. The coordinate system is created according to the head position concerning the object. Early user tests showed that the decoupling of head and smartphone rotation and position while translating objects is very intuitive and coherent to the user.

For the *rotation* around the y-axis, we selected a common approach known from map interactions with smartphones. Rotation is initiated by touching the screen with two fingers and then continuously rotate them around each other. To adjust the *scale*, we adopted the pinch gesture done with two fingers. The space between the fingers is translated directly to the size of the object that is modified. During object manipulation, no information is visible on the display. The interaction design is represented in figure 8.2. For comparison, we depicted the elaborated mid-air gestures supported by commercial AR HMDs in figure 8.3.

8.2.2 Secondary Screen Support

Supplementary to object modifications, AR applications often require input on free-floating or space anchored 2D GUIs. Our smartphone supported approach leverages two different possibilities to interact with these kinds of interfaces.

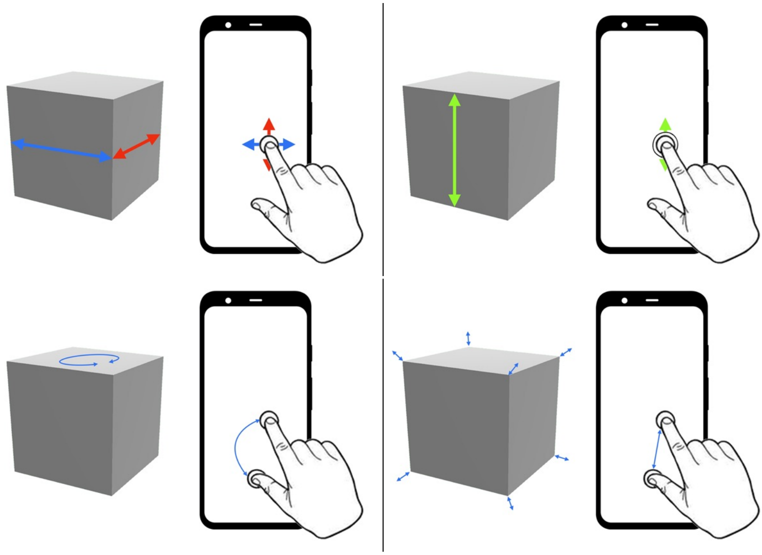


Figure 8.2: Interaction concept for object manipulation with a smartphone. From left to right, top to bottom: 1) single tap + swipe: translation along x- and z-axes (left/right, back/forth) 2) double tap + swipe: translation along y-axis (up/down) 3) two finger rotation: rotation 4) pinch: scale.

First, by implementing a simple remote-like controller similar to the object manipulation. Users can swipe and tap to interact with the in AR display space presented user interface elements like buttons, slides, or checkboxes. In the second approach, the entire user interface is transferred on demand onto the smartphones' display and allows seamless and direct touch and swipe interaction with the represented elements.

8.3 Implementation

To investigate the unique features of using a smartphone as ubiquitous input and output controller in AR environments, we implemented our apparatus incorporating a Microsoft HoloLens and a Google Pixel 2 XL Smartphone.

Smartphone Application To provide seaminess in- and output, we developed a native Android application. Users touch input, swipes, and gestures are

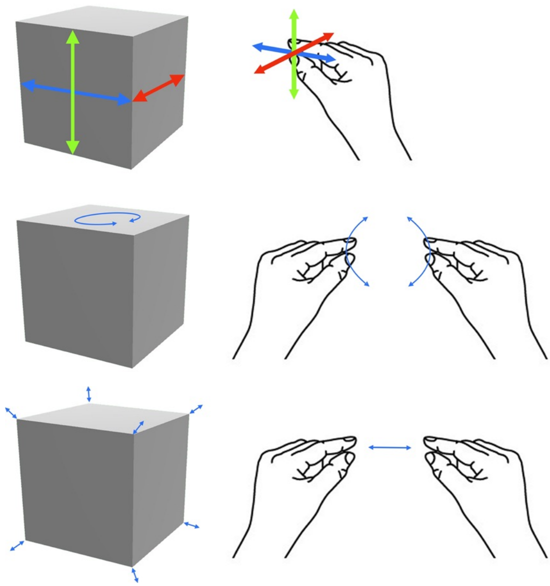


Figure 8.3: Default mid-air gesture interaction concept for object manipulation. From top to bottom: 1) Air tap and hold + move hand along the corresponding axis: translation along with hand movement (left/right, back/forth, up/down) 2) two hand air tap and hold + counter hand movement: rotation 3) two hand air tap and hold + move apart: scale.

sensed and directly sent to the AR HMD. User interface elements can be displayed on request of the AR HMD. Any GUI input is forwarded to the HMD accordingly.

For bidirectional communication between both devices, a Bluetooth radio frequency communication channel is established, since it does not require additional infrastructure. Touches, swipes, and gestures are transferred via this channel to manipulate objects. Besides, bidirectional status messages are transferred for UI input, states of the application, or haptic feedback control commands. For fast and platform-neutral serialization of the structured data, we used Google's Protocol Buffer.

HMD Application The HoloLens displays the AR environment and processes any incoming interaction commands from the Smartphone application. For a smooth and enhanced user experience, any continuum user-input is low pass

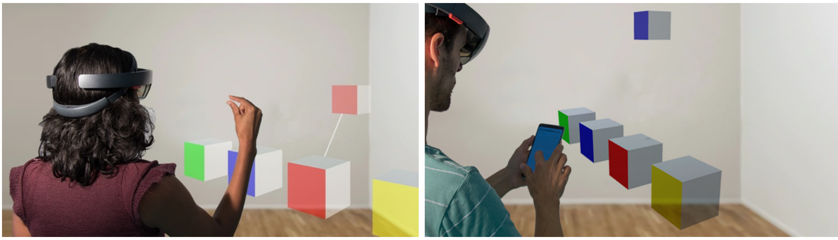


Figure 8.4: The first experiment includes manipulation of the position, rotation, and size of colored cubes. Arrows and lines help to compensate for the limited field of view and enhance orientation in space. The first condition includes manipulation via mid-air gestures (left). The second condition of the first experiment involves multi-touch as an input modality. Participants can manipulate colored cubes by swiping and multi-touch gestures on the smartphone.

filtered to remove any jitter. If selected virtual objects contain context menus, the presentation of these user interfaces is triggered on the connected smartphone.

Both, the HoloLens application and the AR environment are implemented using the Unity game engine 2018.4.7 and the Mixed Reality Toolkit v2.

8.4 Method

Our approach enables immersed users to manipulate virtual objects in AR environments with a smartphone. The goal of this study is to evaluate the effect of a touch screen as input and output modality in contrast to state-of-the-art mid-air gestures on object manipulation performance. We further investigate the effect of task complexity on overall task completion time and workload.

Two different tasks were elaborated to understand the qualities of a touchscreen as an in- and output controller regarding performance, user experience, and cognitive load. First, participants rearranged virtual objects in 3D space, followed by a set of modification tasks requiring to interact with a free-floating context menu.

We used a repeated-measures within-subject design with a within-subject IV INPUT. In the first experiment, INPUT has two levels: *mid-air gesture* as baseline and *multi-touch* since we focus only on the input capabilities of the smartphone.

In the second experiment, the IV has the additional level *multi-touch display* since we are also using the screen of the smartphone as output.

8.4.1 Subjects

We recruited 24 participants (12 female, 12 male) via our universities' mailing lists. Participants received either 10 EUR or course credits as compensation for their participation. Four had previous AR experience, one of them used AR glasses for professional purposes. The study received ethics clearance according to the ethics and privacy regulations of our institution.

8.4.2 Apparatus

The apparatus for this study comprised a Microsoft HoloLens and the Google Pixel 2 XL running Android 9. The smartphone has a bright, high-resolution display (538 PPI) with a presentation and interaction area of 136x68 mm. Both devices are connected via Bluetooth and run our developed applications presenting the different stimuli and logging the data. The developed smartphone application serves as an input controller as well as a secondary display. For the baseline, only the HoloLens with the build-in mid-air hand gesture support was facilitated. Our experiment was conducted in a room with controlled light conditions for consistent visibility of holograms and a free interaction area of approximate 3x3 m.

8.4.3 Procedure

After welcoming the participants, we asked them to sign the consent form. We explained the course of the study, all devices, and the interaction concepts to the participants. Afterward, we adjusted the AR glasses to the participant's head and ensured that the participant can comfortably perceive the entire display area. In the last preparation step, we ask the participants to start our application which guides the participant through the study. The application starts with a tutorial and aids the participant in getting used to the different interaction concepts. During the tutorial, participants could freely practice the first input modality until they understand them and feel comfortable. Specific questions regarding the input and study were answers, and the main task was explained. Participants were requested to finish the tasks as fast and as accurate as possible. Then participants started with the object manipulation task. After manipulating all 12 objects (three sets of four objects), they had to fill out the NASA-TLX [95] questionnaire on a dedicated laptop. To minimize fatigue, participants could take a break at this point since all tasks were performed in a standing position. Eventually, they continued with the second task using the same input modality. Again, a tutorial was presented introducing the new task and allow the participant to practice and familiarize themselves with the different function. Subsequent to accomplishing the new tasks, participants filled out the NASA-TLX questionnaire. Finally, they repeated both of the experiments with the other input modality.

The input modality and the tasks were presented in a counterbalanced order using a full Latin square to prevent any sequence or learning effects. Throughout the study, we logged all interactions with the system for subsequent offline analyses. After successfully finishing both experiments, we asked for comments about the user experience and which input modality they ultimately prefer. Participants completed the study including the debriefing within 40 to 55 minutes.

8.4.4 Experiment 1: Object manipulation

For the first experiment, we designed an object manipulation task in which participants had to select, translate, rotate, and scale different colored cubes. Four cubes with a side length of 50 cm were placed in front of the participant. After the selection of a cube, a white line guides the participant to the ghost representation of the cube representing the target. Participants were asked to perform the manipulations using the given input modality as precise and fast as possible. Both conditions are visualized in Figure 8.4. Task complexity was increased through three sets of four cubes. The first set only includes translation to achieve the target transformation. The second set of four cubes additionally includes rotation. Lastly, participants had to translate, rotate, and scale each of the cubes respectively. We measured the accumulated TCT starting from the first modification till the last modification of each cube since we were only interested in the object manipulation performance. Further, we recorded the accuracy measured by the euclidean distance, absolute rotation, and scale offset. Finally, we assessed the cognitive workload through the RTLX. In the first experiment, we recorded a total of 578 object manipulations.

8.4.5 Results 1: Object Manipulation

We analysed the TCT, and the accuracy with respect to position, rotation and scale deviation. We further assessed an analysed the cognitive workload using the RTLX questionnaire. For statistical comparison we used a repeated measures t-test. All significance levels of both experiments are at $\alpha = .05$. The results of the first experiment are graphically depicted in Figure 8.5.

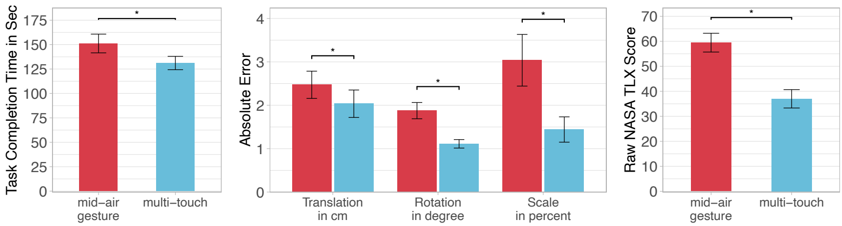


Figure 8.5: Mean values for TCT, relative error, and RTLX Score for each condition of the first experiment. Error bars show standard error of the mean (SE). Asterisk indicate statistically significant differences between conditions.

Task Completion Time

The aggregated TCT for all manipulations of *mid-air gesture* input ($M = 453.47, SE = 34.84$) was significantly higher than *multi-touch* input ($M = 393.53, SE = 25.01$), $t(23) = 2.341, p = .028, r = 0.438$. The effect size estimate indicates that the difference in TCT created by the input modality was a medium effect. To understand the strength of the smartphone input we subsequently analysed each of the three task complexities individually using t-tests. We found significant differences between *mid-air gesture* input ($M = 173.47, SE = 15.74$) and *smartphone input* ($M = 135.42, SE = 10.52$), $t(23) = 2.799, p = .010, r = 0.571$. for the medium complex task including translation and rotation but no significant differences for the others tasks (all $p > .05$).

Accuracy

For statistical analysis we split the accuracy measure into translation, rotation, and scaling error. The differences between the conditions of all metrics were not normally distributed, as assessed by the Shapiro-Wilk test (all $p < .001$), therefore, we analyzed the data using Wilcoxon signed-rank tests. The accuracy

Figure 8.6: The conditions of the second experiment. Participants are either controlling the GUI via mid-air gestures (left), touch gestures on the smartphone or the GUI is displayed on the smartphone itself (center/right) while the task and elements are displayed in AR.

using *multi-touch* input was significantly improved compared to *mid-air gesture* input in all subcategories (all $p < .001$). We measured the largest effect size for the scaling factor ($Z = -3.848662, r = 0.467$), followed by the translation error ($Z = -5.32343, r = 0.365$) and rotation error ($Z = -4.114372, r = 0.350$).

Cognitive Workload

We used the RTLX as subjective, multidimensional assessment tool to rate the perceived workload of each input condition. The RTLX scores were not normally distributed ($p = .010$). The Wilcoxon signed-rank test showed that using *mid-ari gesture* input ($M = 59.46, SE = 3.77$) elicit a statistically significant change in the perceived workload in comparison to our *multi-touch* input ($M = 37.00, SE = 3.68$), $Z = -4.273, p < 0.001$). Indeed, the effect size $r = 0.617$ suggests a large practical significance.

We summarize that the utilized input modality has a significant effect on the object manipulation performance measured using the relative error concerning translation, rotation, and scale. Further, the task completion time can, dependent on object manipulation complexity, significantly be reduced. The RTLX score was the lowest when using the multi-touch display of the smartphone.

8.4.6 Experiment 2: GUI Interaction

For the second experiment, we designed a menu with a list of modification options for a virtual 3D object. The participant had to interact with the menu and change the settings according to the presented information. Therefore, the target state was textual shown next to the object to modify. The menu comprises drop-down items, radio-buttons, slider, and regular buttons.

In this experiment, INPUT has three levels. Additional to *mid-air gesture* and *multi-touch* we introduce *multi-touch display*. In that condition, the menu was directly presented on the smartphones' display, and the participant could directly interact via touch. In the other conditions, the menu was anchored in space, facing the user and had to be operated via mid-air or multi-touch gestures. Both visualizations are represented in Figure 8.6. In total, participants had to complete 36 tasks, 12 in each condition.

Again, we measured TCT and RTLX. Since the correct input for all parameters was required to complete the task, we did not measure an error rate. TCT was measured from first to last input for each sub-task individually. In total, we recorded 864 menu interaction.

8.4.7 Results 2: GUI Interaction

We statistically compared the TCT, and the RTLX, using a one-way repeated measures analysis of variance (ANOVA). The results of the second experiment are displayed in Figure 8.7.

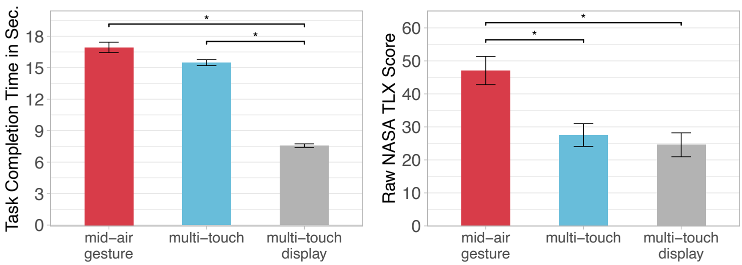


Figure 8.7: Mean values for the TCT and RTLX Score for all three conditions of the second experiment. Error bars show standard error of the mean (SE). Asterisk indicate statistically significant differences between conditions.

Task Completion Time

We analysed the average time the participants needed to adjust the 3D model according to the specifications without any error. We measured the highest TCT in the mid-air gesture condition ($M = 16.93, SD = 8.43$). Time decreased in the multi-touch input condition ($M = 15.48, SD = 4.81$), and was lowest in the multi-touch display condition ($M = 7.57, SD = 2.85$).

A repeated measures ANOVA with a Greenhouse-Geisser correction revealed a statistically significant difference between TCT measurements, $F(1.51, 428.06) = 240.25, p < .001$, partial $\eta^2 = .46$. Bonferroni-adjusted post-hoc analysis revealed a significant difference ($p < .001$) in completion times of *mid-air gesture* input and *multi-touch display* input (9.38, 95%-CI[8.17,10.59]) and *multi-touch display* input and *multi-touch* input (7.89, 95%-CI[7.16,8.24]).

Cognitive Load

We conducted further analyses to assess how the participants subjectively perceived the workload while interacting in AR. Using the smartphones' display for input led to a lower perceived workload ($M = 24.58, SD = 17.75$) compared to the *multi-touch* input ($M = 27.54, SD = 16.92$) and *mid-air gesture* input ($M = 47.08, SD = 21.00$).

A repeated measures ANOVA revealed a significant difference between the input modalities for the GUI interaction task, $F(2, 46) = 60.56, p < .001$ with a large estimated effect size (partial $\eta^2 = .73$). Bonferroni-adjusted post-hoc analysis revealed a significant difference in perceived workload between *mid-air gesture* input (both $p < .001$) and *multi-touch* (19.54, $95\% - \mathrm{CI}[13.43, 25.66]$) as well as *multi-touch display* input (22.50, $95\% - \mathrm{CI}[16.43, 28.56]$).

We encapsulate that using the smartphone as a controller significantly reduced the perceived workload independent of the utilized input method. However, only the utilization of both capabilities, the multi-touch screen as input, and the display as output is significantly reducing the required interaction time.

Overall Preference

In the overall ranking, 16 out of 24 participants preferred the *multi-touch* input for object manipulation. Participants clarify that interactions felt more accurate and less frustrating, especially during rotation and scaling of objects. Only one participant was in favor of *mid-air gesture* input, while seven participants had no explicit preference. For the GUI interaction experiment overwhelming 22 participants preferred the *multi-touch display* interaction. Exact modifications of the GUI slider subjectively result in the most exasperation using *mid-air gestures*. Hence, none preferred this input modality for GUI interactions.

8.5 Discussion

Our results show that with the smartphone as a tangible multi-touch input controller, immersed users can modify the virtual environment significantly faster compared to state-of-the-art mid-air gesture input. Due to tangible direct multi-touch interaction, the perceived cognitive workload is significantly reduced. Our approach benefits in particular from using multi-touch interactions that are not overloaded. Results indicate that using the two-finger movement for rotation in parallel to two-finger pinching ads level of complexity that slowed down the overall interaction speed. Following, we examine several factors that are potentially responsible for the overall advances in interaction.

In all conditions, the systems provide visual and auditive feedback. However, only in the smartphone-based conditions, direct kinesthetic and tactile feedback is given. Users benefit in general from the well-known device and can perform gestures with ease. Contrary, there is no haptic feedback while performing mid-air gestures and users are less trained. Added physical fatigue during long-lasting interaction further impairs interaction. This gorilla arm effect is also confirmed by several works evaluating mid-air interaction [29, 106].

Interestingly we could not observe any adverse effects for the multi-touch display condition. Usually switching attention to a secondary display causes overhead for the user [204]. Participants mentioned that switching and focusing between the AR display and the smartphone display feel unnatural in the first place but became quickly intuitive. Data supports this through the lowest perceived demand. We assume that the potential overhead is minimized since virtual information was approximately displayed at an optical distance of 2.0 m away from the user. Thus, perceptual conflicts, as space misperception, or in particular the vergence-accommodation conflict was minimized.

Setting users' performance, workload, and accuracy into contrast, our results suggest to deploy controller-based multi-touch interaction for simple and complex modifications and interaction in AR environments. We assume that users benefit through a relaxed body posture in particular during extended interaction sessions. For future AR systems that require modification of the virtual elements, our findings implicate that the support of a smartphone comprising a multi-touch screen offers the best results.

8.5.1 Limitations

Currently, our smartphone controller approach supports only a subset of interaction necessary to fully interact in generic AR applications. However, we demonstrated the feasibility and potential of multi-touch and a secondary touch display in AR environments. Extending our approach with already existing positional tracking solutions [34, 178] would enable fluent, precise, and convenient input in mobile or spontaneous interaction scenarios. To further increase the interaction space, additional sensors like accelerometer, gyroscope, or proximity could be incorporated.

Our approach uses a well developed high precision touch interface for interaction. In contrast, mobile mid-air gesture tracking is more complex and less accurate through different factors as spacial and temporal camera resolution. Our participants are used to smartphone interaction and gestures on multi-touch displays. In everyday interaction with computing systems, mid-air gestures are less common. Hence, our results may not be generalizable for users with extensive hand gesture experience that may be less prone to fatigue.

In our study scenario, the relationship between the secondary handheld display and augmented space is evident. In more complex scenarios, methods need to be implemented to keep it comprehensible to the user, when and how information is displayed on the secondary screen.

8.6 Conclusion

In this chapter, we have shown the potential of transforming unmodified smartphones into ubiquitous controllers and extensions for mobile MREs. Our approach comprises a smartphone paired with an MR headset. Thus users can effortlessly interact with the virtual content via multi-touch gestures, and we can further extend the interaction space through a handheld high-resolution touchscreen.

With two experiments, we investigated how to interact with a digital artifact in mobile mixed reality environments. We researched the interaction performance with the smartphone as a controller in contrast to mid-air gestures as a baseline. In both experiments, we found significant differences in task completion time, accuracy, and workload. Users benefited from using the smartphone as an input controller for fine-grained manipulations, the haptic feedback, and, ultimately, the tangible display. Our work contributes to an empirical study showcasing the viability of smartphones as a multimodal and ubiquitous input device in mobile AR experiences.

Chapter 9

A Taxonomy for Mixed Reality Experiences

Since the very beginning of human-computer interaction, the possibilities of input and output modalities have expanded tremendously. The continually increasing capabilities of sensing technologies and output modalities for AR and VR systems create numerous new viable use-cases in which these systems could augment the environment and entertain, train, or support users.

A new taxonomy is required to overcome this separation of AR and VR experiences and group them under the overarching theme of MR experiences. Our taxonomic space is a semantic structure to analyze and classify current and future MR experiences. The overall goal of this taxonomy is to provide a framework for classification that is generic enough to structure new emerging systems and experiences in the domain of MR, yet simple enough to apply. We propose the following by adapting from the technical taxonomy of MR visual displays [175] and Buxton's taxonomy for input devices [36], and incorporate results and knowledge gained from the development of the presented MR research-probes.

In this chapter, we first discuss related taxonomies that inform our design. We then present the scheme of the taxonomy we have developed for classifying Mixed Reality experiences. Finally, we show how to classify MR experiences in the visualization of our taxonomy.

9.1 Related Taxonomies

Researchers have proposed several taxonomies to classify augmented, virtual, and mixed reality systems. They also include envisioned and developed interaction techniques [191]. The currently existing taxonomies for augmented or virtual reality experiences can be divided into technical and conceptual taxonomies.

With a *conceptual taxonomy*, Jacob et al. [118] proposed a unifying concept to classify emerging interaction methods beyond the notion of desktop computing. With their descriptive framework, interaction designs can be analyzed and compared based on themes as physical, human body, environment, and social setting. Dubois et al. [54] followed an interaction-centered approach to classify AR and VR systems. This generic framework models the interaction between object, person, or adapter and the system, and allows the classification of existing AR systems. Later Dubois extended this taxonomy and presented a notation to describe mobile systems that combine physical and digital entities and the physical relationship [53]. In contrast, Mackay [163] used a simplified approach purely based on the design of the augmentation, with the user, the physical object itself, or the environment instrumented to provide augmentations. Nevertheless, this taxonomy cannot be used to discriminate between different input and output modalities. Focusing on the human sensory system, Lindeman and Noma [153] proposed a classification framework based on where the fusion of real and digital stimuli happens. They presented the pathways for stimuli and three possible locations for fusion: the environment, the sensory subsystem, or the computer. Each system can be classified by the human sense it is stimulating and the location of mixing. The proposed taxonomy allows a detailed classification based on the delivered output to the human sensory system. However, any input modalities are not represented by this taxonomy. Addressing the changes in input modalities, Benford et al. [23] described a design framework that allows the classification based on the movement of the user. The framework distinguishes between captured, desired, and expected movements of the users, yet is not specialized for MR systems.

Based on a literature review and interviews, Speicher et al. [232] proposed a conceptual framework comprising seven dimensions to distinguish MR experiences unambiguously. The dimensions are the number of environments, number of users, level of immersion, level of virtuality, degree of interaction, and the two lower-level dimensions input and output. Each dimension has flexible characteristics allowing for precise specification of each dimension.

Milgram et al. [175] proposed a *technical taxonomy* known as the Reality-Virtuality (RV) continuum, which describes a dynamic transition from the real environment to the virtual one. However, one-dimensional taxonomies are not suited to emphasize the different factors of MR systems. Hence, Milgram presents a three-dimensional taxonomic framework based on the RV continuum [176], where mixed reality systems are structured depending on display and technology performance rank. The three-dimensional taxonomic framework comprises the extent of world knowledge, reproduction fidelity, and extent of presence metaphor describing the system's knowledge about the environment, the quality of the stimuli presented to the user, and the users' sense of being present in the augmented or virtual world. Also, Newman et al. [188] used the RV continuum as the foundation for their taxonomy. The authors combined it with Weiser's vision of ubiquitous computing and created another one-dimensional continuum ranging from ubiquitous to monolithic computing. Placing both at right angles creates the two-dimensional Milgram-Weiser continuum and allows for the classification of mixed reality and ubiquitous computing applications. Even more precise accuracy of discrimination is realized by the new topology of Normand et al. [191]. They describe four-axis as a base to classify AR applications. These axes include the DOF of the supported tracking technology, the augmentation type that discerns optical-see-through and video-see-through displays, and the temporal base that includes, past, current, or future content. The last axis incorporates the rendering modalities based on the human senses.

Currently, there is no universal definition of mixed reality [232]. The same holds for taxonomies trying to classify different MR experiences. Many of the existing taxonomies distinguish between VR and AR and do not allow them to categorize hybrid systems. Besides, it is complex to visualize three or more dimensions graphically [191]. Hence, we aim for a two-dimensional grid-based representation of our taxonomy.

9.2 Scheme

The leading factor for our taxonomy for the classification of MR experiences is the individual human. We formulate our two-dimensional taxonomy based on the human sensory system and human capabilities of articulation. The first dimension is ground on the observed input events of the sensors of the experience. The second dimension emanates from the output stimuli the experience is presenting to the user. Both dimensions have their specific characterizations. Together they represent the capabilities and features of the MR experience on a conceptual level. In the following, we discuss each of our taxonomies' dimensions that classify existing and future MR experiences. In Figure 9.1, we visualize the matrix representation of our taxonomy.

9.2.1 Output

The first and fundamental dimension for our classification of MR experiences is the presented stimuli. Even though current MR experiences are mostly visual, we discriminate MREs based on the full human sensory system. We can split this dimension into five categories: visual, auditory, haptic, gustatory, and olfactory (cf. Section 2.4). Each of these categories is classified according to which degree of virtuality the MRE is stimulating the human sensory system based on the Milgram RV continuum [175]. Unlike the Milgram RV continuum, the degree of virtuality does describe the origin of the stimulus rather than the environment. Original stimuli from the real world are represented at one end of the spectrum and virtual, synthetically generated stimuli at the opposite side. Our taxonomy can map the degree of virtuality for each stimulus received by the human sensory system and is not limited to visual stimuli transmitted via conventional video displays. Hence, this dimension alone offers already a high discriminative power. To distinguish between real and virtual stimuli, we consider the following: A simulation always synthesizes the experience of a virtual stimulus. In our taxonomy, we consider real stimuli as a natural sensation without artificial modifications. The original object always emits a real stimulus. In this sense, oppressing a signal is also considered as modification, e.g. noise cancellation.

9.2.2 Input

Even though the output dimension alone already offers a high discriminative power, it does not differentiate experiences regarding the input opportunities. The interaction modalities of an MRE are a fundamental aspect to consider. Hence, the second core dimension is any input sensed by the MRE. The first level of discrimination is the kind of input performed by the user. In our taxonomy, we split users' interaction into embodied input and interaction performed with a physical item (controller). Any input involving the human body to interact with the MR experience is considered as embodied. To further separate embodied input, we present the six subclasses: head pose, body pose, location, gesture, gaze, and voice. We include these six classes as they cover what humans are capable of articulating consciously. Each of the different embodied input modalities for MR systems is described in detail in Section 2.4.2. All other input that requires a physical item acting as a controller is classified in the remaining group of physical controller input. To further categorize the different potential of the input modalities, we divide them according to the supported degree of freedom similar to Buxton's tableau of continuous input devices [36]. For simplicity, we extract the division into position, motion, and pressure, but we increase the degrees of freedom. Hence, the sum of any combination of input is registered in the corresponding group. We abstract here and consider the degrees of freedom as an open-ended scale, starting from one degree for simple slides to two degrees, e.g. gaze on a display or location on a map to full body pose or flexible voice input.

As a final class to distinguish the kind of input, we split system input into explicit and implicit interaction. Implicit interaction comprises the user's natural interaction with the computing systems that autonomously react to context and activity. In contrast, explicit input embraces all active and conscious input or task solving efforts into the computing system.

9.3 Taxonomy Visualization

In- and output technology for MR systems is changing rapidly. However, the human sensory system is a constant parameter in this dynamic field. Through the course of the thesis, we present six different MR research probes that are discussed in detail in the corresponding chapters (cf. Table 1.2 for an overview). To put them retrospectively in the broader field of MR research, we classify the probes in the visual representation of the previously outlined taxonomy that is presented in Figure 9.1.

Two probes extend the perception of reality. The quadcopter mounted projector visually augments the environment with navigational cues. Based on the users' location, different spatial information is displayed in the environment. The augmented physics lab experiment combines real physical probes with real-time data visualization on a smartphone screen. Implicit and explicit input is sensed through the smartphones' orientation and touch input.

With the two virtuality enhancing probes, we foster visual, auditory, and haptic output. Two probes enable typing in (mobile) VR. Both output virtual visual output, while the auditory and haptic output are still delivered from a physical

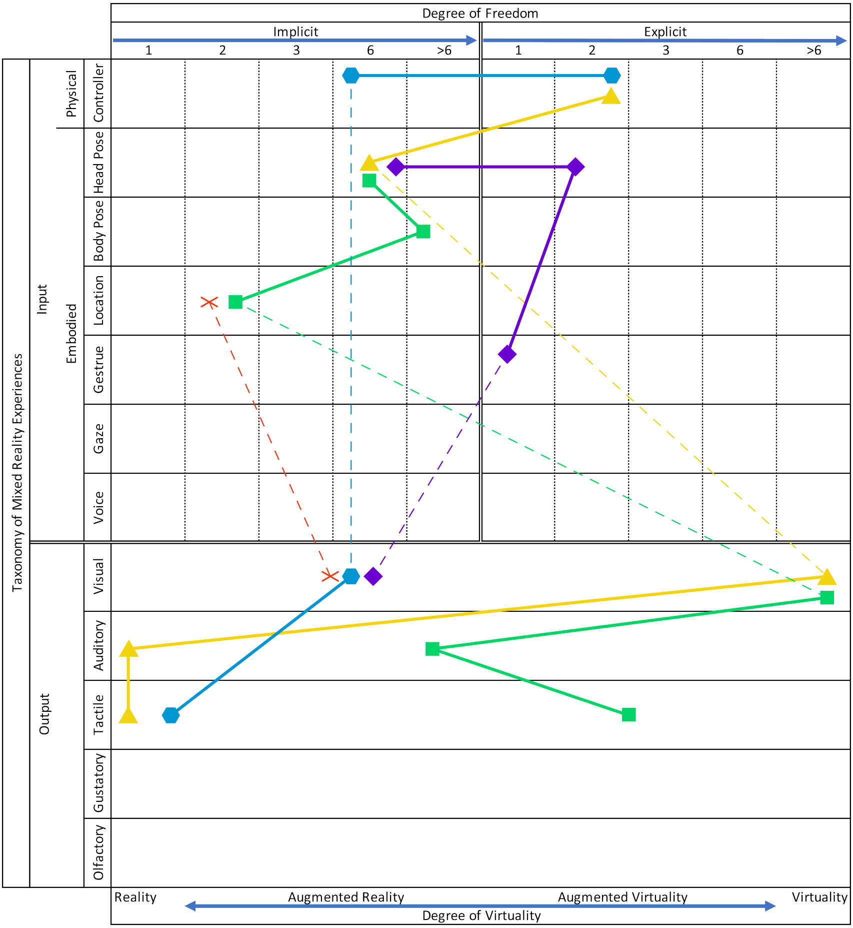


Figure 9.1: Visual representation of the taxonomy of MREs including the six research probes (color-coded) presented throughout this thesis. Enhancing Reality: projected in-situ navigation cues (red cross) and thermal experiment (blue hexagon). Enhancing Virtuality: Typing in (mobile) VR (yellow triangle) and *TactileDrone* (green square). Interaction in MR: *SmARtphone Controller* (violet rhombus).

keyboard. In contrast, the *TactileDrone* research probe outputs augmented and virtual visual, haptic, and auditory stimuli. Last, the *SmARtphone Controller* fuses different visual displays while utilizing a known tangible controller. Input is exclusively provided through the repurposed smartphone.

9.4 Conclusion

In this chapter, we presented a new taxonomy for classifying MREs. We derived the taxonomy by adapting and fusing Milgram's taxonomy for the RV continuum [175] and Buxton's relevant characteristics for input devices [36]. The main objective of our taxonomy is to overcome the separation of AR and VR experiences and group MR experiences with their unique properties and importance. Our taxonomy can help classify new trends in the evolution of MR experiences or identifying novel and underexposed combinations of input and output modalities, thus enhancing the interaction opportunities in MR experiences.

Chapter 10

Design Recommendations

In this thesis's broader context, we developed numerous prototypes and research probes, conducted studies, executed focus groups, and held various workshops to examine different scenarios and user groups. All concepts, applications, prototypes, and research probes are informing the design of Mixed Reality Experiences (MREs). In this chapter, we provide overarching design recommendations that tackle the input, output, and technological challenges that arose during the development of MREs. Although we draw the recommendations from content-specific applications, we seek generalisability and urge their consideration when developing future MR systems. We present design recommendations regarding input and output modalities as well as technology-centered design recommendations.

10.1 User-Centered Design Recommendations

The latest advances in research and technology allow for novel ways to interact and present digital content. Established guidelines targeting conventional computing interfaces [227] cannot cover the flexibility and unique features of MRE. However, MR best-practice-guides are not sufficiently evolved to lead developers [208]. On a conceptual level, it is essential that regardless of the utilized output modalities, the simulated stimulus blends seamlessly into the real environment. Visuals, sounds, or odors need to integrate as believable unified components of the experience. In this section, we extend these universally applicable guidelines and highlight specific design recommendations regarding the input and output modalities of MR systems.

Consider the Entire Environment

When designing for MR, it is essential to understand and respect the entire environment and the immediate surrounding of the user with all characterizing aspects. In its essence, the MR application's environment has a significant impact on the practicality of explicit and implicit interaction concepts and all output modalities.

Social acceptability and the practicalities of operation of the different modalities in shared spaces are issues still to be resolved. For example, voice input may not reliably work in environments with plenty of ambient noise. Further, social acceptability is a crucial criterion when designing input modalities used in public environments. Direct mid-air gestures may work reliably and intuitively, but could be socially inappropriate or create a feeling of indisposition. On the other hand, presenting digital information in MRE could cause information overload (cf. Chapter 3). In particular, visual clutter and highly immersive experiences could potentially create dangers to the user by overlaying hazards in the physical world, so it is crucial that virtual information should always be presented in context. Lastly, the environment has a significant influence on the feasibility of the technology (cf. Chapter 4), for example, projectors do not perform well in bright environments.

We recommend considering the entire environment in each design decision. The flexibility and possibilities of MR systems can be overwhelming, but experiences that take the context into account can significantly increase user engagement and satisfaction.

Prevent Interaction Causing Fatigue

When analyzing the developed taxonomy for MRE several explicit and implicit input possibilities were identified. In each category, a large variety of input concepts can be implemented, so it is essential to recognize the ergonomics and design interaction concepts accordingly. To sustain high throughput text input for prolonged periods, physical keyboards (cf. Chapter 6), as a known interface, will outperform other modes of input.

Unfamiliar or varying actions can cause fatigue. Mid-air gestures, for example, are prone to causing fatigue if performed repeatedly. Pushing several buttons simultaneously on handheld controllers can be challenging for users with small hands. The same applies to holding MR devices or wearing heavy and unconformable HMDs. We recommend preventing or at least reducing actions that can create user fatigue.

Support Natural Interaction

Interacting with new computing systems like MR requires the user to learn different ways to interact with the interface. This change could cause frustration or risk of cognitive overload. To develop successful MREs, designers and developers should take advantage of the human sensory and ergonomic potential. If the physical environment and utilized hardware allow, MREs should be explored by implicit natural movement. For excellent usability and a seamless blend of virtual and physical interaction, we recommend integrating familiar concepts and supporting natural interaction known from the physical world. This includes but is not limited to interaction with interfaces we already know very well, like the keyboard (cf. Chapter 6) or smartphone (cf. Chapter 8), but also natural touch exploration.

Synchronicity before Modalities

It is still challenging to create a sense of presence in MR. However, carefully designed MR experiences that follow the notion of plausibility can create this sensation. Through specified technology, each of the human senses can be presented with synthesized information. When increasing the immersion by blending these different output modalities, a coherent and consistent stimulus is fundamental. Only when stimuli are presented synchronously, and no mismatches occur, will the sense of presence increase rather than disrupt (cf. Chapter 7).

When designing prototypes, technical limitations can cause different output modalities to be out of synchronization. These temporal or spatial offsets can be observed by the user and potentially destroy the sense of presence. In these cases, working with less, but synchronized stimulating outputs modalities can create a more compelling experience. However, the additional complexity when integrating additional outputs with potential offsets is more likely to be disorienting to the user. When designing MREs, synchronicity is preferable to the number of output modalities; Following the less is more approach can help to create MREs inducing a strong sense of presence.

Involve the User

Interactive Systems, and in particular MR systems, require developers and experienced designers to consider the users. Only in this way can pleasant usability, a high sensation of presence, and a plausible experience be created. On the one hand, this holds for the interaction concept: Dependent on a user's body shape, the interaction volume changes. On the other hand, users' attitudes can influence the representation of their avatars, which could own unique personal characteristics within the MRE. Knowing the users' preference allows designing highly customized MREs. However, if this detailed information is unavailable, we recommend designing for a flexible and gender-neutral avatar representation.

10.2 Technology-Centered Design Recommendations

The developed research probes also allowed us to supplement technology-centered design recommendations in addition to the user-centered design recommendations. Although technology is currently evolving rapidly and MR devices are becoming more powerful, capable, and comfortable, we emphasize the following technology-centered recommendations.

Instrument the Environment

Today, MR systems can, depending on the environmental limitation, be implemented and installed either as stationary setups in the environment or as wearable devices in a mobile context. The prototypes and research probes are based on different platforms realizing stationary [144], mobile [137], and hybrid [132] approaches to present visual, auditory, or haptic feedback in mixed realities.

Based on the developed MREs and conducted studies, we currently recommend to instrument the environment with technology as far as the scenario allows. In a stationary setting, equipping the environment ensures excellent comfort and flexibility for the user and greater reliability through minimizing the constraints which are valid for mobile devices. Stationary tracking or projection systems still outperform their mobile counterparts. Further, constraints like weight or power consumption are in stationary setups less relevant. By today, projection-based mixed reality has become the most mature technology that can create compelling stationary experiences. Also, it enables a more human-centric experience since flexible multi-user support can easily be realized.

However, we are confident that with the next iteration of wearable MR devices, these limitations will be tackled. Then, personalized and mobile MREs can be realized and offer similar comfort and extended flexibility.

Overcome Hardware Limitations

When developing MR applications, the selected hardware components have a significant impact on their quality, complexity, and feasibility. Each technology comes with a specific set of limitations and constraints. Optical see-through AR headsets have a limited FOV, VR headsets are often wired, vibrotactile haptic feedback though controllers limits expressiveness, and mobile devices have limited processing capabilities. It is essential to know and minimize certain limitations according to the use-case when designing MREs, and to plan studies.

We recommend overcoming the limitations prevalent in the hardware by designing around them. Specifically created shaders and geometries can compensate for limited processing capabilities. Limitations in the FOV or quality of the output modality can be counterbalanced by creative adaptations of the virtual content and environment (cf. Chapter 7). For qualitative studies in VR, we emphasize again to favor stationary setups that enable higher performance in all domains. Early circumvention of limitations can reduce the futile feedback regarding already known hardware limitations.

Preserve and Protect Privacy

Independent of the use-case, users, industry, and stakeholders are concerned about the multitude of sensors embedded into mixed reality devices. Though individual interviews (cf. Chapter 3) and field studies in classrooms (cf. Chapter 5) we identified considerable concerns regarding personal privacy. Although more and more context-understanding sensors like always-on microphones and cameras are gradually becoming part of the users' environment, there is still a pronounced skepticism against ubiquitous monitoring.

The deployed sensors and processed recordings should always offer the best trade-off between privacy and added value. Hence, decisions need to be made responsibly. When designing mixed reality systems, we recommend relying on a minimal viable selection of sensors and to precisely communicate sensing, recording, and processing.

Chapter 11

Conclusion

This thesis explores the impact of input and output modalities on Mixed Reality Experiences (MREs). To examine the potential and to understand the requirements for extended Mixed Reality (MR), we developed several research probes following a user-centered design process. The goal is to support researchers, developers, and designers in creating excellent MREs by overcoming the current challenges and providing guidance through the manifold design decisions on input and output modalities.

In this final chapter, we summarize our research contributions and answer the research questions that were addressed throughout the thesis. We close with a perspective towards future research directions and a final statement about their value for stakeholders of mixed realities.

11.1 Summary of Research Contributions

The contributions of this thesis are situated in the field of mixed reality. First, we identified the potential of MREs in a user-centered design process and identified existing challenges for the widespread use of this novel technology. Second, we presented seven research probes developed to explain the impact of modalities and interaction techniques on the user experience. Third, we derived design recommendations based on these and developed new probes and contributed insights to elevate the design of MREs. Lastly, we provided a new taxonomy for

MREs, setting our research probes into perspective, showcasing the versatility of MREs, and lay out a classification scheme for current and future MREs.

11.1.1 Research Probes

Throughout this thesis, we presented seven research probes to investigate the impact of input and output modalities on MR systems. The diversity of the presented research probes shows the tremendous possibilities of how to envision, design, and implement MREs. We showcased the different approaches for realizing MREs based on varying conditions and environments, such as handheld and wearable experiences in stationary or mobile settings. We highlighted informed design decisions including whether to instrument the environment or the user, and taking into account the underlying interaction concepts of implicit and explicit inputs.

Besides being instrumental in answering the outlined research questions, the provided research probes contribute to the following aspects: On a conceptual level, we describe how to seamlessly superimpose visual information in-situ in the physical world for mobile and stationary experiences. On a methodology level, we created a system for offline AR annotations, and on a technological level, we created prototypes of real-time presentation of information that amplify the human sensory system. We present the architecture and facilitated frameworks to realize MREs on a technology-independent abstraction level using the example of wearable HMDs and smartphones.

With the developed probes, we enabled effective and dynamic interaction within MREs. Two probes, for example, enabled users to efficiently type in Virtual Reality (VR) (cf. Chapter 6). Here, we contribute two solutions for stationary and mobile interaction in MR:an apparatus for text input on a traditional keyboard while being immersed in MR, and a robust and high precision finger tracking solution for VR. With our second portable implementation, based on a smartphone, we presented additional insights on the quality, strength, and shortcomings of low fidelity solutions.

Finally, our research probes contribute a detailed elaboration on how to integrate and blend tangibles within an MRE. The presented probes utilize tangibles and physical objects to either enhance the interaction modality or, in more immersive experiences, to increase the sense of presence through the additional haptic component.

11.1.2 Design Recommendations

We presented eight design recommendations derived from the experience gained through the design, development, and evaluation of the research probes. The design recommendations reflect the current state of technology and support the development of meaningful MREs. We grouped the recommendations into technology-centered and user-centered recommendations to emphasize the human factor when developing technology-driven MREs. The user-centered recommendations serve as a guide for choosing appropriate input and output modalities for satisfactory interaction and presentation of stimuli. Further, the recommendations address social interests regarding privacy and social norms. With the technology-centered design recommendations, we highlight the diversity and entailing challenges of environments where users can explore MREs, and present guidelines to choose the most suitable technology and implementation.

11.1.3 Taxonomy of Mixed Reality Experiences

We contributed with a new taxonomy for classifying MREs by adapting and fusing Milgram's taxonomy for the Reality-Virtuality (RV) continuum and Buxton's relevant characteristics for input devices. We avoid the separation of AR and VR experiences and group MREs with their unique characterizing input and output attributes into one unified space. The taxonomy can be implemented to categorize MREs based on the human sensory system and human capabilities of articulation. Thus, it can help identify new trends in the evolution of MREs as well as recognize novel and underexposed combinations of input and output modalities. To showcase the taxonomy's principle and adaptability, we classified the stated research probes within our new taxonomy.

11.1.4 Limitations

The area of MR is substantial, and contemporary technology offers high versatility and diversity. It is essential to utilize this technology concisely and sensibly. Each of the probes we realized and each study we carried out has a particular set of limitations. These are considered in detail in each chapter.

Moreover, the research probes introduced in this thesis are composed of current off-the-shelf components. Hence, the probes fundamentally differ from prospective products designed for the consumer market. During the studies, certain recurring hardware-related limitations like apparatus weight, Field of View (FOV) or comfort became predominant. These could have a potential effect on the results. Nevertheless, we expect many of these limitations to be solved soon through ongoing development and advances in technology. We believe that the obtained knowledge and the generalized outcomes of each study will be valid for prospective MREs and remain valid through upcoming technological changes.

11.2 Research Questions

In the first chapter of this thesis we elaborated eight research questions assigned to *User-Centered*, *Mixed Reality Experience-Centered*, or *Interaction-Centered* focused research. During the research presented in this thesis, we acquired the knowledge to answer these questions. In the following, we present the condensed answers that are explored in more detail in the respective chapters.

11.2.1 User-Centered

With a user-centered approach, we first explored the potential of mixed reality experiences in domestic environments. Through an online survey and a technology probe, we gained in-depth knowledge to answer the following research questions (cf. Chapter 3).

What are the users' attitudes towards the introduction of mixed reality at home?

The workshops, interviews, and online surveys showed that users are generally positive towards MREs as a supportive part of their everyday lives. Experiences are expected to be an integral part and to be deployed comprehensively for a variety of basic tasks. Nevertheless there was some hesitation or underlying concern expressed about security and about the social considerations to do with how it would be perceived when being used in public. For a successful introduction of MR in a domestic environment, designers, developers, and stakeholders must take the users' concerns regarding privacy and social context into account.

What are the usage scenarios that are most promising for domestic mixed reality applications?

We showed that users are eager to benefit from MREs across many locations at home. We identified MR as a way to provide contextualized assistance as being most promising. Further, users envisioned scenarios of a enhanced human sensory system to perceive the world in a new and amplified way, and also highlighted the potential for augmented remote presence to participate at social activities. Lastly, the opportunity to enhance household artifacts with additional features was considered favorably.

What are the constraints and opportunities for future systems?

During the development and evaluation of the research probes, we showed the enormous potential of future MREs. With this work, we recognized the potential, in particular in learning and domestic environments. However, before social acceptance can rise and MR can become a success, today's hardware limitations need to be tackled, and existing privacy concerns need to be addressed. Thus, MR devices will accompany smartphones before slowly replacing them.

11.2.2 Mixed Reality Experience-Centered

As a result of the envisioned, developed, and evaluated research probes, we were able to derive the design guidelines and taxonomy presented earlier. Further, we built the research probes to gain knowledge to answer the following research questions.

How to enhance reality by blending information in the real world?

We built and evaluated research probes to analyze the effects of blending virtual and synthesized information into the real world, thus extending different output modalities including projection (cf. Chapter 4), handheld displays (cf. Chapter 5), or MR Head-Mounted Displays (HMDs) [135]. Our in-situ projection does not require the user to wear or hold the device; however, it reduces at the same time the user-control. In the studies, we showed that users value the natural availability of information and the avoided attention shifts. Further, our experiments confirmed that the in-situ presentation of information causes a significantly higher ability to observe the real-world. Enhancing the real world's perception by presenting additional real-time environmental information to the user and thus amplifying the human senses and cognition is considered valuable in learning scenarios.

How to enhance virtual reality by adding tangibility?

We advanced conventional audiovisual VR environments by adding haptic feedback. Levitating tangibles remedy the lack of haptic sensation in the virtual environment. Since the levitating tangibles are active and tracked, they can be used as input and output modality. In the conducted studies, users showed an increased sense of presence when tangibles facilitated haptic feedback. Further, we obtained detailed insights into the integration of tangibles as an input modality. With our research probe, we showed that tangibles utilized as input modalities in the form of keyboards (cf. Chapter 6) and touchscreens (cf. Chapter 8) can significantly improve the overall interaction performance.

How to provide flexible haptic input and output in mixed reality?

We presented a novel approach (cf. Chapter 7) by blending levitating tangibles, more specifically miniature quadcopters, within the mixed reality environment. This allowed us to introduce a haptic component into MRE and tackle the lack of haptic feedback.Using these versatile tangibles as haptic input and output enabled us to gain knowledge on the possibilities and drawbacks. Depending on the visual stimulus, the provided force of the haptic output is not always adequate. However, we presented generalizable approaches to circumvent occurring breaks in presence through distinct adaptations of the virtual environment.

11.2.3 Interaction-Centered

The analysis of existing interaction concepts enabled us to build on top of this and to envision novel interaction concepts for MR environments. Evaluating these allowed us to answer the following research questions.

How to interact with digital artifacts in mobile mixed reality?

The possibilities to interact in MR environments are manifold and content-specific. Based on two research projects, we studied the possibilities to enable intuitive input in virtual and augmented reality. The results of the conducted studies show that the efficient blending of well-established input modalities like the keyboard with the virtual world can offer a high-throughput input channel. Interweaving a user's smartphone as an input and output device into the MRE offers excellent potential for improved input and exciting new hybrid arrangements of information (cf. Chapter 8).

How to design a taxonomy for mixed reality experiences?

We created a taxonomy to chart existing and future MR experiences (cf. Chapter 9). The foundation of the taxonomy is the separation and categorization of input and output modalities. We exploit Milgram's Reality-Virtuality continuum [176] for classification of output on the one hand while adapting Buxton's taxonomy [36] of input devices for implicit and explicit input modalities in MR environments on the other. To showcase the utility of the constructed taxonomy, all research probes developed throughout this thesis are representatively incorporated into the taxonomic space.

11.3 Future Work

This thesis presents a substantial understanding of the impact of input and output modalities and provides a common ground for future research in the domain of MREs. During the development of the research probes and the elaboration of the taxonomy, additional challenges and exciting research spaces were identified. However, addressing these challenges is beyond the scope of this thesis. In the following section, we envision several small and large research projects and present future research directions.

11.3.1 The Mobile Virtual Office of the Future

With our research, we address the underlying challenge of text input while being immersed in a mixed reality environment. After successfully addressing the challenge of generic input into a computing system while the user is immersed in a virtual environment, MR working environments could be realized.These potentially lead to enhanced performance due to a controlled environment with less audiovisual noise and distractions. Moreover, future virtual offices can overcome the physical limitations of rectangular two-dimensional screen spaces. Future workspace can be arranged freely in three-dimensional space for new means of data visualization or collaboration.

Several research challenges need to be addressed to accomplish the vision of a VR office that sustains pleasant long-term working experiences. On the one hand, challenges like the lack of visual fidelity of the HMD must be resolved. Other technological challenges include weight, wearing comfort or portability, and mobile runtime.

On the other hand, there are interesting questions from a human-computer interaction perspective. The computing paradigm of windows, icons, menus, and pointers is still valid. It remains an open question how this translates to virtual realities. The visualization of the working environment, potential distractions, and threats existing in the physical world are still unclear. These questions unfold an interacting design space that requires further investigation. Working in virtual worlds using a HMD also creates a new paradigm for collaborating and communicating with coworkers. It is still actively being researched how communication and collaboration can efficiently be recreated in VR.

In a future mobile setting, additional factors, including motion sickness, situational awareness, and the masking of the physical world, contribute to the user experiences. Future research is required to determine the relevant answers and potential solutions to create pleasant and productive working environments.

11.3.2 Amplification of Senses

Within the context of this thesis, several prototypes extend or amplify different human senses. The sight was extended beyond the spectrum visible to the human eye [6, 7]. In another prototype, we amplified the temporal resolution of the human visual perception [139], and used in-situ projections to enhance the memorability of points of interest in the physical world (cf. Chapter 4.

This line of research can be continued in two directions. First, various other senses could be extended or amplified to support users in better understanding of the environment. Referring back to the developed taxonomy, underexplored senses can be identified at a glance. Second, amplified senses could be explored in other application domains. We identified that amplified senses can support students in the field of physics. This concept could be extended to either other fields of Science, Technology, Engineering and Mathematics (STEM) as well as entirely different areas to investigate real-world curiosities. In sightseeing scenarios, a tourist could use a tailored MRE to explore the temporal course of a monument. Jumping back in time while observing an interesting spot or area could lead to new and engaging journeys.

11.3.3 Towards Mixed Reality in the Wild

The majority of studies presented in this thesis were conducted in a controlled lab environment. This kind of evaluation offers high internal validity and allows for generalization. However, the next step is to conduct evaluations in the wild. The upcoming technological advances and the widespread availability of commercial MR hardware will allow for this necessary step. In the research context of this thesis, we already studied MR systems in less controlled classroom environments and through field studies. Still, the border between the lab and the wild needs to be shifted further towards evaluation in the real world.

We propose to deploy minimal viable MR systems in the wild for long-term evaluations. These evaluations will allow for novel insight into how users interact with MR systems in an uncontrolled and realistic scenario. Moreover, these studies further increase the ecologic validity and allow for new in-depth knowledge about the ultimate benefits of mixed reality experiences.

11.4 Concluding Remarks

With this thesis, we investigated the impact of input and output modalities on mixed reality experiences and discussed some of the fundamental challenges that arise when designing and developing MR systems. The involved technology is slowly leaving its infancy, and we are confident that upcoming generations of devices and experiences will have a significant impact on how we interact with, consume, and create digital information. This thesis makes a combined contribution to the fields of Human-Computer Interaction and Mixed Reality, thereby considering relevant aspects that advance MRE. As showcased in our presented taxonomy and outlined future directions, additional research is needed to draw a complete picture.

Today, technology companies are promoting commercial virtual reality headsets and augmented reality applications for smartphones. However, technology and developments have not yet progressed far enough to become truly ubiquitous. Ultimately, being ubiquitous, understanding the characteristics of input and output modalities regarding advanced mixed reality systems is crucial to create novel and engaging experiences.

BIBLIOGRAPHY

[1] Annexing reality: Enabling opportunistic use of everyday objects as tangible proxies in augmented reality. pages 1957–1967, 2016.

[2] Aryzon Cardboard. Aryzon: Augumented Reality powered by your smartphone, `https://www.aryzon.com/`, 2019.

[3] DHL Logistics, DHL Supply Chain Deploys Latest Version of Smart Glasses Worldwide, May 2019.

[4] ISO 9241-210:2019 Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems, 2019.

[5] M. H. I. Abdalla. Exploring implicit and explicit camera control for fixed camera third person view. *Bachelor thesis*, Munich, Aug. 2019.

[6] Y. Abdelrahman, P. Knierim, P. W. Woźniak, N. Henze, and A. Schmidt. See through the fire: evaluating the augmentation of visual perception of firefighters using depth and thermal cameras. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 693–696. ACM, 2017.

[7] Y. Abdelrahman, A. Schmidt, and P. Knierim. Snake View: Exploring Thermal Imaging as a Vision Extender in Mountains. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 1067–1071. ACM, 2017.

[8] Y. Abdelrahman, P. W. Woźniak, P. Knierim, N. Henze, and A. Schmidt. Exploration of Alternative Vision Modes Using Depth and Thermal Cameras. In *Proceedings of the 17th International Conference on Mobile*

and Ubiquitous Multimedia, pages 245–252, New York, NY, USA, 2018. Association for Computing Machinery.

[9] M. Abdullah, M. Kim, W. Hassan, Y. Kuroda, and S. Jeon. HapticDrone: An Encountered-Type Kinesthetic Haptic Interface with Controllable Force Feedback: Initial Example for 1D Haptic Feedback. In *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 115–117, New York, NY, USA, 2017. ACM.

[10] P. Abtahi, D. Y. Zhao, J. L. E, and J. A. Landay. Drone Near Me: Exploring Touch-Based Human-Drone Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), Sept. 2017.

[11] A. Achberger. Drohnen als Tangible User Interface zur Veranschaulichung von Experimenten. *Bachelor thesis*, Stuttgart, 2016.

[12] Apple Inc. ARKit 3, ARKit - Augmented Reality - Apple Developer, `https://developer.apple.com/augmented-reality/arkit/`.

[13] B. Araujo, R. Jota, V. Perumal, J. X. Yao, K. Singh, and D. Wigdor. Snake Charmer: Physically Enabling Virtual Objects. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 218–226. ACM, 2016.

[14] Arcadian Virtual Reality L.L.C. *Virtuality - Stand Up Units - SU2000*. arcadianvr.

[15] M. Avila, M. Funk, and N. Henze. DroneNavigator: Using Drones for Navigating Visually Impaired Persons. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 327–328, New York, NY, USA, 2015. Association for Computing Machinery.

[16] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1968–1979, New York, NY, USA, 2016. Association for Computing Machinery.

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

[18] H. Bai, G. A. Lee, and M. Billinghurst. Freeze View Touch and Finger Gesture Based Interaction Methods for Handheld Augmented Reality

Interfaces. In *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*, pages 126–131, New York, NY, USA, 2012. Association for Computing Machinery.

[19] R. Ballagas, J. Borchers, M. Rohs, and J. G. Sheridan. The Smart Phone: a Ubiquitous Input Device. *IEEE Pervasive Computing*, 5(1):70–77, Jan. 2006.

[20] J. Baus, R. Wasinger, I. Aslan, A. Krüger, A. Maier, and T. Schwartz. Auditory perceptible landmarks in mobile navigation. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pages 302–304. ACM, 2007.

[21] O. Baus and S. Bouchard. Exposure to an unpleasant odour increases the sense of Presence in virtual reality. *Virtual Reality*, 21(2):59–74, Oct. 2016.

[22] E. Beheshti, D. Kim, G. Ecanow, and M. S. Horn. Looking Inside the Wires: Understanding Museum Visitor Learning with an Augmented Circuit Exhibit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1583–1594, New York, NY, USA, 2017. ACM.

[23] S. Benford, B. Gaver, A. Boucher, B. Walker, S. Pennington, A. Schmidt, H. Gellersen, A. Steed, H. Schnädelbach, B. Koleva, R. Anastasi, C. Greenhalgh, T. Rodden, J. Green, A. Ghali, and T. Pridmore. Expected, sensed, and desired. *ACM Transactions on Computer-Human Interaction*, 12(1):3–30, Mar. 2005.

[24] H. Benko, C. Holz, M. Sinclair, and E. Ofek. NormalTouch and TextureTouch: High-fidelity 3D Haptic Shape Rendering on Handheld Virtual Reality Controllers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 717–728. ACM, 2016.

[25] M. Billinghurst and A. Duenser. Augmented Reality in the Classroom. *Computer*, 45(7):56–63, July 2012.

[26] M. Billinghurst and H. Kato. Collaborative Augmented Reality. *Communications of the ACM*, 45(7):64–70, July 2002.

[27] J. Blake and H. B. Gurocak. Haptic Glove with MR Brakes for Virtual Reality. *IEEE/ASME Transactions On Mechatronics*, 14(5):606–615, 2009.

[28] K. Boehner, J. Vertesi, P. Sengers, and P. Dourish. How HCI Interprets the Probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1077–1086, New York, NY, USA, 2007. ACM.

[29] S. Boring, M. Jurmu, and A. Butz. *Scroll, Tilt or Move It: Using Mobile Phones to Continuously Control Pointers on Large Public Displays*. ACM, New York, New York, USA, Nov. 2009.

[30] D. A. Bowman, C. J. Rhoton, and M. S. Pinho. Text Input Techniques for Immersive Virtual Environments: An Empirical Comparison. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(26):2154–2158, Nov. 2016.

[31] J. Brooke. SUS-a Quick and Dirty Usability Scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[32] F. H. Brynie. *Brain Sense*. The Science of the Senses and How We Process the World Around Us. American Management Association, Sept. 2009.

[33] R. Budhiraja, G. A. Lee, and M. Billinghurst. Interaction techniques for HMD-HHD hybrid AR systems. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 243–244, Oct. 2013.

[34] R. Budhiraja, G. A. Lee, and M. Billinghurst. Using a HHD with a HMD for mobile AR interaction. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–6, Oct. 2013.

[35] G. C. Burdea. Haptics Issues in Virtual Environments. In *Computer Graphics International, 2000. Proceedings*, pages 295–302. IEEE, 2000.

[36] W. Buxton. Lexical and Pragmatic Considerations of Input Structures. *ACM SIGGRAPH Computer Graphics*, 17(1):31–37, Jan. 1983.

[37] T. P. Caudell and D. W. Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, pages 659–669 vol.2, 1992.

[38] L.-P. Cheng, P. Lühne, P. Lopes, C. Sterz, and P. Baudisch. Haptic turk: a motion platform based on people. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 3463–3472. ACM, 2014.

[39] L.-P. Cheng, T. Roumen, H. Rantzsch, S. Köhler, P. Schmidt, R. Kovacs, J. Jasper, J. Kemper, and P. Baudisch. Turkdeck: Physical virtual reality based on people. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 417–426. ACM, 2015.

[40] G. R. Chris, C. Chiu, E. Kusternig, M. Kusternig, and H. Witzmann. iOrb - Unifying Command and 3D Input for Mobile Augmented Reality. In *Proceedings of the IEEE VR Workshop on New Directions in 3D User Interfaces*, pages 7–10.

[41] B. A. Cliath, O. D. F. Rialtais, T. S. Alliance, S. T. Laighean, F. Rialtais, and A. R. Post-tráchta. Learning for Life: White Paper on Adult Education. Technical report, 2000.

[42] A. Colley, L. Virtanen, P. Knierim, and J. Häkkilä. Investigating Drone Motion as Pedestrian Guidance. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, pages 143–150, New York, NY, USA, 2017. Association for Computing Machinery.

[43] G. Cortes, E. Marchand, G. Brincin, and A. Lécuyer. MoSART: Mobile Spatial Augmented Reality for 3D Interaction With Tangible Objects. *Frontiers in Robotics and AI*, 5:13, Aug. 2018.

[44] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, June 1992.

[45] A. Dancu, Z. Franjcic, and M. Fjeld. Smart flashlight: map navigation using a bike-mounted projector. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 3627–3630. ACM, ACM, 2014.

[46] T. Denning, Z. Dehlawi, and T. Kohno. In Situ with Bystanders of Augmented Reality Glasses: Perspectives on Recording and Privacy-mediating Technologies. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, pages 2377–2386, New York, NY, USA, 2014. ACM.

[47] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. I. Swan. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5:161, 2018.

[48] T. Dingler, P. Wahju, P. Knierim, and A. Schmidt. Memory Support through In-Home Display Deployment. *IT Professional*, 18(6):42–49, 2016.

[49] A. A. DiSessa. *Changing minds: Computers, learning, and literacy*. Mit Press, 2001.

[50] T. Djajadiningrat, P.-Y. Chao, S. Kim, M. Van Leengoed, and J. Raijmakers. Mime: An AR-based System Helping Patients to Test Their Blood at Home. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pages 347–359, New York, NY, USA, 2016. ACM.

[51] D. Dobbelstein, C. Winkler, G. Haas, and E. Rukzio. PocketThumb: a Wearable Dual-Sided Touch Interface for Cursor-based Control of Smart-Eyewear. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(2):9–17, June 2017.

[52] T. Döring, M. Brade, T. Fröhlich, M. Honauer, P. Knierim, and A. Reinschlüssel. 10. Workshop Be-greifbare Interaktion. *Mensch und Computer 2017-Workshopband: Spielend einfach interagieren*, 13:325, 2018.

[53] E. Dubois, P. Gray, and L. Nigay. ASUR++: A Design Notation for Mobile Mixed Systems. In *Human Computer Interaction with Mobile Devices*, pages 123–139, Heidelberg, 2002. Springer Berlin Heidelberg.

[54] E. Dubois, L. Nigay, J. Troccaz, O. Chavanon, and L. Carrat. Classification Space for Augmented Surgery. an Augmented Reality Case Study. In *13 International Conference on Human-Computer Interaction (INTERACT'99)*, page 353, Edinburgh, UK, 1999.

[55] M. Dunleavy, C. Dede, and R. Mitchell. Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of science Education and Technology*, 18(1):7–22, 2009.

[56] Epson. Epson. Moverio BT-200 Smart Glasses, Model: V11H560020, 2019.

[57] J. Falk, J. Redstr o m, and S. Bj o rk. Amplifying Reality. In *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, pages 274–280, Berlin, Heidelberg, 1999. Springer-Verlag.

[58] K. Fan, J. Huber, S. Nanayakkara, and M. Inami. SpiderVision: extending the human field of view for augmented awareness. In *Proceedings of the 5th Augmented Human International Conference*, page 49. ACM, 2014.

[59] S. Feiner, B. MacIntyre, M. Haupt, and E. Solomon. Windows on the World: 2D Windows for 3D Augmented Reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*,

pages 145–155, New York, NY, USA, 1993. Association for Computing Machinery.

[60] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal Technologies*, 1(4):208–217, Dec. 1997.

[61] A. M. Feit, D. Weir, and A. Oulasvirta. How We Type: Movement Strategies and Performance in Everyday Typing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4262–4273, New York, NY, USA, 2016. Association for Computing Machinery.

[62] K. J. Fernandes, V. Raja, and J. Eyre. Cybersphere - the fully immersive spherical projection system. *Commun. ACM*, 46(9):141, 2003.

[63] A. Field and G. Hole. *How to design and report experiments*. Sage, 2002.

[64] N. D. Finkelstein, W. K. Adams, C. J. Keller, P. B. Kohl, K. K. Perkins, N. S. Podolefsky, S. Reid, and R. LeMaster. When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Phys. Rev. ST Phys. Educ. Res.*, 2005.

[65] S. S. Fisher, M. McGreevy, J. Humphries, and W. Robinett. Virtual Environment Display System. In *Proceedings of the 1986 Workshop on Interactive 3D Graphics*, pages 77–87, New York, NY, USA, 1987. Association for Computing Machinery.

[66] S. Follmer, D. Leithinger, A. Olwal, A. Hogge, and H. Ishii. InFORM: Dynamic Physical Affordances and Constraints through Shape and Object Actuation. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, pages 417–426, New York, NY, USA, 2013. Association for Computing Machinery.

[67] M. Frainzl. Extending Everyday Objects with Augmented Reality. *Master thesis*, Munich, Nov. 2018.

[68] P. Fröhlich, A. Oulasvirta, M. Baldauf, and A. Nurminen. On the Move, Wirelessly Connected to the World. *Communications of the ACM*, 54(1):132–138, Jan. 2011.

[69] M. Funk. Human-drone Interaction: Let's Get Ready for Flying User Interfaces! *interactions*, 25(3):78–81, Apr. 2018.

[70] M. Funk, A. B a chler, L. B a chler, T. Kosch, T. Heidenreich, and A. Schmidt. Working with Augmented Reality?: A Long-Term Analysis of In-Situ Instructions at the Assembly Workplace. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 222–229, New York, NY, USA, 2017. ACM.

[71] M. Funk, T. Kosch, and A. Schmidt. Interactive Worker Assistance: Comparing the Effects of in-Situ Projection, Head-Mounted Displays, Tablet, and Paper Instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 934–939, New York, NY, USA, 2016. Association for Computing Machinery.

[72] M. Funk, S. Mayer, M. Nistor, and A. Schmidt. *Mobile In-Situ Pick-by-Vision: Order Picking Support using a Projector Helmet*. Order Picking Support using a Projector Helmet. ACM, New York, New York, USA, June 2016.

[73] B. Gaver, T. Dunne, and E. Pacenti. Design: Cultural Probes. *interactions*, 6(1):21–29, Jan. 1999.

[74] W. W. Gaver, A. Boucher, S. Pennington, and B. Walker. Cultural Probes and the Value of Uncertainty. *interactions*, 11(5):53–56, Sept. 2004.

[75] P. Gerth. Visualisierung der Prognostizierten Trajektorie eines Tischtennisballs. *Bachelor thesis*, Stuttgart, 2016.

[76] G. Gibbs and A. Jenkins. *Teaching large classes in higher education: How to maintain quality with reduced resources*. Routledge, London, 1992.

[77] A. Gomes, C. Rubens, S. Braley, and R. Vertegaal. BitDrones: Towards Using 3D Nanocopter Displays as Interactive Self-Levitating Programmable Matter. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 770–780, New York, NY, USA, 2016. Association for Computing Machinery.

[78] G. González, J. P. Molina, A. S. García, D. Martínez, and P. González. Evaluation of Text Input Techniques in Immersive Virtual Environments. In *New Trends on Human-Computer Interaction*, pages 1–10. Springer London, London, Feb. 2009.

[79] Google LLC. ARCore. Build new augmented reality experiences that seamlessly blend the digital and physical worlds, https://developers.google.com/ar/, 2020.

[80] N. Gorm and I. Shklovski. Participant Driven Photo Elicitation for Understanding Activity Tracking: Benefits and Limitations. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1350–1361, New York, NY, USA, 2017. Association for Computing Machinery.

[81] S. W. Greenwald, W. Corning, M. Funk, and P. Maes. Comparing Learning in Virtual Reality with Learning on a 2D Screen Using Electrostatics Activities. *J. UCS*, 24(2):220–245, 2018.

[82] J. Groschopp. Office to go: A Mobile Mixed Reality Office for the Pocket. *Master thesis*, Munich, Apr. 2019.

[83] T. A. Grosse-Puppendahl, J. Scott, S. E. Hodges, J. F. M. Helmes, S. A. Taylor, N. Y.-C. Chen, P. Knierim, G. Wood, J. Fromm, A. Schmidt, and others. Updating displays based on attention tracking data. US Patent 15058012, 2017.

[84] T. A. Grosse-Puppendahl, J. W. Scott, S. E. Hodges, J. F. M. Helmes, N. Y.-C. Chen, S. A. Taylor, P. Knierim, J. Fromm, G. Wood, A. Schmidt, and others. Data Transfer Between Host and Peripheral Devices. US Patent 15089177, 2017.

[85] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Text Entry in Immersive Head-Mounted Display-Based Virtual Reality Using Standard Keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 159–166, Mar. 2018.

[86] X. Gu, Y. Zhang, W. Sun, Y. Bian, D. Zhou, and P. O. Kristensson. Dexmo: An Inexpensive and Lightweight Mechanical Exoskeleton for Motion Capture and Force Feedback in VR. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1991–1995, New York, NY, USA, 2016. Association for Computing Machinery.

[87] J. Gugenheimer, D. Dobbelstein, C. Winkler, G. Haas, and E. Rukzio. FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 49–60, New York, NY, USA, 2016. ACM.

[88] J. Gugenheimer, P. Knierim, C. Winkler, J. Seifert, and E. Rukzio. UbiBeam: Exploring the Interaction Space for Home Deployed Projector-Camera Systems. In J. Abascal, S. Barbosa, M. Fetter, T. Gross,

P. Palanque, and M. Winckler, editors, *Human-Computer Interaction – INTERACT 2015*, pages 350–366, Cham, 2015. Springer International Publishing.

[89] J. Gugenheimer, D. Wolf, E. R. Eiriksson, P. Maes, and E. Rukzio. Gyrovr: Simulating inertia in virtual reality using head worn flywheels. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 227–232. ACM, 2016.

[90] J. Gugenheimer, D. Wolf, G. Haas, S. Krebs, and E. Rukzio. Swivrchair: A motorized swivel chair to nudge users' orientation for 360 degree storytelling in virtual reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1996–2000. ACM, 2016.

[91] P. Gurevich, J. Lanir, B. Cohen, and R. Stone. TeleAdvisor: A Versatile Augmented Reality Tool for Remote Assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 619–622, New York, NY, USA, 2012. Association for Computing Machinery.

[92] T. Ha and W. Woo. ARWand: Phone-Based 3D Object Manipulation in Augmented Reality Environment. In *2011 International Symposium on Ubiquitous Virtual Reality (ISUVR)*, pages 44–47. IEEE, June 2011.

[93] N. Haas. Evaluation of "In-VR-Questionnaires". *Bachelor thesis*, Stuttgart, 2017.

[94] J. Hardy and J. Alexander. Toolkit Support for Interactive Projected Displays. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, New York, NY, USA, 2012. Association for Computing Machinery.

[95] S. G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, Nov. 2016.

[96] M. Hassenzahl, M. Burmester, and F. Koller. AttrakDiff work model and measuring tool, 1998.

[97] P. Havig, J. McIntire, and E. Geiselman. Virtual reality in a cave: limitations and the need for HMDs? In P. L. Marasco and P. R. Havig, editors, *Head- and Helmet-Mounted Displays XVI: Design and Applications*, pages 58–63. International Society for Optics and Photonics, SPIE, 2011.

[98] Z. He, F. Zhu, A. Gaudette, and K. Perlin. Robotic Haptic Proxies for Collaborative Virtual Reality. *arXiv preprint arXiv:1701.08879*, 2017.

[99] M. L. Heilig. Sensorama Simulator. U.S. Patent 3.050.870, Aug. 1962.

[100] D. Hein. Interaction Techniques via Smartphone Input in the Context of Task-Based Augmented Reality for Smart Glasses. *Master thesis*, Munich, Aug. 2019.

[101] J. F. M. Helmes, J. Scott, S. E. Hodges, U. Konttori, N. Y.-C. Chen, S. A. Taylor, P. Knierim, G. Wood, and J. W. Fromm. Dual display device. US Patent WO2017053174, 2017.

[102] J. F. M. Helmes, J. Scott, S. A. Taylor, S. E. Hodges, N. Y.-C. Chen, T. A. Grosse-Puppendahl, J. Fromm, P. Knierim, G. Wood, A. Schmidt, and others. Generation of a Modified Ui Element Tree. US Patent 15089268, 2017.

[103] S. Heo, C. Chung, G. Lee, and D. Wigdor. Thor's Hammer: An Ungrounded Force Feedback Device Utilizing Propeller-Induced Propulsive Force. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–4, New York, NY, USA, 2018. Association for Computing Machinery.

[104] W. Heuten, N. Henze, S. Boll, and M. Pielot. Tactile wayfinder: a non-visual support system for wayfinding. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, pages 172–181. ACM, ACM, 2008.

[105] H. Hile, R. Vedantham, G. Cuellar, A. Liu, N. Gelfand, R. Grzeszczuk, and G. Borriello. Landmark-based pedestrian navigation from collections of geotagged photos. In *Proceedings of the 7th international conference on mobile and ubiquitous multimedia*, pages 145–152. ACM, ACM, 2008.

[106] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, X. Guo, P. Moghadasian, and P. Irani. *Consumed endurance: a metric to quantify arm fatigue of mid-air interactions*. a metric to quantify arm fatigue of mid-air interactions. ACM, New York, New York, USA, Apr. 2014.

[107] J. D. Hincapié-Ramos, K. Ozacar, P. P. Irani, and Y. Kitamura. *GyroWand: IMU-based Raycasting for Augmented Reality Head-Mounted Displays*. IMU-based Raycasting for Augmented Reality Head-Mounted Displays. ACM, New York, New York, USA, Aug. 2015.

[108] T. Hirzle, J. Rixen, J. Gugenheimer, and E. Rukzio. *WatchVR: Exploring the Usage of a Smartwatch for Interaction in Mobile Virtual Reality*. Exploring the Usage of a Smartwatch for Interaction in Mobile Virtual Reality. ACM, New York, New York, USA, Apr. 2018.

[109] S. Hodges, J. Scott, N. Chen, S. Taylor, J. Helmes, P. Wright, L. Walker, N. Trim, T. Grosse-Puppendahl, P. Knierim, G. Wood, and T. Denney. Enhancing Smartphone Productivity and Reliability with an Integrated Display Cover. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 349–352, New York, NY, USA, 2018. Association for Computing Machinery.

[110] H. G. Hoffman. Physically touching virtual objects using tactile augmentation enhances the realism of virtual environments. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, pages 59–63. IEEE, 1998.

[111] S. Honey and K. Milnes. The annotated america's cup. *IEEE Spectrum*, 50(9):36–42, Sept. 2013.

[112] A. H. Hoppe, K. Westerkamp, S. Maier, F. van de Camp, and R. Stiefelhagen. Multi-user Collaboration on Complex Data in Virtual and Augmented Reality. In *International Conference on Human-Computer Interaction*, pages 258–265. Springer, 2018.

[113] M. Hoppe. Adjusting virtual worlds to real world feedback limitations while using quadcopters. *Master thesis*, Stuttgart, Aug. 2017.

[114] M. Hoppe, J. Karolus, F. Dietz, P. W. Woźniak, A. Schmidt, and T.-K. Machulla. *VRsneaky: Increasing Presence in VR Through Gait-Aware Auditory Feedback*. Increasing Presence in VR Through Gait-Aware Auditory Feedback. ACM, New York, New York, USA, May 2019.

[115] M. Hoppe, P. Knierim, T. Kosch, M. Funk, L. Futami, S. Schneegass, N. Henze, A. Schmidt, and T. Machulla. VRHapticDrones: Providing Haptics in Virtual Reality through Quadcopters. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pages 7–18, New York, New York, USA, Nov. 2018. ACM.

[116] A. Ibrahim, B. Huynh, J. Downey, T. Höllerer, D. Chun, and J. O'Donovan. ARbis Pictus: A Study of Language Learning with Augmented Reality. *CoRR*, abs/1711.11243, 2017.

[117] B. Irom. Virtual Reality and the Syrian Refugee Camps: Humanitarian Communication and the Politics of Empathy. *International Journal of Communication*, 12(0):23, Sept. 2018.

[118] R. J. K. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-Based Interaction: A Framework for Post-WIMP Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 201–210, New York, NY, USA, 2008. Association for Computing Machinery.

[119] J. Jang and T. Bednarz. HoloSensor for Smart Home, Health, Entertainment. In *ACM SIGGRAPH 2018 Appy Hour*, New York, NY, USA, 2018. Association for Computing Machinery.

[120] S. Jeong, N. Hashimoto, and S. Makoto. A Novel Interaction System with Force Feedback Between Real - and Virtual Human: An Entertainment System: "Virtual Catch Ball". In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, pages 61–66, New York, NY, USA, 2004. ACM.

[121] J. Jung, J. Hong, S. Park, and H. S. Yang. *Smartphone as an augmented reality authoring tool via multi-touch based 3D interaction method*. ACM, New York, New York, USA, Dec. 2012.

[122] A. Kaiser. Entwicklung einer Schnittstelle zwischen Augmented Reality und physikalischen Objekten. *Bachelor thesis*, Stuttgart, Nov. 2016.

[123] H. Kato. ARToolKit. ARToolKit Home Page, `http://www.hitl.washington.edu/artoolkit/`. *hitl.washington.edu*, 1999.

[124] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pages 85–94. IEEE Comput. Soc, Oct. 1999.

[125] D. Kember and L. Gow. Orientations to Teaching and Their Effect on the Quality of Student Learning. *The Journal of Higher Education*, 65(1):58–74, 1994.

[126] D. Kharlamov, B. Woodard, L. Tahai, and K. Pietroszek. TickTockRay: Smartwatch-Based 3D Pointing for Smartphone-Based Virtual Reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 365–366, New York, NY, USA, 2016. Association for Computing Machinery.

[127] B. Kim, H. Y. Kim, and J. Kim. Getting Home Safely with Drone. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 117–120, New York, NY, USA, 2016. Association for Computing Machinery.

[128] Y. R. Kim and G. J. Kim. HoVR-Type: Smartphone as a Typing Interface in VR Using Hovering. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 333–334, New York, NY, USA, 2016. Association for Computing Machinery.

[129] F. Kiss, P. W. Woundefinedniak, V. Biener, P. Knierim, and A. Schmidt. Vum: Understanding requirements for a virtual ubiquitous microscope. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia (MUM)*, page 259–266, New York, NY, USA, 2020. Association for Computing Machinery.

[130] K. Klamka and R. Dachselt. ARCord: Visually Augmented Interactive Cords for Mobile Interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, New York, NY, USA, 2018. Association for Computing Machinery.

[131] H. Klinke. Evaluation of Mid-Air Gestures for Interacting with Large High-Resolution Displays. *Bachelor thesis*, Stuttgart, 2015.

[132] P. Knierim, M. Funk, T. Kosch, A. Fedosov, T. Müller, B. Schopf, M. Weise, and A. Schmidt. UbiBeam++: Augmenting Interactive Projection with Head-Mounted Displays. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, New York, NY, USA, 2016. Association for Computing Machinery.

[133] P. Knierim, D. Hein, A. Schmidt, and T. Kosch. The SmARtphone Controller: Leveraging Smartphones as Input and Output Modality for Improved Interaction within Mobile Augmented Reality Environments. *i-com Journal of Interactive Media*, 20(1):49–61, 2021.

[134] P. Knierim, F. Kiss, M. Rauh, and A. Schmidt. Tangibility is overrated: Comparing learning experiences of physical setups and their virtual equivalent in augmented reality. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia (MUM)*, 2020.

[135] P. Knierim, F. Kiss, and A. Schmidt. Look Inside: Understanding Thermal Flux Through Augmented Reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 170–171, Oct. 2018.

[136] P. Knierim, T. Kosch, A. Achberger, and M. Funk. Flyables: Exploring 3D Interaction Spaces for Levitating Tangibles. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 329–336, New York, NY, USA, 2018. Association for Computing Machinery.

[137] P. Knierim, T. Kosch, J. Groschopp, and A. Schmidt. Opportunities and Challenges of Text Input in Portable Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, New York, NY, USA, 2020. Association for Computing Machinery.

[138] P. Knierim, T. Kosch, M. Hoppe, and A. Schmidt. Challenges and Opportunities of Mixed Reality Systems in Education. In *Mensch und Computer 2018 - Workshopband*, Bonn, 2018. Gesellschaft für Informatik e.V.

[139] P. Knierim, T. Kosch, G. LaBorwit, and A. Schmidt. Altering the Speed of Reality? Exploring Visual Slow-Motion to Amplify Human Perception using Augmented Reality. In *Proceedings of the 1st Augmented Humans International Conference*, New York, NY, USA, 2020. ACM.

[140] P. Knierim, T. Kosch, V. Schwind, M. Funk, F. Kiss, S. Schneegass, and N. Henze. Tactile Drones - Providing Immersive Tactile Feedback in Virtual Reality through Quadcopters. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 433–436, New York, NY, USA, 2017. Association for Computing Machinery.

[141] P. Knierim, S. Maurer, K. Wolf, and M. Funk. Quadcopter-Projected In-Situ Navigation Cues for Improved Location Awareness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, New York, NY, USA, 2018. Association for Computing Machinery.

[142] P. Knierim and A. Schmidt. The virtual office of the future: Are centralized workplaces obsolete? *New Future of Work*, August 2020.

[143] P. Knierim, A. Schmidt, and T. Kosch. Demonstrating thermal flux: Using mixed reality to extend human sight by thermal vision. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia (MUM)*, page 348–350, New York, NY, USA, 2020. Association for Computing Machinery.

[144] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–9, New York, NY, USA, 2018. Association for Computing Machinery.

[145] P. Knierim, P. W. Woźniak, Y. Abdelrahman, and A. Schmidt. Exploring the Potential of Augmented Reality in Domestic Environments. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, USA, 2019. Association for Computing Machinery.

[146] T. Kosch, M. Funk, D. Vietz, M. Weise, T. Müller, and A. Schmidt. DroneCTRL: A Tangible Remote Input Control for Quadcopters. *Proceedings of the 31th Annual ACM Symposium on User Interface Software and Technology*, 2018.

[147] T. Kosch, P. Knierim, P. W. Woźniak, and A. Schmidt. Chances and Challenges of Using Assistive Systems in Education. In *Mensch und Computer 2017 - Workshopband*, Regensburg, 2017. Gesellschaft für Informatik e.V.

[148] J. Kreimeier, S. Hammer, D. Friedmann, P. Karg, C. Bühner, L. Bankel, and T. Götzelmann. Evaluation of Different Types of Haptic Feedback Influencing the Task-Based Presence and Performance in Virtual Reality. In *PETRA '19: The 12th PErvasive Technologies Related to Assistive Environments Conference*, pages 289–298, New York, NY, USA, June 2019. ACM.

[149] A. Kropielnicka. Data Visualization in Augmented Reality. *Bachelor thesis*, Stuttgart, July 2018.

[150] M. W. Krueger, T. Gionfriddo, and K. Hinrichsen. VIDEOPLACE—an Artificial Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 35–40, New York, NY, USA, 1985. Association for Computing Machinery.

[151] M. Kyt o, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, New York, NY, USA, 2018. Association for Computing Machinery.

[152] J.-W. Lin, P.-H. Han, J.-Y. Lee, Y.-S. Chen, T.-W. Chang, K.-W. Chen, and Y.-P. Hung. Visualizing the Keyboard in Virtual Reality for Enhancing Immersive Experience. In *ACM SIGGRAPH 2017 Posters*, pages 35:1–35:2, New York, NY, USA, 2017. ACM.

[153] R. W. Lindeman and H. Noma. A Classification Scheme for Multi-Sensory Augmented Reality. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*, pages 175–178, New York, NY, USA, 2007. Association for Computing Machinery.

[154] R. W. Lindeman, R. Page, Y. Yanagida, and J. L. Sibert. Towards Full-Body Haptic Feedback: the Design and Deployment of a Spatialized Vibrotactile Feedback System. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 146–149. ACM, 2004.

[155] R. W. Lindeman, Y. Yanagida, H. Noma, and K. Hosaka. Wearable vibrotactile systems for virtual contact and information display. *Virtual Reality*, 9(2-3):203–213, 2006.

[156] C. Liu, S. Huot, J. Diehl, W. Mackay, and M. Beaudouin-Lafon. Evaluating the Benefits of Real-time Feedback in Mobile Augmented Reality with Hand-held Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2973–2976, New York, NY, USA, 2012. ACM.

[157] M. C. J. Lizandra, M. A. Raya, C. Monserrat, C. Botella, R. M. ı. a. Baños, and B. Guerrero. Using Augmented Reality to Treat Phobias. *IEEE Computer Graphics and Applications*, 25(6):31–37, 2005.

[158] T. Lokki and M. Grohn. Navigation with auditory cues in a virtual environment. *IEEE MultiMedia*, 12(2):80–86, 2005.

[159] B. Long, S. A. Seah, T. Carter, and S. Subramanian. Rendering Volumetric Haptic Shapes in Mid-Air Using Ultrasound. *ACM Transactions on Graphics (TOG)*, 33(6):181–10, Nov. 2014.

[160] P. Lopes, A. Ion, and P. Baudisch. Impacto: Simulating physical impact by combining tactile stimulation with electrical muscle stimulation. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 11–19. ACM, 2015.

[161] P. Lopes, S. You, L.-P. Cheng, S. Marwecki, and P. Baudisch. Providing Haptics to Walls and Heavy Objects in Virtual Reality by Means of Elec-

trical Muscle Stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1471–1482. ACM, 2017.

[162] R. Luckin and D. S. Fraser. Limitless or pointless? An evaluation of augmented reality technology in the school and home. *International Journal of Technology Enhanced Learning*, 3(5):510–524, 2011.

[163] W. E. Mackay. Augmented Reality: Linking Real and Virtual Worlds: A New Paradigm for Interacting with Computers. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 13–21, New York, NY, USA, 1998. Association for Computing Machinery.

[164] I. S. MacKenzie and R. W. Soukoreff. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, pages 754–755, New York, NY, USA, 2003. ACM.

[165] L. Manning. The Man Who Awoke. *Wonder Stories*, (10), Mar. 1933.

[166] M. Matrosov, O. Volkova, and D. Tsetserukou. LightAir: A Novel System for Tangible Communication with Quadcopters Using Foot Gestures and Projected Image. In *ACM SIGGRAPH 2016 Emerging Technologies*, pages 16:1–16:2, New York, NY, USA, 2016. ACM.

[167] T. Mattelmäki and K. Battarbee. Empathy probes. In *PDC*, pages 266–271, 2002.

[168] S. Maurer. Augmenting Movement with Navigational Aids from a Projector Drone. *Master thesis*, Stuttgart, 2016.

[169] A. J. May, T. Ross, S. H. Bayer, and M. J. Tarkiainen. Pedestrian navigation aids: information requirements and design implications. *Personal and Ubiquitous Computing*, 7(6):331–338, 2003.

[170] S. Mayer, P. Knierim, P. W. Wozniak, and M. Funk. How drones can support backcountry activities. In *21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 6, 2017.

[171] S. Mayer, L. Lischke, J. E. Grønbæk, Z. Sarsenbayeva, J. Vogelsang, P. W. Woźniak, N. Henze, and G. Jacucci. *Pac-Many: Movement Behavior when Playing Collaborative and Competitive Games on Large Displays*. Movement Behavior when Playing Collaborative and Competitive Games on Large Displays. ACM, New York, New York, USA, Apr. 2018.

[172] D. Maynes-Aminzade. Edible Bits: Seamless Interfaces between People, Data and Food. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, 2005.

[173] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2143–2152, New York, NY, USA, 2015. ACM.

[174] M. McGill, A. Ng, and S. Brewster. I Am The Passenger: How Visual Motion Cues Can Influence Sickness For In-Car VR. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5655–5668, New York, NY, USA, 2017. ACM.

[175] P. Milgram and F. Kishino. A Taxonomy of Mixed Reality Visual Displays. *IEICE TRANSACTIONS on Information and Systems*, E77-D(12):1321–1329, Dec. 1994.

[176] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: a class of displays on the reality-virtuality continuum. In *Telemanipulator and Telepresence Technologies*, pages 282–292. International Society for Optics and Photonics, Dec. 1995.

[177] M. Mine, A. Yoganandan, and D. Coffey. *Making VR work: building a real-world immersive modeling application in the virtual world*. building a real-world immersive modeling application in the virtual world. ACM, New York, New York, USA, Oct. 2014.

[178] P. Mohr, M. Tatzgern, T. Langlotz, A. Lang, D. Schmalstieg, and D. Kalkofen. TrackCap: Enabling Smartphones for 3D Interaction on Mobile Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, New York, NY, USA, 2019. Association for Computing Machinery.

[179] S. Moon, C.-K. Lee, S.-W. Nam, C. Jang, G.-Y. Lee, W. Seo, G. Sung, H.-S. Lee, and B. Lee. Augmented reality near-eye display using Pancharatnam-Berry phase lenses. *Scientific Reports*, 9(1):1–10, Apr. 2019.

[180] B. Morschheuser, M. Riar, J. Hamari, and A. Maedche. How games induce cooperation? A study on the relationship between game features and we-intentions in an augmented reality game. *Computers in human behavior*, 77:169–183, 2017.

[181] F. F. Mueller and M. Muirhead. Jogging with a Quadcopter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2023–2032, New York, NY, USA, 2015. Association for Computing Machinery.

[182] F. Müller, J. McManus, S. Günther, M. Schmitz, M. Mühlhäuser, and M. Funk. *Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays*. Assessing Foot-Taps for Interacting with Head-Mounted Displays. ACM, New York, New York, USA, May 2019.

[183] K. Nagai, S. Tanoue, K. Akahane, and M. Sato. Wearable 6-DoF wrist haptic device SPIDAR-W. In *SIGGRAPH Asia 2015 Haptic Media And Contents Design*, page 19. ACM, 2015.

[184] H. Nakamura and H. Miyashita. Augmented Gustation Using Electricity. In *Proceedings of the 2nd Augmented Human International Conference*, New York, NY, USA, 2011. Association for Computing Machinery.

[185] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, H. Hörtner, and C. Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, Mar. 2006.

[186] J. L. Nasar and D. Troyer. Pedestrian injuries due to mobile phone use in public places. *Accident Analysis & Prevention*, 57:91–95, 2013.

[187] M. Nebeling, J. Nebeling, A. Yu, and R. Rumble. *ProtoAR: Rapid Physical-Digital Prototyping of Mobile Augmented Reality Applications*. Rapid Physical-Digital Prototyping of Mobile Augmented Reality Applications. ACM, New York, New York, USA, Apr. 2018.

[188] J. Newman, A. Bornik, D. Pustka, F. Echtler, M. Huber, D. Schmalstieg, and G. Klinker. Tracking for Distributed Mixed Reality Environments. In *Workshop on Trands and Issues in Tracking for Virtual Environments at the IEEE Virtual Reality Conference (VR'07)*, 2007.

[189] Newzoo. Global Mobile Market Report. Technical report, Newzoo, Sept. 2019.

[190] F. Nieuwenhuizen. Evaluation of typing performance in virtual reality on a physical keyboard. *Bachelor thesis*, Stuttgart, 2017.

[191] J.-M. Normand, M. Servières, and G. Moreau. A New Typology of Augmented Reality Applications. In *Proceedings of the 3rd Augmented Human International Conference*, New York, NY, USA, 2012. Association for Computing Machinery.

[192] A. Nowak, P. Knierim, A. Romanowski, A. Schmidt, and T. Kosch. What does the Oscilloscope Say?: Comparing the Efficiency of In-Situ Visualisations during Circuit Analysis. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2020. ACM.

[193] H. Nozaki. Flying Display: A Movable Display Pairing Projector and Screen in the Air. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, pages 909–914, New York, NY, USA, 2014. ACM.

[194] M. Obaid, O. Mubin, C. A. Basedow, A. A. c. a. U nl u er, M. J. Bergstr o m, and M. Fjeld. A Drone Agent to Support a Clean Environment. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, pages 55–61, New York, NY, USA, 2015. ACM.

[195] H. Oh, S.-Y. Yoon, and C.-R. Shyu. How Can Virtual Reality Reshape Furniture Retailing? *Clothing and Textiles Research Journal*, 26(2):143–163, Apr. 2008.

[196] J. Paavilainen, H. Korhonen, K. Alha, J. Stenros, E. Koskinen, and F. Mayra. *The Pokémon GO Experience: A Location-Based Augmented Reality Mobile Game Goes Mainstream*. A Location-Based Augmented Reality Mobile Game Goes Mainstream. ACM, New York, NY, USA, May 2017.

[197] T. Page. Application-based Mobile Devices in Design Education. *International Journal of Mobile Learning and Organisation*, 8(2):96–111, June 2014.

[198] M. Pfeiffer, T. Dünte, S. Schneegass, F. Alt, and M. Rohs. Cruise Control for Pedestrians: Controlling Walking Direction using Electrical Muscle Stimulation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2505–2514. ACM, 2015.

[199] K. Pietroszek, J. R. Wallace, and E. Lank. *Tiltcasting: 3D Interaction on Large Displays using a Mobile Device*. 3D Interaction on Large Displays using a Mobile Device. ACM, New York, New York, USA, Nov. 2015.

[200] C. Pinhanez. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In G. D. Abowd, B. Brumitt, and S. Shafer, editors, *Ubicomp 2001: Ubiquitous Computing*, pages 315–331, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[201] B. Rabowsky. *Interactive Entertainment: A Videogame Industry Guide*. Radiosity Press, 2009.

[202] I. Radu, B. McCarthy, and Y. Kao. Discovering educational augmented reality math applications by prototyping with elementary-school teachers. In *2016 IEEE Virtual Reality (VR)*, pages 271–272, Mar. 2016.

[203] N. Ranasinghe, A. D. Cheok, O. N. N. Fernando, H. Nii, and G. Ponnampalam. Electronic Taste Stimulation. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, pages 561–562, New York, NY, USA, 2011. Association for Computing Machinery.

[204] U. Rashid, M. A. Nacenta, and A. Quigley. *The cost of display switching: a comparison of mobile, large display and hybrid UI configurations*. a comparison of mobile, large display and hybrid UI configurations. ACM, New York, New York, USA, May 2012.

[205] M. V. J. Rauh. Smart.lab: A mobile augmented reality application to substitute an experiment on heat conduction. *Bachelor thesis*, Munich, Oct. 2018.

[206] K. Rehrl, E. Häusler, S. Leitinger, and D. Bell. Pedestrian navigation with augmented reality, voice and digital map: final results from an in situ field study assessing performance and user experience. *Journal of Location Based Services*, 8(2):75–96, 2014.

[207] Y. Rogers. Moving on from Weiser's Vision of Calm Computing: Engaging Ubicomp Experiences. In *Proceedings of the 8th International Conference on Ubiquitous Computing*, pages 404–421, Berlin, Heidelberg, 2006. Springer-Verlag.

[208] S. Rokhsaritalemi, A. Sadeghi-Niaraki, and S.-M. Choi. A Review on Mixed Reality: Current Trends, Challenges and Prospects. *Applied Sciences*, 10(2):636, 2020.

[209] H. Saidi, M. Serrano, P. Irani, and E. Dubois. *TDome: A Touch-Enabled 6DOF Interactive Device for Multi-Display Environments*. A Touch-Enabled 6DOF Interactive Device for Multi-Display Environments. ACM, New York, New York, USA, May 2017.

[210] M. E. C. Santos, A. Chen, T. Taketomi, G. Yamamoto, J. Miyazaki, and H. Kato. Augmented Reality Learning Experiences: Survey of Prototype Design and Evaluation. *IEEE Transactions on Learning Technologies*, 7(1):38–56, Jan. 2014.

[211] M. E. C. Santos, A. Chen, T. Taketomi, G. Yamamoto, J. Miyazaki, and H. Kato. Augmented Reality Learning Experiences: Survey of Prototype Design and Evaluation. *IEEE Transactions on Learning Technologies*, 7(1):38–56, Jan. 2014.

[212] J. Scheible and M. Funk. In-situ-displaydrone: facilitating co-located interactive experiences via a flying screen. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays*, pages 251–252. ACM, 2016.

[213] J. Scheible, M. Funk, K. C. Pucihar, M. Kljun, M. Lochrie, P. Egglestone, and P. Skrlj. Using Drones for Art and Exergaming. *IEEE Pervasive Computing*, 16(1):48–56, 2017.

[214] A. Schmidt, S. Schneegass, K. Kunze, J. Rekimoto, and W. Woo. Workshop on Amplification and Augmentation of Human Perception. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 668–673, New York, NY, USA, 2017. ACM.

[215] D. Schmidt, R. Kovacs, V. Mehta, U. Umapathi, S. Köhler, L.-P. Cheng, and P. Baudisch. Level-ups: Motorized stilts that simulate stair steps in virtual reality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2157–2160. ACM, 2015.

[216] S. Schneegass, F. Alt, J. Scheible, A. Schmidt, and H. Su. Midair displays: Exploring the concept of free-floating public displays. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2035–2040. ACM, ACM, 2014.

[217] S. Schneegass and A. Voit. *GestureSleeve: using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches*. using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches. ACM, New York, New York, USA, Sept. 2016.

[218] J. Scholz and A. N. Smith. Augmented reality: Designing immersive experiences that maximize consumer engagement. *Business Horizons*, 59(2):149–161, Mar. 2016.

[219] B. Schwerdtfeger, D. Pustka, A. Hofhauser, and G. Klinker. Using Laser Projectors for Augmented Reality. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, pages 134–137, New York, NY, USA, 2008. Association for Computing Machinery.

[220] V. Schwind, P. Knierim, L. Chuang, and N. Henze. “Where’s Pinky?”: The Effects of a Reduced Number of Fingers in Virtual Reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 507–515, New York, NY, USA, 2017. Association for Computing Machinery.

[221] 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*, pages 1–12, New York, NY, USA, 2019. Association for Computing Machinery.

[222] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. “These Are Not My Hands!”: Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1577–1582, New York, NY, USA, 2017. Association for Computing Machinery.

[223] A. Sellen, Y. Rogers, R. Harper, and T. Rodden. Reflecting Human Values in the Digital Age. *Communications of the ACM*, 52(3):58–66, Mar. 2009.

[224] M. Serino, K. Cordrey, L. McLaughlin, and R. L. Milanaik. Pokemon Go and augmented virtual reality games: a cautionary commentary for parents and pediatricians. *Current Opinion in Pediatrics*, 28(5):673–677, 2016.

[225] R. Shewbridge, A. Hurst, and S. K. Kane. Everyday Making: Identifying Future Uses for 3D Printing in the Home. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, pages 815–824, New York, NY, USA, 2014. ACM.

[226] B. Shneiderman. The limits of speech recognition. *Communications of the ACM*, 43(9):63–65, Sept. 2000.

[227] B. Shneiderman and C. Plaisant. *Designing the User Interface - Strategies for Effective Human-Computer Interaction (5. ed.).* Addison-Wesley, 2010.

[228] B. Signer and T. J. Curtin. Tangible Holograms: Towards Mobile Physical Augmentation of Virtual Objects. *arXiv preprint arXiv:1703.08288*, 2017.

[229] 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*, pages 3307–3316. ACM, 2015.

[230] M. Slater and S. Wilbur. A framework for immersive virtual environments five: Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, Dec. 1997.

[231] R. W. Soukoreff and I. S. MacKenzie. Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 113–120, New York, NY, USA, 2003. ACM.

[232] M. Speicher, B. D. Hall, and M. Nebeling. What is Mixed Reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–15, New York, NY, USA, 2019. Association for Computing Machinery.

[233] M. A. Srinivasan and R. H. LaMotte. Tactual discrimination of softness. *Journal of Neurophysiology*, 73(1):88–101, Jan. 1995.

[234] J. Steuer. Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4):73–93, Dec. 1992.

[235] M. L. Steven. Virtual Reality. Cambridge University Press, 2016.

[236] E. Strickland. These Researchers Want to Send Smells Over the Internet. *IEEE Spectrum*, Oct. 2018.

[237] M. P. Strzys, S. Kapp, M. Thees, P. Klein, P. Lukowicz, P. Knierim, A. Schmidt, and J. Kuhn. Physics holo.lab learning experience: using smartglasses for augmented reality labwork to foster the concepts of heat conduction. *European Journal of Physics*, 39(3):035703, May 2018.

[238] M. P. Strzys, S. Kapp, M. Thees, J. Kuhn, P. Lukowicz, P. Knierim, and A. Schmidt. Augmenting the thermal flux experiment: A mixed reality approach with the HoloLens. *The Physics Teacher*, 55(6):376–377, Sept. 2017.

[239] Q. Sun, L.-Y. Wei, and A. Kaufman. Mapping Virtual and Physical Reality. *ACM Transactions on Graphics (TOG)*, 35(4), July 2016.

[240] I. E. Sutherland. The Ultimate Display. In *Proceedings of the IFIP Congress*, pages 506–508, 1965.

[241] I. E. Sutherland. A Head-Mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, pages 757–764, New York, NY, USA, 1968. Association for Computing Machinery.

[242] D. Swapp, V. Pawar, and C. Loscos. Interaction with co-located haptic feedback in virtual reality. *Virtual Reality*, 10(1):24–30, 2006.

[243] S. A. Taylor, N. Y.-C. Chen, J. F. M. Helmes, S. E. Hodges, J. Scott, A. Schmidt, P. Knierim, G. Wood, J. W. Fromm, T. Grosse-Puppendahl, and others. Persistent Display Device with Power Harvesting. US Patent 14874896, 2017.

[244] B. Thomas, B. Close, J. Donoghue, J. Squires, P. D. Bondi, and W. Piekarski. First Person Indoor/Outdoor Augmented Reality Application: ARQuake. *Personal and Ubiquitous Computing*, 6(1):75–86, Feb. 2002.

[245] S. Thomas. Future Ready Learning: Reimagining the Role of Technology in Education. 2016 National Education Technology Plan. *Office of Educational Technology, US Department of Education*, 2016.

[246] I. Thomas A Furness. The Super Cockpit and its Human Factors Challenges:. *Proceedings of the Human Factors Society Annual Meeting*, 1986.

[247] R. Thornton and D. Sokoloff. Heat and Temperature Conceptual Evaluation (HTCE). *PhysPort*, 2001.

[248] K. Tsukada and M. Yasumura. Activebelt: Belt-type wearable tactile display for directional navigation. In *International Conference on Ubiquitous Computing*, pages 384–399. Springer, 2004.

[249] V. Vanden Abeele, E. Hauters, and B. Zaman. Increasing the Reliability and Validity of Quantitative Laddering Data with LadderUX. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, pages 2057–2062, New York, NY, USA, 2012. Association for Computing Machinery.

[250] S. Wakamiya, H. Kawasaki, Y. Kawai, A. Jatowt, E. Aramaki, and T. Akiyama. Lets Not Stare at Smartphones while Walking: Memorable Route Recommendation by Detecting Effective Landmarks. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016.

[251] J. Walker, S. Kuhl, and K. Vertanen. Decoder-assisted typing using an HMD and a physical keyboard. In *Extended Abstracts of the the ACM Conference on Human Factors in Computing Systems, CHI*, 2016.

[252] J. Walker, B. Li, K. Vertanen, and S. Kuhl. Efficient Typing on a Visually Occluded Physical Keyboard. In *Proceedings of the 2017 CHI Conference*

on Human Factors in Computing Systems, pages 5457–5461, New York, NY, USA, 2017. ACM.

[253] Y. Wang, F. B. Neubauer, H. R. Lüscher, and K. Thurley. GABAB receptor-dependent modulation of network activity in the rat prefrontal cortex in vitro. *European Journal of Neuroscience*, 31(9):1582–1594, May 2010.

[254] S. G. Weinbaum. *Pygmalion's Spectacles*. The Floating Press, 1949.

[255] M. Weiser. The Computer for the 21st Century. *ACM SIGMOBILE Mobile Computing and Communications Review*, 3(3):3–11, July 1991.

[256] T. Weißer. Anpassung der Routenberechnung zur Stressreduktion beim Autofahren. *Diplom*, Stuttgart, 2016.

[257] P. Wellner. Interacting with Paper on the DigitalDesk. *Communications of the ACM*, 36(7):87–96, July 1993.

[258] C. Wheatstone. XVIII. Contributions to the physiology of vision.—Part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London*, pages 371–394, 1838.

[259] A. Wilson, H. Benko, S. Izadi, and O. Hilliges. Steerable Augmented Reality with the Beamatron. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, pages 413–422, New York, NY, USA, 2012. Association for Computing Machinery.

[260] C. Winkler, J. Seifert, D. Dobbelstein, and E. Rukzio. Pervasive Information through Constant Personal Projection: The Ambient Mobile Pervasive Display (AMP-D). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 4117–4126, New York, NY, USA, 2014. Association for Computing Machinery.

[261] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, June 1998.

[262] K. Wolf, M. Funk, P. Knierim, and M. Löchtefeld. Survey of Interactive Displays through Mobile Projections. *International Journal of Mobile Human Computer Interaction*, 8(4):29–41, Oct. 2016.

[263] K. Wolf, M. Funk, P. Knierim, M. Löchtefeld, and S. Boring. Interactive Displays through Mobile Projection. In *Proceedings of the 17th*

International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pages 1126–1129, New York, NY, USA, 2015. Association for Computing Machinery.

[264] P. W. Woźniak, P. Knierim, and M. Hoppe. Health, fun, and engagement. *XRDS: Crossroads, The ACM Magazine for Students*, 25(4):16–20, July 2019.

[265] H.-K. Wu, S. W.-Y. Lee, H.-Y. Chang, and J.-C. Liang. Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62:41–49, 2013.

[266] R. Xiao, C. Harrison, and S. E. Hudson. WorldKit: Rapid and Easy Creation of Ad-Hoc Interactive Applications on Everyday Surfaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, pages 2889–2890, New York, NY, USA, 2013. Association for Computing Machinery.

[267] K. Yamaguchi, G. Kato, Y. Kuroda, K. Kiyokawa, and H. Takemura. A Non-grounded and Encountered-type Haptic Display Using a Drone. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, pages 43–46. ACM, 2016.

[268] C. Yu, Y. Gu, Z. Yang, X. Yi, H. Luo, and Y. Shi. Tap, Dwell or Gesture? Exploring Head-Based Text Entry Techniques for HMDs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4479–4488, New York, NY, USA, 2017. Association for Computing Machinery.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

München, den 22.6.2020

Pascal Knierim